

PCC-Task Group on Metadata Application Profiles

Final Report

Submitted to PCC Policy Committee
May 1, 2020

[Executive Summary](#)

[Introduction](#)

[Application Profiles in PCC Context](#)

1. [Define MAPs in the PCC Context](#)
 - a. [Use cases and Requirements](#)
 - b. [MAPs that Serve Both MARC and Linked Data](#)
 - c. [Shareable MAPs in the PCC Context](#)
2. [Base Model Assumptions](#)
 - a. [Data Standard - BIBFRAME and Extensions](#)
 - b. [RDA as Content Standard](#)
 - c. [Open MAP Model](#)

[Stakeholders in MAP Development](#)

1. [Library System Vendors](#)
2. [Data Remediation Vendors](#)
3. [OCLC](#)
4. [Other Profile Creators](#)
5. [Consumers of the Library Data](#)

[Developing and Consolidating Expertise](#)

1. [Experiences and Skills Needed](#)
2. [Different Skills for Different Tasks](#)
3. [Identifying Expertise Inside and Outside of the Community](#)
4. [Available Resources](#)

[Maintenance/Governance Models](#)

[Prior Art, Linked Data and MAPs](#)

[Conclusion and Consideration](#)

[Acknowledgements](#)

[Task Group Members](#)

Executive Summary

From May 2019 to April 2020, the Task Group on Metadata Application Profiles (MAPs) worked to identify and understand the issues and practices associated with the creation and management of MAPs within the PCC context. This final report seeks to define MAPs in the context of PCC, including use cases and requirements, base-model assumptions, and maintenance/governance models. It also recommends actions critical to the successful development, implementation, and maintenance of PCC MAPs.

Linked data is an evolving area for the library community. The PCC strategy should be based on continuing assessment of use cases, clear articulation of goals and assumptions, a robust governance framework, ongoing cultivation of skills and experience, and collaboration with key partners. The Task Group makes the following recommendations:

1. PCC context MAPs should be expected to be in a human readable format for use by metadata professionals and semantically adaptable for machine processing by programmers to meet RDF semantic validation requirements.
2. PCC MAPs should support both MARC and linked data workflows as part of the library community's transition strategy from MARC to linked data.
3. PCC MAPs should be open to support element extensions with other ontologies as needed.
4. PCC should form a task group to focus on the creation, maintenance, and management of MAPs, and to cultivate experience in those areas. The task group will closely work with the PCC standing committees and other relevant bodies.
5. PCC should develop a set of competencies for librarians to work with linked data and MAPs. To do this, it will need to identify expertise both within and outside the PCC community.
6. PCC cannot achieve its linked data objectives alone. PCC should establish and maintain open communication channels with partner communities to advance shared objectives. Stakeholders include LC, OCLC, other profile creators, library systems vendors, data remediation vendors, and consumers of library linked data.

Introduction

A primary role of the Program for Cooperative Cataloging (PCC) is to work with the PCC membership to define shared metadata standards in support of metadata interoperability and rich descriptions of library resources. The PCC currently provides human readable metadata application profiles (MAPs) via the [BIBCO \(Bibliographic Cooperative Program\) Standard Record RDA Metadata Application Profile](#) and the [CONSER \(Cooperative Online Serials\) Standard Record RDA Metadata Application Profile](#) to guide cataloger practice within the scoped semantics and workflows of MACHine-Readable Cataloging (MARC) and Resource Description and Access (RDA). As the PCC community transitions toward implementing linked data for resource description, the extreme flexibility of the Resource Description Framework (RDF) requires development of machine-actionable MAPs in order to state the community's desired semantics explicitly for tool developers and data consumers. The PCC will also need to develop human-readable metadata application profiles for RDF descriptions, but these are more easily generated from machine-actionable MAPs.

The Task Group on Metadata Application Profiles 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. According to the PCC Metadata Application Profile Task Group charge:

“A metadata application profile (MAP) is a set of recorded decisions about a shared data target for a given community. MAPs declare what models are employed (what types of entities will be described and how they relate to each other), what controlled vocabularies are used, the cardinality of fields/properties (what fields are required and which fields have a cap on the number of times they can be used), data types for string values, and guiding text/scope notes for consistent use of fields/properties. A MAP may be a multipart specification, with human-readable and machine-readable aspects, sometimes in a single file, sometimes in multiple files (e.g., a human-readable file that may include input rules, a machine-readable vocabulary, and a validation schema).”¹

PCC MAPs will support shared data practices to ensure interoperability in a broader semantic web environment as they will provide a definition of a “floor-level” description with primary rules for each property used to describe a particular entity in the machine-readable cataloging description standard, especially since data created by cooperative cataloging program are meant to be reused by anyone with minimal modification in support of cooperative cataloging. With community defined MAPs in place, the PCC can begin to establish community created and maintained data that is easily consumable in individual library's local environments.

This report attempts to define MAPs in PCC context based on use cases and base-model assumptions, and recommends actions critical to successful development, implementation, and maintenance of PCC

¹ <https://www.loc.gov/aba/pcc/taskgroup/Metadata-Application-Profiles.html>

MAPs. The Task Group is aware that the field is in flux and that the PCC at this moment is without community standards for MAPs. This report is written with the assumption that those standards will be developed based on community needs.

Application Profiles in PCC Context

Define MAPs in the PCC context

By majority, PCC is made up of research libraries. The profiles created by the PCC will undoubtedly focus on the needs of the PCC libraries, and yet some decisions and practices made in this context may benefit or benefit from wider MAP activities.

Use Cases and Requirements

As alluded to in the definition in the introduction, there are many uses and audiences for MAPs. The following (select) use cases and functional requirements are roughly in the order that the Task Group would prioritize based on current implementation needs and would very likely shape the approach PCC takes for creating and maintaining MAPs.

- **Human Readable:** Provide human-readable guidance to implementers, including catalogers, tool developers, and trainers.²
 - Specify the entities to be described and the properties/elements identified for use:
 - What must be said about them (required data elements)
 - What may be said about them (optional data elements)
 - Specific instructions on the encoding of data elements
 - Define “a floor” baseline description that may be further extended locally or by specific communities/domains
- **Further specify or constrain modeling for entities or entity types:** Provide further specifications or constraints for modeling complex entities or entity types.
 - Give direction when there is more than one way to do something (e.g., the modeling of aggregates³ can often be done in multiple ways even within one model or cataloging form configuration).
 - Highlight type-specific elements (e.g., using a duration property with time-based content and a different extent pattern for text-based content).
- **Cataloging workflow configuration:** A machine parsable output serving as the basis for cataloging workflow configurations.
 - All RDF tooling, whether ontology agnostic or specific to a particular model, requires that the output be explicitly defined. (Note: there isn’t a standard form configuration

² Intended audience and likely partnerships for PCC MAPs are expanded upon below in the stakeholder section. Please see the Section “Relationships with Stakeholders.”

³ [JFLA LRM](#), p. 93.

specification; machine-actionable MAPs will likely require conversion for specific applications).

- If workflows are not explicit and repurposable, there is a greater chance the output will not conform to expectations.
- **Semantic validation:** A machine-parsable schema serving as the basis for data/semantic validation.
 - If a machine-readable MAP is not created natively, such as in [Shape Expressions \(ShEx\)](#) or [Shapes Constraint Language \(SHACL\)](#), it should be possible for the MAP to be converted to either these RDF validation standards, or or any equivalent validation standard.
- **Versioning:** MAPs will inevitably change, making versioning important.
 - Versioning will aid in allowing implementers to opt into specific versions, and cross-walk from one version to another.
 - The frequency with which new versions are created will likely be determined by staffing models, the amount of resources allocated to maintain these profiles, the urgency and number of changes needed, and the need to address any changes made to related standards/models that the PCC MAPs build on. While a regular frequency would signal to the community that the PCC MAP is an actively maintained standard with clear schedules for engagement, this would need to be balanced with urgency of the need for specific changes and the needs of communities extending the PCC MAPs.
- **Aid the long transition from MARC to linked data**
 - Members of a community often use a variety of encoding standards, and there is a need to specify metadata requirements in a schema-agnostic way in order to promote convergence of practice, and with it compatibility and reusability of data. A special case of this is communities like the library cataloguing community which is in a long-term transition from MARC to BIBFRAME. See section below titled “Account for the need to address both MARC and linked data workflows”.

MAPs that Should Serve Both MARC and Linked Data Workflows

Libraries are entering an extended transitional period where some institutions begin to adopt BIBFRAME while others remain with MARC. While full round-tripping of BIBFRAME and MARC data is neither feasible nor desirable, the economies of shared library metadata production require that data produced in BIBFRAME can be converted for use in MARC systems, and vice-versa. Conversion and reuse, including downstream consumption of data, is most easily achieved if there is commonality between BIBFRAME and MARC practice wherever possible in areas such as selection of metadata elements, choice of ontology, syntax, and cardinality. These common requirements can be captured in application profiles.

- **Issue 1: Drawn out and uneven movement toward linked data**

There will remain a need for support for MARC records for the foreseeable future since there will be some libraries that may not change to direct creation of linked data. There are many possible reasons for this: lack of usable and affordable infrastructure, lack of documentation, lack of training, dependence

on library systems that do not support linked data, skepticism toward linked data, need to concentrate on other training (e.g., RDA) and dependence on OCLC, Inc. (OCLC) for most of their catalog records. As the broader library community transitions to full linked data descriptions, even libraries that implement linked data descriptions will be making use of MARC data that was originally created in a BIBFRAME-based ontology and converted to MARC, for example by the Library of Congress (LC) converter or by OCLC. Many of these converted records will be coded as PCC records. Because of this, there needs to be compatibility in standards between MARC and BIBFRAME so the PCC community does not have two definitions of the same PCC standard.

There will be a corresponding need on the part of libraries who have moved to BIBFRAME to be able to make use of data converted from existing MARC records (whether on the fly or through a batch process) in their copy cataloging workflows. Additionally, it would also be highly desirable for updates made to BIBFRAME data to be propagated to the corresponding MARC records, and vice-versa. Again, this suggests that a compatibility in standards that takes this need for conversion into account is necessary. Because current converters tend to expect certain input, both metadata and modeling of that metadata, MAPs really are necessary because they contain the information for conversion.

- **Issue 2: Library systems' lag in technology advances and the greater systems environment**

Very few library systems provide access to linked data creation. MARC is still very much part of the basic structure of these systems, with orders, payments, holding and circulation data, and authorities all attached to a MARC record. The dependence of most libraries on a commercial system for any metadata creation or import suggests that most will not be moving quickly to direct creation of linked data and may not have the means to do so for some time — libraries do not change library systems often or easily.

A few systems do provide conversion to linked data from MARC records. This brings in another possible need—to codify conversion of MARC records to linked data so that they are easily compatible with other conversions and newly created linked data.

For those libraries who are moving to linked data the question remains of what to do with data currently included with MARC records in library systems. For most of those libraries, the emphasis is currently the direct creation of bibliographic records. The larger infrastructure of a library system — acquisitions, circulation, holdings, financial data, reports — has not been explored yet. While this is out of scope for the PCC, it needs to be recognized that even libraries moving toward linked data may still be dependent on an “operational” MARC record to fulfill and link these tasks if their system requires it. This is not the case with all systems, as some have an “inventory” or otherwise more neutral database as their base.

As an example, Stanford is moving toward linked data in stages. Their current integrated library system (ILS) uses the MARC record as its link to all the different ILS modules and is necessary for report generation. Copy cataloging is done using MARC records. These records are stored in the local system and are converted to linked data by an outside vendor. In the near term, original cataloging will take

place outside the ILS in Sinopia⁴; the cataloging data will be then converted to MARC and linked to the ILS through the converted MARC records. A discovery layer will still use the MARC data until a new phase of the linked data transition is implemented.

Shareable MAPs in the PCC Context

Consideration will need to be given to how reusable PCC MAPs will be in different contexts. This means asking the questions - What is the appropriate level of agreement among PCC Libraries? What aspects of the created data should the PCC remain flexible on? The following table (starting from the bottom) illustrates how basic decisions about the shared data practices are fundamental and therefore are the areas where the PCC should focus its attention. With these decisions in place, related communities and institutions can enhance floor descriptions with profile extensions. From there the PCC can support and optimize specific workflows, and configure applications accordingly. Note that the boundaries between the different levels are less distinct than the table might suggest.

Table 1: Levels of Agreement in Metadata Application Profiles

Application	The syntax the profile is captured in	Likely to change depending on the app.
	Decisions around blank nodes	Escalates into a community discussion quickly
	Which look up service are we using?	Usually up to developers, but the decision needs to be recorded
	Labels and order of fields	May need greater consideration at the Community level
Workflow	Default values for the institution	True across the board for institution
	Default values for specific workflows	True for specific workflows and/or content types
	Levels of description	Think MARC encoding levels
	Administrative metadata	Deciding how to track data flows
Institution	Preferred vocabularies	LC Authorities only? FAST, Wikidata, ISNI, RDA vocabularies, local
	How we say something	E.g. <code>bf:originDate</code> vs. Contributions with dates, both?
	Caps on certain assertions	Upper limits on the quantity assertions using a particular property
	Decisions on community optionals	Shoring up the floor
Data	What are the things we're describing?	Community decided and maintained
	What can/should we say about things?	Often considered "the floor"
	Cardinality, repeatability, datatypes	Usually pretty loose to account for practicalities
	How tied to AAPs and RDA are we?	A consideration for BF-> MARC conversions

While the PCC will want machine-actionable MAPs for the reasons stated in the Use Cases and Requirement section above, as a community the PCC has not, and may never, reach consensus on *the single best* data format for expressing MAPs. Rather than allowing this to paralyze the progress, the PCC should balance desires for specific outputs with existing skills in the PCC community and its highest prioritized use cases. (Skills required to maintain MAPs are addressed in more detail below.) To facilitate maintenance of MAPs, PCC may commit to spreadsheets as the "source of truth" and convert to other formats like [JavaScript Object Notation \(JSON\)](#), SHACL, and/or ShEx. This would not only allow for

⁴ Sinopia linked data creation environment: <https://sinopia.io/>

catalogers to make changes to the spreadsheet without having to understand other serializations⁵, spreadsheets serve as a relatively human-readable document. If starting with spreadsheets proves unfeasible, the PCC could start from other more expressive formats and convert to spreadsheets or html for human readability, but this would mean fewer people could interact directly with the shared profiles until user friendly tooling for authoring these other formats are available. Through experience the PCC will better understand challenges converting between different serializations of MAPs, and over time standards will emerge, allowing to iterate with subsequent changes.

As has already been seen in the early experiments with profiles in the Linked Data for Production: Pathway to Implementation ([LD4P2](#)) context, organizing core parts of PCC MAPs in a modular way is going to be really important for sustainability and maintenance. For example, if the PCC decides that labels for entities across its descriptions are going to use a single property (like `rdfs:label`), it may want to state that only once within a MAP so that it can be easily implemented (or changed as the case may be). A more obvious example might be to have one profile for geographic entities that can be reused for both locations as subjects and/or publication information.

Base Model Assumptions

In a linked data context, PCC data will be stored in one or more RDF serialization(s), using a BIBFRAME-based data model and [RDA](#) as the content standard.⁶ This necessitates a two-pronged approach to metadata application profile base models for the PCC community as shown in table 2, an approach that will require some mediation between conflicting requirements of the data models and the content standard.

MAPs for library linked data are new to PCC and most librarians. The Task Group’s recommendations are subject to change as *OUR* community gains more experience, both theoretical and hands-on.

Table 2: Recommendations for PCC context MAPs Base model

Data Standard	BIBFRAME with RDF, RDFS, OWL, BFLC as required
Extensions	<ul style="list-style-type: none"> ● BIBFRAME-based library-developed extensions for specific domains/communities (e.g., Art & Rare Materials (ARM), Performed Music Ontology (PMO)) ● BIBFRAME-based extensions for use with service providers (e.g., OCLC, Share-VDE, GoogleBooks) ● Integration of post-3R Project RDA relationship elements ● Possible additions from schema.org for web

⁵ "In computing, serialization...is the process of translating data structures or object state into a format that can be stored...or transmitted...and reconstructed later." Wikipedia, <https://en.wikipedia.org/wiki/Serialization>. There are several formats for RDF serialization, which is also sometimes called "notation."

⁶ Use of the term ontology versus vocabularies often time is intermingled. In this document, the term "RDF vocabularies" is chosen to express the emphasis of the language Resource Description Framework for a library linked data environment. For more information, refer to: <https://www.w3.org/standards/semanticweb/ontology> and <https://dcmi.github.io/ldci/D2695955/#a-rdf-vocabularies-and-application-profiles>.

Value Vocabularies	<ul style="list-style-type: none"> ● id.loc.gov vocabularies ● RDA vocabularies ● External value vocabularies endorsed for PCC use⁷
Content Standard	RDA (instructions for determining instance value data, e.g., where to take a title from)

Data Standard Assumptions- BIBFRAME & Extensions

BIBFRAME is a library-centered RDF-based ontology developed by the Library of Congress with input from a number of libraries in North America and Europe. Libraries worldwide have been actively experimenting with the BIBFRAME data model and with projects such as Zepheira’s work with the archival community. These projects have served to further BIBFRAME development and adoption, even beyond the library community.⁸ Through the Linked Data for Production ([LD4P](#)) effort, BIBFRAME is being used by more than 20 libraries in North America, most of which are PCC member libraries.⁹ In Europe, implementations have taken place or are in process at the National Library of Sweden, the National Library of Hungary, the National Library of Norway, and the National Library of Finland. All these libraries have already converted their bibliographic data to some flavor of BIBFRAME. Along with the converted and newly-created data from LC and LD4P, there is now a large pool of library linked data in BIBFRAME. PCC, as an organization with members participating in LD4P, and with its close relationship with LC, has already invested time and expertise in BIBFRAME. The formation of a PCC pool of BIBFRAME data, built on a robust application profile, will only increase the availability of high quality, reusable library linked data.

The use of one ontology as a base model, with any necessary extensions for other vocabularies such as RDA, will also ease conversion to MARC – which is likely necessary for the foreseeable future. In addition to the PCC base model MAPs, PCC will also have to address the need for more specialized extensions for specific bibliographic domains (e.g., rare materials, music, art, etc.). The PCC should work with those groups to ensure compatibility with the base model and support data created through these extensions. Vendors providing various services based on linked data will also find it necessary to create their own extensions to meet their requirements and to provide their services to users. For instance, Share-VDE is developing the “opus”, a type of bf:Work that gathers other bf:Works under a single node. Similar to this is LC’s experimentation with “hubs”, the data from which they intend to make available for download in the coming year. Also, OCLC recently received a grant to develop an entity management system for libraries through a local Wikibase. It is likely that they will also need to develop some extensions to BIBFRAME. PCC MAPs will need to take all of these into consideration as well as the base model and domain extensions.

⁷ Some examples of vocabularies that may be evaluated for PCC use are listed in [Formulating and obtaining URIs: a guide to commonly used vocabularies and sources](#), issued by the PCC URIs in MARC Task Group.

⁸ Zepheira has modeled different types of materials, such as archival resources, using BIBFRAME version 1 on <http://bibfra.me>.

⁹ The BIBFRAME implementation of the LD4P participating libraries is in conjunction with the Casalini and @Cult’s project, [Share-VDE](#) linked data platform.

Content Standard Assumptions - RDA

Since 2015, PCC libraries have followed the BIBCO Standard Record (BSR) and CONSER Standard Record (CSR) RDA Metadata Application Profiles in describing library collections in the MARC format. In the [PCC Position Statement on RDA](#) in August 2019, the Policy Committee affirmed the use of RDA as a loose content standard and a starting point for developing standards in a linked data context as well. It does, however, state that RDA will likely need to co-exist with non-RDA practices in a linked data environment, and that guidance is necessary for when the RDA and BIBFRAME models conflict. Retaining the BSR and CSR, and expanding their contexts to include linked data, will assist in developing linked data best practices and compatible data models, while retaining a strong connection to the modeling of MARC records. RDA also offers a much-needed enrichment of BIBFRAME ontologies, particularly for work-to-work relationships and modeling of aggregates and diachronic works.

MAP Model Assumptions - Open MAP Format

A MAP may be extensible, and thus “open”, or it may be a fixed standard, and thus “closed”. Because of the above recommendations, the Task Group recommends an open MAP format. The MAP would consist of a base standard defined in such a way that accommodates element extensions with other ontologies as the needs arise. A closed MAP where cardinality and constraints are applied within a prescribed ontology with little or no extensibility is contrary to the PCC’s ultimate mission. PCC as an organization strives to be one community where standards and best practices are shared.

Stakeholders in MAP development

In addition to LC, OCLC will be an important partner for PCC, especially PCC participation in the [OCLC Entity Management Infrastructure Mellon Grant](#). Other important stakeholders in the development of PCC MAPs are library system vendors, cataloging and data remediation services, non-PCC profile developers, and other consumers of library linked data.

Library System Vendors

As providers of a significant library technology infrastructure, library system vendors must have a strong understanding of the structure and cardinality of the metadata they are handling. PCC MAPs will inform vendors in these areas and enable them to develop functionalities facilitating various library workflows. From the perspective of library staff, the integrity, consistency, and predictability of data from heterogeneous sources impact their daily workflow. It is desirable that library system vendors develop features including:

- The ability to import and export machine readable MAPs in standard formats and/or the ability to create and edit machine readable MAPs locally, in support of the reuse and extensibility of existing MAPs to provide flexibility at the consortium and individual institution levels.

- A set of constraints, analogous to minimum MARC field requirements in current systems, to ensure data exchange and discovery
- A data validation function capable of evaluating imported and exported data for compliance with *expected* Galleries, Libraries, Archives, and Museums (GLAM) metadata standards and of identifying what data remediation is necessary for compliance with those standards.

While library staff should not expect that data in any library system would be homogenous, the deviation should be within a set boundary. The data validation function available through a MAP ensures the data stored and available inside the system are compliant to future PCC MAPs at the minimum, with the ability to accommodate possible extensions used by the GLAM community. Contrary to the consistency and integrity on the network level, there will be needs for flexibility and modification on both consortium as well as individual institutional levels. PCC MAPs will serve as one of the many guiding documents for vendors in addressing data integrity, flexibility, and extensibility issues when developing their products.

Data Remediation Vendors

Cataloging and data remediation vendors will need to make use of PCC MAPs. They will also need to apply custom extensions to account for local practice for cataloging form creation and data validation. If working with data from a variety of library system vendors, they will need conversion tools for MAPs created in different systems. The custom nature of much of this work means that these vendors will be important partners in developing flexible MAPs that nevertheless produce PCC-compliant data outputs.

The Library of Congress

The Library of Congress, as the originator and maintenance agency of various cataloging-related standards and policies, will be influential in shaping PCC MAPs. Since the incubation of BIBFRAME, LC has been the driving force behind its conceptual modeling. Any changes to the data modeling will have overarching ramifications to the design of any metadata application profile and software applications that use BIBFRAME data (e.g., Sinopia). In recent years, LC has converted most of its controlled vocabularies into an open linked data format, and these vocabularies are accessible at <http://id.loc.gov>. Given that a large proportion of them have been used by libraries in cataloging, their continual relevance in the BIBFRAME era is assured. As LC develops LC-PCC Policy Statements for post-3R Project RDA, their decisions will impact how RDA is implemented and encoded in both MARC and BIBFRAME. Since PCC libraries catalog using RDA, decisions made in the policy statements will influence all aspects of the PCC MAPs. In addition, any divergence between LC and PCC in implementation may have implications on the interoperability between LC and non-LC PCC data.

As a major contributor to the current PCC bibliographic and authority data files, and with their staff participation in internal BIBFRAME testing, LC will continue to be a major contributor of BIBFRAME data during the production-level implementation phase. Their on-the-ground experience with BIBFRAME will influence element cardinality and extensibility of PCC MAPs based on the unique needs of

material-specific formats. Besides creating BIBFRAME data natively, legacy data will need to be converted into BIBFRAME and vice versa to prevent abrupt interruption of workflows in individual libraries in the U.S. and around the globe. With its leading role in setting specifications for conversions between BIBFRAME and MARC, LC has been submitting MARC discussion papers and proposals to the MARC Advisory Committee for consideration based on needs related to the conversion. Decisions on mappings and conversion from LC, and decisions on the creation and management of PCC MAPs, will have significant impacts on each other as well as implications on the interoperability of MARC and BIBFRAME data in the future.

OCLC

OCLC has completed a number of pilot projects with linked data, already provides linked data with Faceted Application of Subject Terminology (FAST) and the Virtual International Authority File (VIAF), and will continue as a major provider of linked data with work to build an Entity Management Infrastructure funded by a Mellon grant, focusing initially on person and work entities. An entity production system will be in place by the end of 2021. By providing the home for the CONSER database and the place where many libraries share their BIBCO contributions, WorldCat will continue to be a place where bibliographic data is widely shared in both MARC and linked data forms. PCC expects OCLC to be able to both ingest and export BIBFRAME data in the future, as well as to ingest many other standard formats of bibliographic data. Development of PCC MAPs will include input from OCLC, just as OCLC currently contributes to PCC standards and to MARC development. Reliable PCC MAPs that are widely utilized will enable OCLC to consistently ingest data for wide sharing within the global library community.

Other Profile Creators

PCC is not alone in the business of creating metadata application profiles in the GLAM community. PCC libraries, though following RDA most of the time, also use non-RDA standards (e.g., [Describing Archives: A Content Standard \(DACS\)](#), [Descriptive Cataloging of Rare Materials \(DCRM\)](#), [Dublin Core Metadata Initiative \(DCMI\)](#), etc.) for special materials which may give rise to the need for extensions or deviating from the baseline MAP. With specific needs to address, MAPs created by others in the GLAM community give potential reference points to areas where extensions or deviations may be needed. Furthermore, MAPs from non-PCC origins may also share some common themes that can inform the PCC MAPs. Any required/recommended elements or preferred controlled vocabularies shared among these non-PCC MAPs imply a consensus that PCC may want to take into consideration when creating any PCC MAPs. Besides MAPs themselves, other profile creators may also have implementation plans on applications and technologies for MAP editing, data entry, conversion, and validation. These could be valuable references that inform the development and implementation of PCC MAPs. With that in mind, PCC should reach out and coordinate with specialized groups in the GLAM community (e.g., [OLAC](#), [MLA](#), [RBMS](#), etc.) to minimize overlapping of effort and potential compatibility issues.

Consumers of the Data

By moving to linked data, library data creators have the potential to increase exposure of their metadata to an audience beyond library vendors, patrons, other GLAM institutions, and bibliographic utilities if they wish. The increase in exposure could raise the discoverability of library resources through Internet search engines and linking to/from other linked data based web resources. Search engine optimization (SEO), including rich results/snippets displayed on search result pages, necessitates that metadata be encoded in certain formats with specific element-level requirements. For example, Google recommends that data should be encoded using the Schema.org vocabulary in [JSON-LD](#) serialization. To fully exploit SEO, the PCC must be aware of these requirements to ensure that *OUR* metadata, created in accordance to PCC MAPs, can be easily repurposed for consumption by search engines.

Moreover, there have been initiatives to reconcile library vocabularies with linked data sets created both inside and outside the GLAM community. As more of these “same as”¹⁰ or “close match”¹¹ relationships are asserted between data sets, the more “connectedness” library metadata will have. This increase in connectivity enables web content providers to harvest, reuse, and integrate library linked data with other content to deliver an enhanced online experience. When choosing or recommending controlled vocabularies for use in data elements of the PCC MAPs, the PCC will need to evaluate its use of ontologies in relation to other linked data datasets in order to maximize the connections.

On the other hand, web service providers and aggregators like Google Books and DPLA have been repurposing library data in prior formats for their services. They have existing processes to manipulate MARC and non-MARC data like MODS and Dublin Core to meet the needs of their services. Moving from these non-linked data formats to BIBFRAME will impact their workflows. In addition to requiring them to handle source data in a different format, this move may also disrupt existing functionalities in their services. Decisions and implementation timeline about the PCC MAPs should be clearly communicated to these content providers.

Discovery layers used by libraries are another prominent consumer of PCC data, and their functions are driven by the underlying metadata. However, various experiments have shown potentials of enhanced discovery experience using linked data, and discovery systems will also need to account for data that contextualizes library resources and their description. Consistent PCC MAPs will ensure that discovery system designers can expect standardized data regardless of the library system that generated it. In addition, PCC MAPs will allow the use and inclusion of vocabularies and ontologies that are not specified in the base model, which will enable a specialized discovery experience.

¹⁰ <https://www.w3.org/TR/owl-ref/#sameAs-def>

¹¹ <https://www.w3.org/TR/skos-reference/#mapping>

Developing and Consolidating Expertise

The success of MAPs within the PCC will depend on a critical mass of people who are equipped with knowledge of semantic web and linked data technologies, as well as knowledge of descriptive standards and practice for cultural heritage resources. One of the charges asked of the Task Group was to “identify the expertise already in PCC and any new experience/skills needed.” The Task Group decided to focus on the latter of the two questions in this charge, reframed as: *What new experiences and skills are needed for librarians – especially cataloging and metadata professionals – to work with MAPs, and how will those skills be obtained?* PCC members will need to develop new skills in semantic web and linked data technologies in order to understand how MAPs will work, particularly machine-actionable ones, in order to contribute constructively to discussions and development of PCC MAPs, and to express their needs in order to improve or extend MAPs developed by others.

As the first step, the Task Group identified the following as critical experiences and skills needed for understanding, creating, and using application profiles.

Competencies and Skills Needed

- Understanding of descriptive and subject cataloging standards
 - Strengths and limitations of traditional library controlled vocabulary practices
- Understanding of the importance of interoperability¹²
- Understanding of semantic web technologies (i.e., RDF and linked data concepts)
 - What is RDF
 - What is linked data
 - Understand the function of ontologies and how to work with multiple ontologies
 - Understanding of data validation tools/metadata application profile encoding schemes (e.g., ShEx, Shacl, etc.)
- Understanding of the impact of semantic web/linked data within the cultural heritage context
 - Vocabularies shared across cultural heritage communities (e.g., VIAF, [ULAN](#) and other Getty vocabularies)
 - Alignment and differences between different data models (i.e., [CIDOC CRM](#); [IFLA LRM](#); [RiC-CM/ICA](#); [RiC-O](#))
- Ability to reformat MAPs into different serializations
 - Depending on the “primary” form the MAPs take, the PCC and partners will need to consider how to support conversion to other formats (e.g., spreadsheets to/from JSON or other desired outputs depending on the direction related tools take)

¹² Zeng, Marcia Lei and Lois Mai Chan. 2006. Metadata Interoperability and Standardization-- A Study of Methodology. Part II: Achieving Interoperability at Record and Repository Levels. D-Lib Magazine, 12(6). <http://www.dlib.org/dlib/june06/zeng/06zeng.html> doi:10.1045/june2006-zeng; Chan, Lois Mai and Marcia Lei Zeng. 2006. Metadata Interoperability and Standardization -- A Study of Methodology. Part I: Achieving Interoperability at Schema Level. D-Lib Magazine, 12(6). <http://www.dlib.org/dlib/june06/chan/06chan.html> doi:10.1045/june2006-chan

Different Skills for Different Tasks

It should be noted that the Task Group thinks that these experiences and skills are not required for everyone who works with MAPs. Depending on the tasks and operational roles, the PCC community can expect two distinct levels of competency, one for those who will be creating resource descriptions based on one or more MAPs, who might need only a general introduction and training on the semantic web, and a second for those who will develop, implement, and maintain metadata application profiles, who will need a strong background in RDF, technical specifications for data validation tools and converters, and multiple ontologies. As tooling for MAP development and editing matures, a strong technical background may become less relevant for basic profile editing and cataloging workflow configuration. Most catalogers will require education on fundamentals of semantic web technologies, though in the future this knowledge might be expected as a baseline for all library professionals.

By the same token, as the library community develops greater familiarity with semantic web technologies, the PCC will need to continuously evaluate training needs for resource description. The Group also thinks that it is important for the PCC to plan and establish training modules for general cataloging and metadata professionals who will use PCC-approved MAPs to create quality data that meets PCC standards.

Identifying Expertise Inside and Outside of the Community

To support development and maintenance of the machine-actionable PCC approved MAPs, the PCC should identify those with expertise in and outside of the PCC community who will participate in the process. The Task Group believes the PCC needs to ask the following questions in order to accomplish this task:

- What relevant expertise and activities currently exist within the PCC, and how can the PCC build on any existing expertise;
- What relevant expertise and activities exist outside of PCC and how can the PCC avail itself of and/or participate in them;
- How can the PCC take advantage of existing expertise, especially from external partners (e.g., appointments to relevant PCC bodies).

Many cultural heritage communities have successfully created their own MAPs (including DCMI, LD4P2, and the Digital Library Federation) but much like the PCC, many of these communities are only starting to develop machine-actionable MAPs. Given the potential overlap of resources being described, the PCC could benefit from partnering with these groups in order to build robust MAPs in service of interoperable data, emphasizing collaboration and reuse across traditionally distinct communities as a core value, rather than as an afterthought.

Another option worth exploring is working with educators in iSchools to develop meaningful curricula that will ensure future library professionals will have relevant skills.

Available Resources

The Group identified two resources that provide relevant training, one open source and the other not:

1. [The DCMI Linked Data Competency Index \(2017\)](#)¹³

There are two parts that discuss application profiles: RDF Vocabularies and Application Profiles and Creating Linked Data Applications. While the Creating Linked Data Applications has only one section, Storing RDF Data, the RDF Vocabularies and Application Profiles section consists of 181 modules encompassing from Finding RDF-based vocabularies to the Designing RDF-based vocabularies.

2. [Library Juice Academy](#)¹⁴

The Library Juice Academy provides several basic courses (such as “RDF, RDFa, and Structured Data Vocabularies”) and more advanced courses (such as “The SPARQL¹⁵ Fundamentals” I, II, and III, “Introduction to JSON and Structured Data,” and “JSON-LD Fundamentals”)

Maintenance/Governance Models

The PCC describes itself as “a cooperative cataloging venture wherein members contribute bibliographic records and related data under a common set of standards and conventions using the bibliographic utilities.”¹⁶ Although the PCC now operates in a more complex standards and technical environment, its goal of promoting common practices so that high quality data can be produced and shared at scale remains a core value.

The PCC’s role will stay the same regardless of data format, continuing to provide a common set of standards and conventions for the linked data realm. To manage this, the main recommendation is to expand the current governing model within PCC to include a structure that ensures coordination among PCC’s stakeholders in the development and maintenance of MAPs. Stakeholders include the standing committees, other PCC groups engaged in linked data activities (e.g., LD4P and the PCC Linked Data Advisory Committee), the Library of Congress, and partner communities wishing to extend PCC MAPs. The structure also needs to enable communication and consultation with the PCC membership as a whole, and with key service and data providers.

¹³ <https://dcmi.github.io/ldci/D2695955/#a-rdf-vocabularies-and-application-profiles>;
<http://explore.dublincore.net/>; <http://explore.dublincore.net/explore-learning-resources-by-competency/>

¹⁴ <https://libraryjuiceacademy.com/courses.php>

¹⁵ SPARQL: SPARQL Protocol and RDF Query Language

¹⁶ <https://www.loc.gov/aba/pcc/>

Justification

- It is assumed that the PCC will create, promote, update, and maintain its own MAPs. Various PCC groups (e.g., standing committees, task groups) will play an important part in the development and maintenance of MAPs. The Library of Congress has historically been a prominent contributor to the development of PCC standards and this can be expected to continue, especially with its role in creating PCC MAPs. The Standing Committee on Standards and the Standing Committee on Applications play important roles in providing PCC member libraries with descriptive standards and technology. It would be appropriate to expand the roles of these groups to MAP maintenance. These committees (or task forces reporting to these committees) will need to work closely with the Standing Committee on Training to provide training for members in using MAPs and the tools that support them. There will likely be a significant number of people seeking guidance on using MAPs or needed new ontologies. The PCC will need to be able to respond to meet those needs in a timely manner. Ensuring open and continuing communication between member libraries will be important as the PCC moves into linked data.
- With the PCC's increasing emphasis on pilot projects that serve as a testbed for new practices, it will be important to have channels for these projects to provide feedback to the PCC on issues that emerge from their work.
- Metadata application profiles will be created not only by the PCC but also by partner communities, including but not limited to MLA, OLAC, and RBMS. MAPs developed by these groups may specify additional or different ontologies from those expected by PCC's MAPs, to best describe the resources covered by their specialized community. It would be desirable for these groups to build on a base level PCC MAP with extensions for their needs rather than invent their own MAPs. Therefore, communication between the PCC and other community groups is more important than ever. Regular communication and cooperation with these groups would be ideal. In addition, there may be technical questions surrounding profile extensions that it will be necessary to explore as part of this collaboration.
- The Task Group expects that PCC MAPs will be used not only by libraries and other cultural heritage communities, but also by service providers, application developers, and researchers who use library data. Incorporating as many of these stakeholders as possible when initially developing MAPs and subsequently into any governance model will greatly benefit the library community by cutting down on the proliferation of similar MAPs.

Recommendations

While there are several emerging technologies that the PCC may embrace that may prove critical to the development of RDF-based MAPs, the PCC should not lose sight of its core value as an entity that provides a common set of standards and conventions for the community. The Task Group respectfully recommends the following as a start on PCC MAP governance:

- Standing Committee on Standards will be responsible for MAP content, just as they are currently responsible for the maintenance of the BSR and CSR.
- Standing Committee on Applications will be responsible for recommending, testing, and moving forward the technology to implement machine-actionable MAPs.
- Standing Committee on Training will be responsible for developing and providing training to PCC members on use of the MAPs that are adopted by PCC.
- Formation of a new task group, reporting to the Standing Committee on Standards, that has representatives from each of the Standing Committees and includes others with relevant expertise. Using the recommendations of the PCC Task Group on Metadata Application Profiles as a basis, this task group will be responsible for creation, maintenance, and management of PCC MAPs in collaboration with the Standing Committees and with relevant PCC groups actively engaged in linked data activities. This group will also be responsible for working with the stakeholders listed in section 3 to meet shared objectives. This group could be charged with ensuring continuing substantive communication with the PCC membership regarding MAP developments. Resolution of concerns and issues identified by MAP stakeholders will also be the responsibility of this group. Additional responsibilities may be determined through consultation with PoCo and the Standing Committees. The new task group will be established with an initial term of two years, with a review at the end of that time.

Environmental Scan - Linked Data and MAPs

As has been previously mentioned, there is no perfect choice from the existing standards and related technologies for capturing MAPs. The following list, which was generated through an environmental scan, should be evaluated when the PCC weighs its needs with the strengths and weaknesses of these options. Also worth considering is what strategies are available for converting between different options.

[BIBFRAME Profiles](#)

BIBFRAME Profiles are described as “primarily a means for an application, e.g. cataloging tool, to guide a cataloger in the creation or modification of a BIBFRAME record. But such Profiles can also be used in other ways such as a formal representation of the constraints associated with a content description standard (e.g., RDA, DACS, etc.) or agreed upon representation for interoperability.” While not an official standard for creating MAPs or defining RDF constraints, the BIBFRAME Profile specification was developed by the Library of Congress for the BIBFRAME community to use to configure cataloging tools that can communicate shared data targets for interoperability. BIBFRAME Profiles are captured using JSON, and LC has provided a profile editor to author these JSON files. The BIBFRAME Editor and Sinopia, based on the BIBFRAME Editor, are the only known cataloging tools to use the BIBFRAME Profile specification, and unless adopted by other tools, BIBFRAME Profiles would likely need to be converted to other serializations for use by other tools.

W3C

[SHACL](#)

Shapes Constraint Language (SHACL) is a W3C standard providing “a language for validating RDF graphs against a set of conditions. These conditions are provided as shapes and other constructs expressed in the form of an RDF graph. RDF graphs that are used in this manner are called ‘shapes graphs’ in SHACL and the RDF graphs that are validated against a shapes graph are called ‘data graphs.’ As SHACL shape graphs are used to validate that data graphs satisfy a set of conditions they can also be viewed as a description of the data graphs that do satisfy these conditions. Such descriptions may be used for a variety of purposes beside validation, including user interface building, code generation and data integration.” SHACL is growing in adoption, but user friendly tools for authoring shapes and cataloging tools that use these shapes remain few.

[ShEx](#)

The Shape Expressions Language was developed by a W3C community group and at the time of this writing is version 2.1 (October, 2019). ShEx was developed in parallel to SHACL and performs many of the same RDF validation functions as SHACL. ShEx validation rules can be written in JSON or in the compact ShExC language. The community group provides open source validation tools in javascript, Java, Scala, Ruby and Python. ShEx is being used in some large medical information environments and has been adopted by the Wikidata community for the expression of their [data schemas](#) and for validation.

[W3C DCAT](#)

DCAT scheme is an RDF vocabulary to describe catalogs of datasets. It is primarily used within EU countries. A general-purpose application profile, DCAT-AP, was developed, but because each country already had its own practices, versions of DCAT-AP (DCAT-AP-DE, DCAT-AP-IT) were developed. This is a living example of profiles of profiles, and the W3C working group, [Data Exchange Working Group](#) (DXWG) is considering a vocabulary to describe relationships between profiles, the [Profiles Vocabulary](#).

[W3C ODRL](#)

This document defines a data model for open digital rights language when expressing policies that require classes and concepts for its assertions. As an actionable vocabulary, ODRL has its rules that govern the relationship of profiles to the policies they profile, where the profile elements may override those of the base policy. This is an example of specific rules governing profile relationship to the base.

Others

[CEDAR](#) (Center for Expanded Data Annotation and Retrieval)

The CEDAR workbench is a suite of open-source, Web-based tools and APIs creating a modular environment for authoring metadata, originally for use with BioPortal, a portal for bioinformatics ontologies led by the National Center for Biomedical Ontology. It allows users to construct metadata templates built from existing ontologies, fill in the metadata using the templates, share & manage the ontologies, the templates, and the data; and leverage available APIs for validating and submitting metadata to external repositories. The tool creates application profiles, and is not an application profile in itself

[DCMI DCAP](#)

The Dublin Core community has a general model, the [Singapore Framework](#), to describe the context for applications profiles. From this a vocabulary was developed, the [Description Set Profile](#). Currently, DCMI has a [working group](#) looking at the development of a simple “core” vocabulary for application profiles that can be described in a tabulary (spreadsheet) format.

[DPLA](#)

DPLA metadata profile is built on the Europeana EDM, based on [Conway’s Law](#), and references 14 different namespaces for its data contributing partners. The primary audience of DPLA is general public and cultural heritage institutions users.

[Europeana EDM](#)

Europeana reuses data elements already exist to facilitate its aggregating data effort from GLAM in Europe. There is the EDM model that data must be provided in. However, there are a variety of flavors, such as EDM external (for additional data enrichment if needed) and EDM internal (all supplied data must use).

[YAML](#)

This is a human-friendly, textual syntax markup language for configurations and can be used for metadata application profile creation, management, and publishing. It is not a new standard of metadata formats. It is intended for domain experts who do not have extensive knowledge on metadata application but an easy to use preprocessor in order to create standard metadata application profiles. An example of a MAP using YAML is [YAMA](#).

Conclusion and Consideration

The Task Group acknowledges throughout the various sections the challenges in creating, implementing, and managing MAPs in the PCC community space. The PCC as an organization does not have an infrastructure that lends itself to accomplishing some of these tasks well. For instance, it relies on membership to provide system support solutions concerning documentation and working documents from committees and working groups. Going forward, the Task Group concludes that the challenges that the community faces can be looked at in two ways, internal and external.

Internally, the Task Group believes that the PCC needs to establish a task group as recommended in the “Maintenance/Governance Model” section to oversee the creation and management of MAPs. This group should usher and facilitate the different tasks associated with MAPs, such as the creation of machine-actionable profiles. Librarians who serve in this group should be the colleagues who possess a deep understanding of the library resource description and who are equipped with a skills set able to address the PCC response and to take appropriate action regarding current and future semantic web technology.

Externally, the Task Group believes that the PCC should build a close working relationship with relevant groups beyond the library domain that are experienced in creation and management of MAPs. Continuous MAPs development is not unique to the PCC community. Other communities described above, such as the W3C and Wikidata, continue refining MAPs for their users. The PCC will benefit greatly by taking into account the lessons learned from these colleagues and work closely with stakeholders alluded to in the previous sections.

Ideally, resources and training materials must have a permanent location under the auspice of the PCC. This ensures knowledge and skill continuance and archival purposes. Many PCC groups, e.g., Standing Committee on Training, task groups, have established curricula and documents for professional development and task-oriented training. The continuing education to realign with MAP development needs, present and future, benefits the PCC community, and libraries at large. The PCC as an organization relies on member institutions to satisfy the storage need. This lack of a system infrastructure, from the Task Group’s perspective is less than ideal in responding to a rapidly changing information environment. The Task Group recommends the PCC community reach a solution that will meet the needs.

The PCC community has carried the responsibilities of distilling standards and providing best practices guidelines for resource description and authority control. Transitions toward implementing linked data for resource description will probably be one of the greatest challenges that all libraries will face in the coming years. Phased- and/or parallel-system transitions, the management of various steps, libraries will look to the PCC community for managing the move from MARC to a linked data environment.

Acknowledgements

This report has benefited greatly from following colleagues who shared their respective metadata application profiles for their community of users:

- Renate Behrens and Christian Aliverti (EURIG)
- Michael Della Bitta (DPLA)
- Valentine Charles (Europeana)
- Thomas Baker (DCMI)

Task Group Members

Corine Deliot (Consultant, British Library)

Cynthia Whitacre (OCLC)

Heather Pretty (Memorial University of Newfoundland, PoCo liaison)

Honor Moody (Harvard)

Jackie Shieh (Co-chair, Smithsonian)

Jodi Williamschen (Library of Congress)

Karen Coyle (Liaison to the DCMI profiles group)

Lucas Mak (Michigan State University)

Les Hawkins (Library of Congress)

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

Nancy Fallgren (Consultant, National Library of Medicine)

Nancy Lorimer (Stanford)

Chew Chiat Naun (Harvard)

Steven Folsom (Cornell)