Presentations

Benjamin Riesenberg

Selection of a presentation format for a MAP will depend on the local needs and applications that it is designed to support. A MAP for use in the creation or transcription of metadata values by catalogers and other staff should provide human-readable guidance, perhaps including detailed instructions for creating, selecting, and formatting values. When using metadata-creation tools, a machine-readable MAP can pass customizations and constraints such as value data type, cardinality, and VESs for the selection of values to a data-entry form. In these cases an additional human-readable format may also be helpful during descriptive work. For some applications, metadata values for digital objects such as file format, size, and checksum may be generated or extracted based on the specifications in a machine-readable MAP, with no need for human-readable guidance at the point of metadata creation.

Natural Language MAPs

To successfully implement any MAP, the requirements for a metadata model will need to be provided in human-readable format at some point, even if only as a set of specifications provided to software developers. A large portion of existing MAPs have been designed expressly for reading and use by humans. This makes sense in the context of libraries, where metadata is often created by staff who need a comprehensive set of instructions to perform their work.

The Metadata Application Profile Clearinghouse Project reflects the prevalence of human-readable MAP presentations. This online resource provides access to MAPs submitted by a variety of organizations, with a focus on those created for use in describing collections in digital repositories. While the Clearinghouse provides only a very small cross section of MAPs, the prevalence of human-readable formats is remarkable: all eighteen organizations that have submitted MAPs to date use one or more human-readable versions, and only one of these has provided versions for machine processing.[1]

Many questions arise for metadata creators in the process of describing information resources. Content standards for library cataloging are complex, and the need for guidance makes it unsurprising that human-readable MAP formats are the most commonly used in libraries, as they can provide the rules for generating metadata alongside examples of properly formed values and other relevant information. The popularity of human-readable MAP formats means that there are abundant examples in this category.

The BIBCO Standard Record

The BIBCO Standard Record (BSR) RDA Metadata Application Profile is a baseline set of elements applicable to the description of a wide variety of resource formats commonly collected by libraries. It was created by the Library of Congress Program for Cooperative Cataloging's (PCC) BIBCO program in support of its work to improve the quality of bibliographic description and support efficient cataloging. The ninety-six (as of fiscal year 2020) BIBCO member institutions and "funnels"—groups of libraries or individual catalogers that work together to contribute records—as well as other institutions that have adopted the standard voluntarily, use it as fundamental guidance for creating catalog records.[2]

The BSR is presented in a human-readable, primarily tabular format and provides information including the following:

- headings enumerating physical and intellectual entities for description—for example, "Identifying

Works & Expressions," "Identifying Manifestations & Items," and "Describing Carriers"
- properties (referred to in the BSR as "elements") grouped under these headings and others

For each element, the following information may be given:

- for each element taken from the Resource Description and Access (RDA) set of properties, a link to detailed online guidance where available
- notes including information about specific formats for which the element is recommended or required, additional instructions for recording values, and information about where to find additional guidance

Wikidata Application Profiles

Wikidata is a free and open linked-data knowledge base and, like the ubiquitous Wikipedia platform, a project of the nonprofit Wikimedia Foundation. As evidenced by a Library of Congress PCC pilot project launched in 2020 to explore its use, with more than seventy institutional pilot participants including many college and university libraries in the United States, its visibility as both an editing tool and publishing platform for bibliographic description is increasing.[3]

Many communities are developing and publishing human-readable MAPs to guide resource description in the Wikidata platform, including MAPs for use describing books, periodicals, and video games. The MAP published by WikiProject Books provides an enumeration of several conceptual entities that may be used to create bibliographic descriptions, including written works; versions, editions, or translations; exemplars; and manuscripts. Properties recommended for use with each are provided in a tabular format. For each property, information including the following is provided:[4]

- a property label and an ID linking to a full definition
- a required data type for values
- a property description providing brief guidance for entering values
- examples of use, linking to existing item descriptions

Encoded MAPs

The continuing evolution of software tools used to create, manage, and serve metadata provides increasing opportunities for MAPs to integrate directly with them. As a result, growing numbers of encoded,

machine-readable MAPs are available as examples. Many of these MAPs have been developed for RDF applications.

BIBFRAME Profiles

BIBFRAME Profiles are MAPs for describing specific kinds of entities: instances of the resource classes defined in the BIBFRAME RDF vocabulary.[5] They encode definitions in a way that can be interpreted by a specific software tool, the BIBFRAME Editor, which uses the encoded information to generate data-entry forms. The form input provided by a user is combined with information from the Profile to create an RDF metadata instance.

For example, the class of each entity described in a generated metadata instance is taken from the profile. While the Editor uses labels for properties to make the data-entry form more readable for catalogers, properties in the instance are identified by the IRIs specified for them in the profile and paired with the value or values entered by users. This combining of user input with information from a MAP may be clearer if we look specifically at the information that can be included in a BIBFRAME Profile:

- Basic metadata about the profile itself, including an author, date, title, and description; this assists users in managing profiles and selecting them for use.
- A label, IRI, and guiding statement for each entity type that can be described using the profile.
- A set of properties, each identified with an IRI, that may be used for each entity type.

For each property, the following information may be specified:

- whether a value is required
- whether multiple values can be entered
- whether the interface will prompt entry of a literal value or an IRI selected using a query interface in the editor
- a VES from which IRI values may be selected
- a data type to be assigned to literal values
- a default value

The Library of Congress has developed the BIBFRAME Profile Editor software package, which is accessible online and can be used to create, view, edit, and export BIBFRAME Profiles.[6]

Sinopia Resource Templates

Like the BIBFRAME Editor, the Sinopia Linked Data Editor is a tool for creating linked-data resource descriptions and requires encoded information to

structure data-entry forms and generate data from user input. These MAPs were originally encoded using a structure very similar to that of a BIBFRAME Profile. Sets of entity types for description, and properties and value constraints for use with each, were defined and encoded as a single file. Following recent development of the Editor software package, this information is now organized by single entity types, resulting in a resource template that includes a single entity definition, properties for use with the defined resource type, and value constraints for each property. These resource templates provide much the same information as that given for an entity type in a BIBFRAME Profile and also include basic metadata about the templates themselves.

Additionally, Sinopia resource templates are themselves sets of RDF triples, constructed using terms from the Sinopia Vocabulary, published in 2020.[7] Because resource templates are constructed as RDF graphs in the same manner as descriptive metadata sets, they may, like metadata sets, be created and edited using the Linked Data Editor.

Hybrid MAPs

MAP information is always required in human-readable form at some point during the implementation process. In many cases human-readable formats play the primary role, providing information essential to knowledge workers creating resource descriptions. At the same time, new and evolving software platforms for metadata creation are providing more opportunities for encoded MAPs to integrate with them directly, enforcing requirements and value constraints more consistently by integrating with processes for metadata creation and management.

But this opportunity may also mean there is a need to create and maintain two separate MAP formats for a single implementation—one to provide guidance to humans, and one to pass information to software. The ability to use a single MAP format for both humans and machines would be ideal in such cases and would avoid duplicating work. In cases where a single format cannot meet all needs, implementers may benefit from tools to convert back and forth so that updates and changes need be made in only one place.

One MAP for Two Purposes

Any MAP encoded for machine processing can be considered human-readable. Realistically, however, this is limited to people with knowledge in multiple areas, including the MAP's domain model, its data serialization and structure, and the systems that will process it. While presenting encoded MAPs as reference for human users may be efficient, the required knowledge

is a significant barrier to readability. This challenge may be lessened by using a machine-readable format that is relatively simple and accessible.

VALIDATION CODE AS MAP

A number of coding languages exist for the purpose of validating data; one well-known example is the XML Schema Definition Language. Languages such as this allow users to encode requirements for metadata instances in a precise way, and this code can be processed along with metadata in order to accurately and efficiently identify portions of an instance that don't conform to requirements.

Chapter 2 discussed the development and implementation of MAPs primarily in terms of work that takes place prior to or during metadata creation. Validation code can be integrated with data-entry tools to enforce constraints at the time of metadata creation, but it is often used for assessment or quality control afterward. Despite this difference, validation code is useful to consider in any discussion of MAPs. Using it, we can record requirements for the same primary facets of a metadata model that we've discussed up to now—entities, properties, and values.

SHAPE EXPRESSIONS LANGUAGE

The Shape Expressions Language (ShEx) is focused specifically on RDF data and allows users to define conditions that an RDF graph should meet. As with other data-validation languages, it can be used with large quantities of data to quickly and accurately identify portions that don't conform to specified conditions.

Shape Expressions Compact Syntax (ShExC) provides a syntax for writing Shape Expressions schemas that is designed for human readability. ShExC is notated using terms that are relatively well-known from the RDF data model, widely used XML Schema data type terms, and intuitive labels for constraints. Because of this, users already familiar with RDF will have relatively little trouble reading and comprehending a ShExC schema (see figure 5.1), although writing one requires more detailed knowledge of the syntax.

Uptake and usage of this language may increase further given the announcement in May 2019 that Shape Expressions schemas would be enabled within the Wikidata platform, including the addition of a new entity type, the EntitySchema.[8] Schema numbers in Wikidata are prefixed with the letter E and encoded using ShExC.

SIMPLE DATA SERIALIZATIONS

The ShEx compact syntax provides one example of a relatively easy-to-read validation language that is

```
PREFIX ex: <http://example.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX schema: <https://schema.org/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX dce: <http://purl.org/dc/elements/1.1/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dcmitype: <http://purl.org/dc/dcmitype/>

ex:a _ photo _ shape {
  rdf:type [ schema:MediaObject ] {1} ;
  schema:contentUrl IRI + ;
  dct:title rdf:langString {1} ;
  dce:creator rdf:langString * ;
  dct:spatial IRI * ;
  dct:date xsd:date ? ;
  dct:type [ dcmitype:~ ] +
}
```

Figure 5.1
Requirements for RDF description of photographic resources expressed in ShExC syntax

machine-processable. While ShExC is used for the specific purpose of defining requirements for RDF data, it may be possible to utilize other general-purpose data formats for MAPs that offer the same combination of machine and human readability. The YAML data serialization language is one such format. This language has been designed for use in a wide variety of software implementations and makes use of an extremely simple syntax that preserves readability for humans.

Yet Another Metadata Application Profile (YAMA) markup language provides an interesting example of YAML syntax implementation. YAMA builds on the YAML syntax specification by creating a structure for recording the components of a MAP, specifying an extensible set of key/value pairs for defining entities, properties, and value constraints.[9] YAMA MAPs are well-suited to machine processing for output of validation code and other derivatives. Additionally, because they use the simple YAML syntax and natural-language key names for recording MAP content, they can be read by humans as well.

Tools for Conversion

For implementers with sufficient technical know-how, it may be desirable to record MAP information only in a machine-readable presentation and use this version alone for reference. For many metadata creators and managers, however, this is not a realistic solution. Even when an encoded format is required to interact with data-entry forms, validate created metadata, or meet other needs, many of us will want a human-readable presentation as reference for ourselves and catalogers and metadata specialists who describe collections, and to make available to others who may wish to reuse our data. In these cases, the availability of

tools to generate machine-readable versions of MAPs from human-readable documents, or vice versa, will be important. Without such tools, we face the challenge of maintaining two distinct versions of the same MAP.

For MAPs that originate in machine-encoded form, various methods exist for generating human-readable documentation, including the use of templating engines written in JavaScript, Python, or other programming languages, and the use of XML stylesheets for MAPs encoded using XML or RDF/XML syntax. These can be used to output the information in an encoded MAP as HTML for viewing in a web browser or in PDF or another document format.

By supplementing or replacing machine-encoded notation with text meant for humans, the transformation process can provide MAP versions that are better suited for use by catalogers and other metadata specialists. Additional text can include natural-language descriptions of MAP components to replace encoding, which is often truncated for efficiency or embedded in syntax that is necessary for interpretation by software but hinders readability. The transformation process can also provide additional text meant for humans, such as instructions for recording values and properly formed examples.

Human-readable MAPs can provide a more detailed description of entities, properties, and value constraints than encoded versions, but these documents usually lack the structure needed for efficient processing using software tools, making the conversion of human- to machine-readable MAPs a more challenging task than moving in the other direction.

Providing human-readable MAPs in a more structured format may make the task of converting them to encoded MAPs easier. At the time of writing, the

DCMI Application Profiles Interest Group is developing a vocabulary for expressing human-readable MAPs in a standardized tabular form, and a Python software package is being developed in support of this work.[10] Two primary use cases for this software are validating the structure of MAPs written using the vocabulary and generating ShEx schemas based on these. These projects may allow a larger group of users to take advantage of data-validation capabilities by creating structured tabular MAPs and converting these to ShEx encoding.

Notes

1. "DLF AIG Metadata Application Profile Clearinghouse Project," Digital Library Federation Assessment Interest Group Metadata Working Group, https://dlfmetadataassessment.github.io/MetadataSpecsClearinghouse/.
2. "PCC Statistics," Program for Cooperative Cataloging, Library of Congress, https://www.loc.gov/aba/pcc/stats.html.
3. "Wikidata:WikiProject PCC Wikidata Pilot/Participants," Wikidata, last updated May 1, 2021, https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Participants.
4. "Wikidata:WikiProject Books," Wikidata, last updated October 6, 2020, https://www.wikidata.org/wiki/Wikidata:WikiProject_Books.
5. "BIBFRAME Model, Vocabulary, Guidelines, Examples, Analyses," Library of Congress, https://www.loc.gov/bibframe/docs/index.html.
6. "BIBFRAME Profile Editor," Library of Congress, http://bibframe.org/profile-edit.
7. "Sinopia Vocabulary," Linked Data Editor, Linked Data for Production 2 (LD4P2), https://sinopia.io/vocabulary.
8. "Shape Expressions arrive on Wikidata on May 28th," Wikidata, https://www.wikidata.org/wiki/Wikidata:Project_chat/Archive/2019/05#Shape_Expressions_arrive_on_Wikidata_on_May_28th.
9. Nishad Thalhath, Mitsuharu Nagamori, Tetsuo Sakaguchi, and Shigeo Sugimoto, "Yet Another Metadata Application Profile (YAMA): Authoring, Versioning and Publishing of Application Profiles," *DC-2019—The Seoul, South Korea, Proceedings*, DCMI International Conference on Dublin Core and Metadata Applications, 114–25, https://dcpapers.dublincore.org/pubs/article/view/4253.
10. "dcmi/dctap," Dublin Core Metadata Initiative, GitHub, https://github.com/dcmi/dctap; Tom Baker, "Tap2shex," GitHub, 2021, https://github.com/tombaker/tap2shex.