



# *Metadata Best Practices Illustrated with the CEDAR Workbench*

November 2018

John Graybeal

[jgraybeal@stanford.edu](mailto:jgraybeal@stanford.edu)



**BMIR**

Stanford Center for  
Biomedical Informatics Research

# *What is the CEDAR Workbench?*

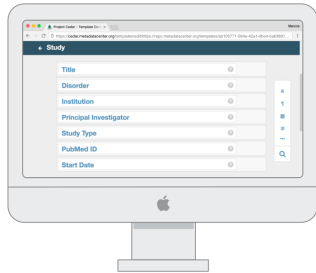


CEDAR

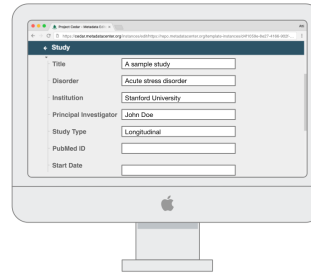
---

[metadatacenter.org](http://metadatacenter.org)

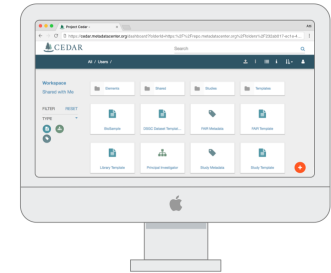
# CEDAR Workbench



**Template Designer**



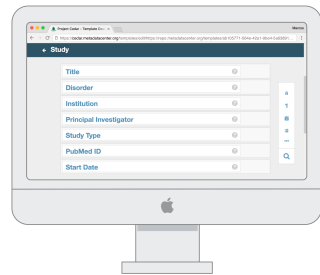
**Metadata Editor**



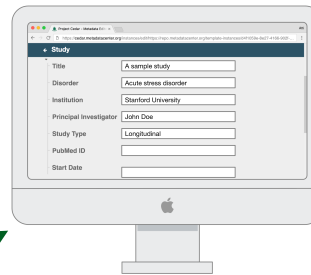
**Resource Manager**

- 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.**

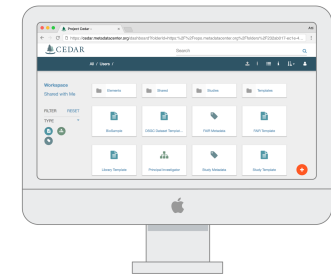
# CEDAR Workbench



Template Designer



Metadata Editor



Resource Manager

**Terms**

**Metadata**



**With semantic services  
(vocabularies) from BioPortal**

**Public Repositories**



**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

The screenshot shows the CEDAR metadata center interface. At the top, the CEDAR logo and 'metadatacenter.org' are on the left. A search bar on the right contains the text 'Drone'. Below the search bar, a breadcrumb trail reads 'All / Users / Sarah Swanz / ESIP Drone Minimum Metadata'. On the left sidebar, there is a 'New +' button and a 'Workspace' section with 'Shared with Me'. Below this is a 'FILTER' section with 'RESET ALL' and a 'TYPE' section with three icons: a document, a cube, and a cube with a tag. At the bottom of the sidebar is a 'VERSION' section. The main area displays a table of metadata templates. The first row, 'Drone Minimum Information', is highlighted with a red box. The table has columns for 'Title', 'Created', and 'Modi...'. The data rows are as follows:

| Title                              | Created | Modi... |
|------------------------------------|---------|---------|
| Drone Minimum Information          | 7/18/18 | 9/4/18  |
| Drone Project Information          | 7/18/18 | 8/30/18 |
| Drone Minimum Information metadata | 7/20/18 | 7/20/18 |
| 1. Introduction to CEDAR Talk      | 7/19/18 | 7/19/18 |
| Drone Platform Information         | 7/18/18 | 7/19/18 |
| Drone Site Information             | 7/18/18 | 7/18/18 |
| Drone Fix Information              | 7/18/18 | 7/18/18 |
| Drone Flight Information           | 7/18/18 | 7/18/18 |
| Drone Sensor Information           | 7/18/18 | 7/18/18 |

## BP1: Target your metadata standards

### Create fields and elements that match your needs

Drone Platform Information

Name

Drone Platform Information

Identifier

Description

Description of the unmanned craft (drone, UAV, RPAS) α

A

\*

Enter Field Name

Platform Type

Enter Preferred Label

Enter Field Help Text

The class of vehicle being described (e.g., submersible, unmanned aerial vehicle)

OPTIONS

VALUES

MULTIPLE

REQUIRED

SUGGESTIONS

HIDDEN

| Name                                      | Type           | Source       | Identifier          | No. Values |  |  |
|---|----------------|--------------|---------------------|------------|--|--|
| Unmanned Aerial Vehicle (Drone)           | Ontology Class | CEDARPC      | ...e9f-005056010073 | 1          |  |  |
| Autonomous Unmanned Aerial Vehicle (AUAV) | Ontology Class | CEDARPC      | ...9c1-005056010088 | 1          |  |  |
| Platform                                  | Branch         | PLATFORM-MMI | Platform            | -          |  |  |

SEARCH






A

#



## BP1: Target your metadata standards

### Include fields and elements in your template


  Drone Minimum Information   




Name


Drone Minimum Information|


Identifier


Description


Information needed to describe all aspects of drone data 


  1 .. N 




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



### 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

Platform Type\*

Dropdown menu for Platform Type:

- TestTetheredConstraintPlatforms
- TestWithCrewPlatform
- TetheredBalloon
- Towfish
- Unmanned Aerial Vehicle (Drone)
- VerticalProfiler



# *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**