

# PCC Task Group on Metadata Application Profiles

## Report to PoCo

October 15, 2019

### Introduction

The Task Group on Metadata Application Profiles (MAPs) was formed in May 2019 by the PCC Policy Committee (PoCo) to help the PCC understand the issues and practices associated with the creation and management of metadata application profiles within the PCC context. The work includes defining MAPs, their purposes, intended audiences, and the cardinal elements and constraints values of MAPs. In addition, the Task Group explores ways that machine-actionable MAPs can be supported so that the data are more interoperable and reusable in the semantic web environment. During the last four months, the Task Group has been conducting bi-weekly virtual meetings to explore and review currently available MAPs for an environmental scan to understand how MAPs were created and used so far. This report outlines the work done so far and next steps the Task Group will undertake in the coming months.

### Progress

The Task Group is approaching the work in two ways thus far; an environmental scan and gathering use cases.

#### Environmental Scan

As a first step, the Task Group started conducting an environmental scan on the application profiles. The goals of the work are learning about the problem space for data targets, profile designs, storage issues, and sharing mechanisms. Gleaning from various application profiles, the Task Group has learned each MAP's stakeholders, design and decision making process, and the management of them (i.e., governance and maintenance models). The actual review was conducted in two ways to better understand MAPs: first, review available document(s), and second, invite guests to learn about the MAPs directly from them. The Task Group developed a list of criteria to conduct a consistent review as included in Appendix 1.

Several of the guests are members of the task group.

- Dublin Core Application Profile (K. Coyle)

In keeping with the Dublin Core mission, the Dublin Core Application Profile work defines a general philosophy for simple application profiles. A current effort is tasked with developing a vocabulary and templates for the creation of application profiles. It is worth noting that the BIBFRAME profile work initiated by Zepheira was based on the earlier work on profiles by DCMI.

- BIBFRAME (J. Williamschen)

The Library of Congress created their BIBFRAME editor and application profiles to facilitate a large-scale cataloging pilot at the Library. These tools are available to all, but they were created for their primary audience -- Library of Congress catalogers. The profiles use some RDA terminology and terms but they are filtered through the BIBFRAME ontology. In addition, some term lists in the BIBFRAME editor incorporate vocabulary terms that are part of the MARC standard. At this time, the application profiles include little validation and documentation is limited, but these areas will receive more attention in the future.

- LD4 ARM (art & rare monographs) (S. Folsom)

The Art and Rare Materials BIBFRAME Ontology extension and related application profiles began as an LD4L Labs work area, and is now a partnership between ARLIS/NA and RBMS. The LD4L Labs' SHACL-based profiles explored patterns for defining display shapes built off from validation shapes; this work was conducted for aspects of the [rare monograph application profile](#). More on the validation shapes is available in the [SHACL maintenance document](#) in the ARM GitHub.

- Digital Public Library of America (DPLA) (M. Della Bitta)

The Digital Public Library of America (DPLA) is designed to build on the Europeana Model and references 14 different namespaces. As its name implies, DPLA is for a general public audience and its primarily cultural heritage content is added by volunteers from DPLA-contributing institutions. DPLA is constructed based on [Conway's Law](#): the idea that systems reflect the design of the people who build them. All DPLA MAPs are human readable PDFs, with no machine readable MAPs available at this time. DPLA metadata itself is expressed in JSON-LD, per [ElasticSearch](#).

- RDA (G. Dunsire)

The RDA Steering Group (RSC) will be forming a task group for RDA application profiles in fall 2019. The issues it will consider include profile architectures, guidance and documentation, and encoding formats. It will follow the Singapore Framework for Dublin Core Application Profiles, with the important difference that it will select only RDA elements. There is no predetermined relationship to the LC-PCC policy task groups

because the latter have to complete their work much sooner. Using RDA as a content standard with BIBFRAME may involve mapping of RDA to BIBFRAME once the latter is stable, or BIBFRAME elements can be specified in an application profile.

- EURIG (R. Behrens)

The EURIG (European RDA Interest Group) RDA Application Profile is the first draft of a new AP for use with post 3R RDA. It is based on the Standardelemente-Set (Application Profile) developed by the German speaking RDA community (DACH), at the time of RDA adoption; the predecessor AP is in PDF form, and is used by the DACH community in conjunction with the current RDA Toolkit and the DACH-AWR (policy statements). EURIG began work on the ERAP in May 2018, and was tasked by the RDA Steering Committee (RSC) with developing template application profiles that could be recommended by the RSC and adapted by other RDA communities for post 3R project use and to foster data exchange among the communities. It was submitted to the RSC in April 2019. The EURIG community believes the finished AP should be freely available, and embedded in the Toolkit. Additional APs for specialized material like music and religious works are to be developed and will build upon the general AP. The working draft is an Excel spreadsheet with the APs each on their own sheet, but they anticipate it being deployed as an online tool of some kind. At the current time the AP is limited to RDA elements and is neutral with regard to data encoding format; there are no current plans for automated validation tools. EURIG hopes to see it widely adopted internationally.

- BIBFRAME Lite + (G. Gonzalez)

BIBFRAME Lite is a suite of ontologies developed by Zepheira in 2014 for its Library Hub (later known as Library.Link Network) partners. There are over 1,400 libraries that currently use BIBFRAME lite ontologies in production. BIBFRAME lite, derived from BIBFRAME v. 1 with extension to specific material types, e.g. books, continuing resources, rare, archival materials, serves users representing over 3,000 library locations in 8 countries. These libraries created over 5 billion links between BIBFRAME Lite resources.

- Sinopia (N. Lorimer)

Sinopia is a cloud-based environment for the creation, editing, and storage of linked data, developed at Stanford University as part of the Linked Data for Production--Pathway to Production (LD4P2) Mellon-funded project. Sinopia includes a profile editor, based on the one developed at the Library of Congress, with which trained catalogers/metadata librarians can create templates for linked data creation; a linked data editor for metadata creation and editing; a connection to the lookup service [Questioning Authority](#), developed at Cornell University; a local database; and, a search

function that currently points to Sinopia, Discogs, and individual holdings of LD4P institutions in SHARE-VDE. Sinopia is not an application profile per se, but the templates serve that function by introducing restrictions in ontology, vocabulary use, cardinality, and whether a property is mandatory. While any ontology may be used in Sinopia, menus for BIBFRAME, RDF, RDFS, BFLC, and MADS-RDF are available within the profile editor. Current experimental profiles/templates include BIBFRAME, RDA, and the various BIBFRAME extensions (BFLC, Art and Rare Materials, Performed Music Ontology).

- Europeana (V. Charles)

The Task Group will set up a meeting with the Europeana to discuss its APs in the near future.

A list of MAPs reviewed and detailed characteristics of each MAP can be found through the link in Appendix 2.

## Tools and Use Cases

The Task Group also started gathering a list of tools used for creating the MAPs and use cases when users want to create a new MAP to better understand how and when MAPs can be created. The work is in progress and the results will be included in the final report to PCC.

## Findings

From the environmental scan, the Task Group has found the following:

1. Most MAPs are developed by their respective user groups whether that is for a specific project or a specific collection. The parameters and rules regarding the use of terms and vocabulary encoding schemes are community specific.
2. The MAPs reflect the use cases of the relevant communities. This applies not only to the collections represented but also other considerations such as audiences served and workflows. For example, aspects of the Europeana and DPLA MAPs reflect their role as hubs aggregating data from a diverse range of contributors.
3. Most MAPs provide general guidelines for how a vocabulary (or vocabularies) should be used.
4. An important issue to understand is how MAPs can be reused, extended, or refined.
5. Some profiles defined cardinality elements and constraint values at a high level, while others were a bit more restrictive. The granularity of the MAPs depends also on the needs of the user group.

6. In some cases, some would want deeper investigation of the possibility of implementing cascading decision processes for preference and options (If (if not)/elseif/else, then what follows).
7. Most profiles we encountered are human readable. Only one profile has deployed Shapes Constraint Language (SHACL) to validate input data, and it would require conversion to another format or loading it into an RDF editor in order to "read" it. However, most MAPs expressed interest in moving toward a machine-actionable function. Invited guests and colleagues said that there would be benefits of machine-parseable APs since these can be more explicit, leaving little for interpretation. Hence the output are more interoperable.
8. In terms of data quality assurance, the order of elements in some APs are built into in its editor.

It is noteworthy to add that the term 'application profile' is used interchangeably with template.

## Next steps

In the months to come, the task group will be working on formulating criteria and best practices for the PCC community to create application profiles when it is needed, based on the data the Task Group has gathered so far. In order to do that, the Task Group will ask the following questions when working with the data and developing the criteria and best practices.

- How to identify primary and secondary audiences, catalogers or programmers? Learn how a multipart specification of a MAP, human-readable vs machine-actionable will impact user's needs and web form configurations.
- How to make a process support an output that meets PCC needs (human readability, validation, machine readability)?
- What are the available tools and technologies to create machine actionable MAPs, and from these MAPs what are plausible workflows for generating human readable documentation?
- What does validation mean and what to validate on (cardinality, ranges of properties in a given context, etc.)?
- Should our MAPs aspire to define a floor description with the possibility to extend (OPEN), or define an exact limit about what can be said in what we consider PCC data (CLOSE)?
- How can Application Profile-based validation provide meaningful feedback to a user editing a set of metadata statements?
- How are existing constraint languages valuable to implementers, particularly if the tools we are building cannot interpret or act on them natively?
- How to support sustainability of creating and governing PCC MAPs?

Lastly, the Task Group will try to identify skill(s) required for creating MAPs and how one can gain such skills.

## Appendix 1

The Task Group interviewed invited guests to discuss the following questions:

- What are your group's objectives and deliverable?
- Intended audience/practitioner
- Purpose of the AP?
- Were you specifically tasked to create an RDA application profile, or are other ontologies also in scope?
- Do you just use one vocabulary, e.g. RDA, why just one and this one, instead of another vocabulary?
- Do you envisage using extensions to RDA? How do you plan to deal with the areas (e.g. "category of work") where RDA makes reference to use of external or "local" vocabularies?
- System implication
- Relationship to other APs?
- What workflows do you imagine to be in place?
- What format do you suspect your AP will use? If the community chooses a not supported format, will you be able to link out to external profiles?
- Do you intend to validate your AP? What tools will you need?
- Will there be documentation of your AP? If so, will it be accessible by anyone?
- Will your AP be an RDF-based schema?
- Will your AP be exportable? In what data formats?
- Since we will remain in a mix MARC/LD environment for some time, will you consider adjusting the AP to accommodate MARC cataloging?

## Appendix 2

[Environmental scan spreadsheet](#)

### Task Group Members

Karen Coyle (W3C, Liaison to the DCMI profiles group)

Corine Deliot (Consultant, British Library)

Nancy Fallgren (Consultant, National Library of Medicine)

Steven Folsom (Cornell)

MJ Han (Co-chair, University of Illinois Urbana-Champaign)

Les Hawkins (Library of Congress)  
Nancy Lorimer (Stanford)  
Lucas Mak (Michigan State)  
Honor Moody (Harvard)  
Chiat Naun Chew (Harvard)  
Heather Pretty (Memorial University of Newfoundland, PoCo liaison)  
Jackie Shieh (Co-chair, Smithsonian)  
Cynthia Whitacre (OCLC)  
Jodi Williamschen (Library of Congress)