

---

---

# Namespace Dictionaries

— Preparing for IM 2.0.0.0 —

---

---

# What we were thinking 8-10 years ago...

- ❑ Discipline Dictionaries would...
    - ❑ Standardize use of discipline terminology across data sources
    - ❑ Stabilize quickly after an initial development phase
    - ❑ Follow the conventions and constraints of the core dictionary closely
    - ❑ Be under configuration control once stable
  
  - ❑ Mission Dictionaries would...
    - ❑ Be created by missions
    - ❑ Be created using some sort of “tool”, TBD
    - ❑ Be allowed latitude regarding core dictionary conventions
-

# Turns out...

## ❓ Discipline Dictionaries do...

- ❓ Standardize use of discipline terminology across data sources, when they exist
- ❓ Usually stabilize quickly after an initial development phase
- ❓ Not always follow the conventions and constraints of the core dictionary closely
- ❓ Not exist under any configuration control at all

## ❓ Mission Dictionaries are...

- ❓ Created by missions with a fair amount of autonomy
  - ❓ Created using some sort of “tool”, either LDDTool or a system overlaying LDDTool
  - ❓ Are allowed some latitude regarding core dictionary conventions for naming and minor structure variations
-

# Surprises

- ❓ Because of the ROSES archiving mandate, most “local” dictionaries will not come from “missions”.
  - ❓ Discipline dictionary development has lagged.
  - ❓ The relationships between and among discipline dictionaries and the core dictionary are not easy to either trace or explain to users.
  - ❓ LDDTool, originally proposed in concept as a stop-gap or internal tool, has become the cornerstone for dictionary development, and provides key features for implementing constraints and recommending best practices.
-

# Red Flags Linked to Dictionaries

- The lack of published development and release schedules, or any public venue for issue reporting, contributes to the perceived opacity reported as an issue for PDS4 generally in the PDS Roadmap.
  - The dependency of dictionaries on the IM version (and thus the core dictionary) has resulted in update dependencies that stewards find onerous and gratuitous.
  - The relation of discipline dictionary versions to IM versions and to other discipline dictionaries, over and above any actual dependencies, has proved to be difficult to document and to explain to users.
-

# CCB-156

This filing raised an issue of inconsistent techniques being applied in various discipline dictionaries to achieve the same result.

The tiger team undertook an analysis of released discipline dictionaries to derive best practices and develop recommendations. Three subsequent CCB issues were filed on “Best Practice”- and “Improved Performance”-type issues. Implementations are already underway.

Discussion of the results of the analysis has lead to a broader review of the state of the development art for both discipline and non-discipline dictionary creators. That review is the focus of this discussion.

---

# Why Now?

1. It's time
  2. PDS Roadmap
  3. Migration
  4. ROSES
  5. Missions
-

# Major Themes Emerging from the Review

Eleven broad areas of concern were identified as a result of the analysis.

Three key themes address nearly all of them.

1. Versioning & Provenance
  2. Build & Release
  3. Development & Configuration Control
-



# Some Terminology for this Discussion

**Dictionary** - The definition of a namespace within the IM; also the set of schema files currently used to define a namespace for label creation and validation purposes

**Discipline Dictionary** - A dictionary developed and maintained by PDS personnel for use in defining metadata that spans node or data source boundaries.

**Project Dictionary** - A dictionary developed as part of a data archiving activity to define contextual metadata specific to the data source or developer

**Mission Dictionary** - A project dictionary developed for a data source which is defined or generally known as a “mission”



# The Role of LDDTool

*LDDTool builds the full information model internally, with the additional dictionary namespace extension, as part of its operation.*

This ensures extension compatibility and consistency with the core IM, within the constraints of the input file and processing. Other tools use LDDTool as their dictionary construction engine.

In conjunction with the *Ingest\_LDD* input structure, LDDTool also ensures that underlying modeling conventions in the core are carried through to the newly defined namespace.

---

# Versioning & Provenance



# Wants

1. Track the development history of a dictionary, independent of IM development
  2. Track the IM development history
  3. Ensure all dictionaries referenced in a single label are part of the same IM version
  4. Locate dictionaries belonging to a specific IM version
-

# Considerations: Versioning

New data preparers may or may not understand the relationship between discipline dictionaries and the core.

It is difficult, even with documentation, for a user to collect a cohesive set of core + discipline schemas for use from the release pages.

It is difficult to determine what was “current” for a previous IM version and discipline dictionaries.

Requiring dictionary stewards to make cosmetic changes in a static dictionary for a scheduled build is not particularly good practice.

---

# Considerations: Dictionary Development History

Change tracking is an essential part of provenance.

Tracking changes formally is a software industry best practice.

We do have the <Modification\_History> class available for (re-)use, or extension for more specialized use.

---

# Build & Release



# Wants

1. Discipline dictionaries released simultaneously with the core IM.
  2. Discipline dictionary references to the IM version namespace managed at build time.
  3. Stewards have to meet build deadlines only when they have a new dictionary being included, or substantive changes to make to an existing dictionary.
  4. Stewards submit the Ingest\_LDD file only - the build and release happens automatically.
-



# Considerations: Building

Currently, stewards are required to build their own schemas and submit them for release.

LDDTool itself is rebuilt for each new IM version. Dictionary stewards must update the LDDTool installation before building their dictionaries against a new IM release.

In the input *Ingest\_LDD*, updating a stable discipline dictionary to build against the latest IM requires changing only references to the IM version, which is generally done manually.

---

# Considerations: Release

Discipline dictionaries are released without regard to either IM builds or other discipline dictionary releases.

There is no mechanism to announce either imminent or *post facto* discipline dictionary releases.

---

# Development & Configuration Control



# Wants

1. All users (internal and external) need a problem reporting mechanism for the core IM and all discipline dictionaries (and related software).
  2. Discipline dictionaries should be included in impact assessments for all proposed changes or additions to the core IM.
  3. Changes and enhancements to discipline dictionaries should be trackable.
  4. Discipline dictionaries should be as rigorous in validation as the core IM.
-

# Considerations: Development

New discipline dictionaries do tend to go through a period of rapid change before stabilizing.

Discipline dictionary creation seems to have become a one-person task, contrary to the original “small team of experts” paradigm.

Discipline dictionary development should be strongly based in validation, which does require some specialized programming skill.

---

# Considerations: Configuration Control

There is no mechanism, formal or informal, for getting wider review and acceptance for discipline dictionaries that span node boundaries.

There is no formal mechanism for reporting issues with a discipline dictionary or requesting changes; or any recourse if the issues are not addressed.

There is no oversight group to consider how changes in discipline dictionaries might affect end-users, data preparers in development, and software either existing or in development designed around existing discipline metadata.

There is no standard for what constitutes “testing” a discipline dictionary for acceptance and release.

---

**Now Consider...**