

Machine-readable Metadata Modeling Specification (M3) Working Group Final Report

Submitted to the Samvera Metadata Interest Group 1/28/2020

The Machine-readable Metadata Modeling Specification (M3) Working group set out to draft a shared specification for a machine-readable format for defining and documenting metadata models. The [full group charge is available on the Samvera Wiki](#).

The near and long term benefits of the work include:

- Automating creation and maintenance of in-sync, up-to-date human-readable documentation and batch data entry templates
- Enable the development of code/gems/microservices which are not tied to specific institutional metadata models and are thus easier to reuse
- Facilitate sharing and reuse of metadata models
- Support formal releases and versioning of metadata models
- Enable the development of more flexible infrastructure through abstraction of metadata
- With supporting infrastructure, simplify the process for adding new metadata elements or changing metadata-based configuration settings, reducing work required and opportunities for errors

The working group was active from February and July of 2019, with work including:

- Identification of information key to describing data elements/properties (or meta-metadata properties) within the scope of the charge
- Establishing names, definitions, and usage/syntax guidelines for these meta-metadata properties
- Reviewing possible encoding formats for the specification
- Testing the developing standard by creating example m3.yml encoded data models, and iteratively revising the standard as issues were identified

The specification and resulting data dictionary provide a common terminology that can be used for data model definition and discussions. This shared vocabulary can help developer and metadata analyst collaborations, as well as enable usage of future work on documentation generation and code configuration.

The group also created a JSON Schema for the M3 specification that defines the meta-metadata properties and the syntax for the creation of YAML encoded M3 metadata models. Creating a JSON

Schema for the specification allows us to assert properties of the specification in a machine-validatable way, making it much easier to ensure consistency and compliance with the specification, and laying the foundation for machine-readability.

Next Steps and Future Work Opportunities

There are several areas of work identified for building on and leverage the initial specification.

The first is establishing a specification maintenance group. We propose that a group be formed along the lines of the URI Management group, under the umbrella of the Metadata Interest Group, to define processes for reviewing contributions and suggestions to the specification, and perform these responsibilities. The working group established a GitHub repository for the specification, which provides us with the mechanics for tracking issues and suggestions, and reviewing and approving changes.

The second opportunity is to develop shared community knowledge and tools for the generation of documentation outputs (e.g. data dictionaries, ingest templates, etc.) from m3.yml data models. Preliminary investigation by the M3 Working Group identified Jekyll as a potential tool for some of the desired outputs. Work needed in this area includes recommendations for set-up and configuration of Jekyll for this purpose, the identification of documentation outputs useful for different audiences and uses, and the creation of templates which can be shared and reused by institutions using the m3.yml encoding specification. This work would have an immediate and tangible benefit for metadata analysts, and by providing this benefit would, we believe, increase uptake of the standard for expressing metadata models, even beyond the Samvera-community. There are two components of this work which require different areas of expertise; development and operations, and metadata description and management. It is recommended that either a mixed expertise working group, or two separate groups are formed. Following the completion of this work, it is recommended that instructional materials or workshops be developed to support the adoption of the specification by metadata analysts.

Third, is connecting the m3.yml encoded data models to the Hyrax codebase for simplified code configuration. This idea has already been demonstrated by several existing projects (Dog Biscuits, Scooby Snacks) and future directions include adoption of the M3 specification and integration into the shared community codebase. The m3.yml specification developed includes many properties which could be utilized for code configuration, with different levels of effort required for different functions and configuration options. An incremental approach could be taken to this work, identifying configurations which could be integrated with minimal code impact and/or have already been tested and demonstrated in existing projects utilizing a data model configuration file. Fuller or more complex functions could be evaluated for addition in the future when the benefits and viability of the approach are better demonstrated.