Descriptive Metadata Specification

The Metadata Implementation Working Group (M-IWG) released a draft of its Descriptive Metadata Specification, which was approved in Nov 2017. The following documents present a summary of the work. Additional documentation will be published as the specification moves into implementation.

Provided by: DLP Metadata Implementation Working Group

Last Revised Date: Nov 2017

Status: Approved

Reviewed by: Content Display FRG, Stakeholder Group, Core Team, Metadata Working Group

Approved by: DLP Steering Group, Nov 2017

- DLP Working Group Context
- Descriptive Metadata Context
- Relationship to Emory Core Metadata
- Process for Analysis
- Scope of Deliverable
- Not Included in this Deliverable
- Data Dictionary Notes
- Links to Descriptive Metadata Specification

DLP Working Group Context

The broad charge for the Metadata Implementation Working Group (M-IWG), a subgroup of the existing Metadata Working Group and of the DLP Project structure, is: to provide metadata requirements and specifications for the Samvera repository (descriptive, technical, preservation, rights, administrative) and provide additional normalization and mappings as needed for migration.

As part of the larger Digital Library Program’s broad charge to migrate to the Samvera framework, our metadata migration planning includes translating current state data from multiple systems, multiple metadata schemas, and multiple data models into a new metadata framework tailored to Fedora 4, the Portland Common Data Model, and RDF.

Descriptive Metadata Context

This document includes specifications for Descriptive Metadata only. Additional types of metadata (e.g. Preservation, Administrative/Workflow) are analyzed as separate Working Group deliverables and are defined in the M-IWG’s team glossary. The Metadata Implementation Working Group uses the following definition to delineate Descriptive Metadata from other types of metadata, based on the Emory Core Metadata definition:

Describes content for search and discovery contexts -- it helps connect users to resources, and provides important context about a resource once it is discovered. This type of metadata drives the ability to search, browse, sort, and filter information.

Note: some other metadata elements may identified to be displayed to end-users, such as selected technical file details; those indications will be documented in separate deliverables.

Relationship to Emory Core Metadata

The DLP Descriptive Metadata Specification reflects a large-scale, implementation-centric application of the Emory Core Metadata standard. While Core Metadata was designed to be standards and implementation agnostic, the DLP specification is designed specifically for implementation in a Samvera-based application and is much more granularly defined. Core Metadata equivalence in some cases is achieved through multiple metadata fields specified for a single Core Metadata element/concept (e.g. Rights/Access Statement). The DLP specification meets and in many cases exceeds Core Metadata requirements.

The group also reviewed Emory Core Metadata for potential revisions, since the standard was first released in 2015 and the DLP will be its largest implementation to date. After review by the larger Metadata Working Group, a separate proposal was developed and approved by Cabinet, which includes clarifications to the Emory Core Metadata standard’s levels of obligation, along with incremental modifications to the standard by lowering obligation levels for Description as well as Subject - Keywords. These adjustments bring greater alignment between the DLP specification and Core Metadata, as well as making it easier for future projects to achieve Core compliance.

Process for Analysis

In order to produce a standardized Descriptive metadata profile for the DLP repository, the team’s goal was to define a metadata property once, provide consistent mappings and documentation for usage, and enable it for re-use across all/any applicable workflow scenarios. As an early artifact, the group established a Data Dictionary to document current and future state metadata needs. The group inventoried a set of six systems, based on the project’s Scope Statement, and pre-normalized metadata across any variant content models within those systems. The resulting inventory initially included over 200 unique metadata element entries, but was eventually reduced to 67. The group collectively reviewed the consolidated inventory of elements in stages based on rough groupings (e.g. Creators/Contributors, Subjects, Titles, Identifiers). During the review process, team members clarified and revised aspects identified in the Data Dictionary for each element based on their respective workflow needs balanced against the broader repository environment context.
Scope of Deliverable

The Descriptive Metadata Profile presented in this document presents a summary view of a consolidated set of “semantic units” (implementation/standards-agnostic metadata containers) along with definitions, levels of requirement, usage and implementation notes, along with equivalencies to Core Metadata, to be implemented in the new repository. These units represent a local Emory cross-walk of descriptive repository metadata to be migrated, relative to the following systems:

- DAMS
- Digitized Books
- ETDs
- Open Emory
- The Keep
- Dataverse

More detailed inventory information and technical detail not presented in this summary is available in the group’s full Descriptive Metadata inventory worksheets (available upon request). Additional, formal documentation will be finalized based on these conventions in DLP’S Implementation Phase as metadata is put into production.

Not Included in this Deliverable

Some limitations were applied to the group’s analysis and resulting deliverables:

- Full MARC data scheme mapping for Digitized Books: analysis of existing data was limited to elements supporting basic discovery (not all possible MARC fields)
- Dataverse: analysis was limited to Citation Metadata only (not all specialized discipline options)
- Full analysis of Rushdie extracted files’ metadata (contingent upon future data models and migration strategies)
- Full metadata specifications for Collections, Files, or Agents (dependent upon larger data management and migration strategies; additional information available upon request)
- Detailed system-level strategies for migration (while examples of original source records are captured in the inventory, we recommend reviewing again on a per-system basis as part of migration planning)
- Final RDF mappings (contingent upon implementation of final data models, and ongoing Samvera community work for Hyrax and MODS to RDF mappings)
- Final end-user display labels and facets (to be confirmed through user research)
- Other categories of metadata (Preservation, Technical, Administrative, and Rights metadata beyond user-facing rights are under development as separate DLP project deliverables)

Data Dictionary Notes

The last section of this document includes links to views of summarized semantic unit definitions (additional inventory data and spreadsheet columns have been omitted for easier review). The views that follow include the following information:

<table>
<thead>
<tr>
<th>Label</th>
<th>Label or name for the semantic unit; final/alternate display labels can be provided in future phases for end-user display</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>Provides a definition of the semantic unit</td>
</tr>
<tr>
<td>Core Metadata Equivalent</td>
<td>Indicates an equivalent Core Metadata element</td>
</tr>
<tr>
<td>Indexed</td>
<td>Indicates if the value is indexed for search</td>
</tr>
<tr>
<td>Obligation</td>
<td>Level of requirement for the metadata entry (based on Core Metadata conventions):</td>
</tr>
<tr>
<td>Required: must be provided</td>
<td></td>
</tr>
<tr>
<td>Required if Applicable: required if the information can be determined and if required for a specific workflow or type of content</td>
<td></td>
</tr>
<tr>
<td>Recommended: not required, but recommended to enrich user experience</td>
<td></td>
</tr>
<tr>
<td>Optional: not required, not widely applicable</td>
<td></td>
</tr>
<tr>
<td>Current Workflow/Usage</td>
<td>Includes field-level usage across systems, cross-referenced to the original inventory entry for each system (e.g. DAMS31, OE27). See the Key to System Numbering Abbreviations below.</td>
</tr>
</tbody>
</table>

Key to System Numbering Abbreviations:

- **DAMS** Emory’s “DAMS” product, also known as Extensis Portfolio Server
- **DB** Digitized Books application
- **OE** Open Emory application
- **ETD** Emory Theses and Dissertations Application
- **KEEP** The Keep application
Links to Descriptive Metadata Specification

View detailed metadata specification

The following additional links provide filtered, summary views to the metadata specification.

By Core Metadata:
- Core Metadata equivalents

By level of requirement:
- Required
- Required if Applicable
- Recommended
- Optional

By current-state system applicability:
- DAMS
- Digitized Books
- ETDs
- OpenEmory
- The Keep
- Dataverse

To exit a filter view, click on the “x” in the black toolbar as shown below.

For more detailed information about any of the metadata summarized here, please contact Emily Porter, M-IWG Convener.