PCC Task Group on URIs in MARC
Year 2 report to PoCo, October 2017
October 23, 2017

INTRODUCTION

This document provides a cumulative overview of the work of the PCC URI Task Group over its first two years, together with an outline of plans for our third and final year.

The work of the Task Group has so far focused on three main areas: addressing structural MARC issues such as provisioning for work entities, relationship URIs, and the distinction between authorities and real world objects; producing documentation to meet critical needs that we have identified, including guidelines for formulating URIs from a range of widely used data sources and a set of tables specifying the subfields corresponding to the object of a statement derived from a MARC field; and engagement with service providers on requirements for tools and services. As the Task Group prepares this report, several members have begun drafting FAQs to provide clarification regarding URI practices.

We will continue to work in all of these areas, but the emphasis in our third year will shift toward implementation issues. These will include development of best practices, which will need to take into account data flow issues involving service providers such as OCLC; outreach to assess community adoption and identify training and tooling needs; aligning our objectives with those of colleagues involved in related efforts under the PCC umbrella, such as implementation of newer vocabularies and exploration of alternative models for authority data; and, crucially, finding a home within PCC for various aspects of our work so that it can be sustained into the future. We propose handing off recommendations concerning best practices to the Standing Committee on Standards. Findings gathered

from the URIs Survey that the Task Group conducted between August and September will be passed on to the Standing Committee on Training.[1]

RECOMMENDATIONS

Below are several areas where the Task Group believes further effort is needed. These include several where the Task Group has begun work, but is likely to need assistance from other groups within PCC. We submit them here as recommendations to the Policy Committee (PoCo).

1. Developing an implementation plan for $0, $1, and $4 in bibliographic records, with OCLC and other data providers and aggregators as essential stakeholders.
2. Developing best practices and related training material for use by the cataloging and vendor community. These should cover significant new MARC provisions such as the 758 field.
3. Taking steps to permit use of $0, $1, and $4 in NACO authority records. A summary of the issues may be found in Appendix 2.
4. Developing a plan for ongoing maintenance of relevant Task Group outputs, including the Formulating URIs and MARC Object Reconciliation documents, and the URIs in MARC FAQs.
5. Developing a process to evaluate and endorse vocabularies for PCC use.
6. Including URIs in any future implementation plans for vocabularies endorsed for PCC use, e.g. ISNI and LCGFT.
7. Clarifying internal PCC protocols for developing and submitting MARC discussion papers and proposals.
8. Identifying issues that arise in related PCC efforts, such as specifications for converting MARC to linked data, and coordinating the work of the relevant groups.
9. Helping to prioritize remaining work areas.

Progress in these areas, as in many others, crucially depends on PCC's ability to follow through on the work of its task groups and committees, and to put their recommendations into effect in a timely manner. The Task Group is grateful to the PCC leadership for its continued support of our work.

SUMMARY OF OUTCOMES TO DATE

Structural MARC issues

MARC 21 is a legacy format with many layers of complexity resulting from the needs of a range of stakeholder communities over several generations.[2] Retooling it to support linked data invariably requires judicious compromise, weighing available options against the likely cost and benefit of implementation. For the last two years the majority of the Task Group's efforts have been on clarifying the syntax and semantics of MARC subfields for URIs. A guiding principle has been to maximize the returns and minimize the disruptions to existing infrastructure in readying MARC data for transformation to RDF triples.

---

[1] FAQs and Survey: MARC Records with RDF URI. Form available at: http://goo.gl/zn55mM

[2] The Task Group also received inquiries from users of UNIMARC and INTERMARC, such as Bibliothèque nationale de France (BNF) and National Diet Library (NDL).

The Task Group's efforts can be divided into two main areas: devising interpretations and best practices that will allow us to work around existing ambiguities in the format; and advancing proposals for new elements or definitions that will allow MARC to support needs it was previously unable to meet. Among the new MARC provisions are the following:

$1 for real world objects. MAC proposal 2017-08 restricts the use of $0s to identifiers from authority sources such as the Library of Congress' linked data service (id.loc.gov). By contrast, $1 was introduced for identifiers that point to real world objects.[3] This results in separating URIs that describe the resource from the ones that point to real world objects.

758 for resource identifiers. MARC had previously lacked a dedicated field for work identifiers, although existing fields such as 787 had been pressed into service for this purpose in some implementations. 758 was defined with linked data principles in mind, so that it deliberately avoids (for example) the elaborate subfielding provided for access points in the 7XX block, and places greater reliance on $4 for specifying the nature of the relationship between instance and work.

$4 for relationship URIs. As a result of a British Library proposal supported by the PCC URI Task Group, $4 can now contain either a relationship code and/or a URI. This remedies what had been a significant limitation of MARC: the inability to specify a relationship unambiguously, and therefore to support use of multiple vocabularies for relationships.

There are now bibliographic records in OCLC that have been enhanced by adding the $4 with a URI to a number of different fields. Some examples are given below for illustrative purposes. The second example also shows the provision of a URI in conjunction with LCDGT terms.

100 1# $a Shapiro, Barbara A., $d 1951- $e author. $4 http://rdaregistry.info/Elements/a/P50195 $4 http://id.loc.gov/vocabulary/relators/aut

386 $4 http://id.loc.gov/vocabulary/relators/fmd $a New Zealanders $a Men $2 lcdgt $0 http://id.loc.gov/authorities/demographicTerms/dg2015060357 $0 http://id.loc.gov/authorities/demographicTerms/dg2015060359

651 #0 $a San Francisco (Calif.), $e setting. $4 http://id.loc.gov/vocabulary/relators/stg

780 00 $4 http://rdaregistry.info/Elements/w/P10226 $t Fishing news international $x 0015-3044 $w (DLC)sn 86012429 $w (OCoLC)1569329

The Task Group also invested a considerable amount of effort in analyzing ambiguities or limitations in existing MARC definitions and proposing interpretations and best practices that would minimize them. For example, although $0 is defined for many MARC fields it is nowhere stated explicitly which parts of the MARC field it pertains to. After extensive discussions and modeling of which MARC fields and subfields would singly or combined represent the object that corresponds to the URI in $0, the Task Group created a set of tables that will help cataloging professionals and programmers understand the application of URIs in $0 and translation of that data into RDF. A sample is shown below.

---

[3] *See* MARC Proposal, 2017-08, Use of Subfields $0 and $1 to Capture Uniform Resource Identifiers (URIs) in the MARC 21 Formats. Available at: https://www.loc.gov/marc/mac/2017/2017-08.html

MARC data field	MARC subfields (singly or combined) equating to RDF Object
100 - Main Entry - Personal Name (NR)	abcdgjq
110 - Main Entry - Corporate Name (NR)	abcdn
111 - Main Entry - Meeting Name (NR)	acdenq
650 - Subject Added Entry - Topical Term (R)	All but numbered subfields 2368
651 - Subject Added Entry - Geographic Name (R)	All but numbered subfields 2368
654 - Subject Added Entry - Faceted Topical Terms (R)	All but numbered subfields 2368
655 - Index Term - Genre/Form (R)	All but numbered subfields 23568
656 - Index Term - Occupation (R)	All but numbered subfields 2368
657 - Index Term - Function (R)	All but numbered subfields 2368
700 - Added Entry - Personal Name (R)	Name: abcdgjq Name/title: abcdfghklmnoprst
710 - Added Entry - Corporate Name (R)	Name: abcdgn Name/title: abcdfghklmnoprstu
711 - Added Entry - Meeting Name (R)	Name: acdegnq Name/title: acdeghqklnpst
730 - Added Entry - Uniform Title (R)	adfghklmnoprst
800 - Series Added Entry - Personal Name (R)	abcdfklmnopqrst
810 - Series Added Entry - Corporate Name (R)	abcdfklmnoprst
811 - Series Added Entry - Meeting Name (R)	acdefklnpqst
830 - Series Added Entry - Uniform Title (R)	adfklmnoprst

Conversely, there are many instances where a URI is not available to represent the entire content of a MARC field. A commonly encountered example is subdivided subject headings where the combination is not explicitly authorized in the LC authority file.

650 #0 $a Prehistoric peoples $z Asia, Central.

In this example, there is a URI for the topical term, "Prehistoric peoples" (http://id.loc.gov/authorities/subjects/sh85080302) and a URI for the geographical region "Asia, Central" (http://id.loc.gov/authorities/subjects/sh85008625) but not a URI representing the entire subject heading as a unit.

After careful consideration, the Task Group recommends against providing URIs that represent only partial entities of a MARC field. Faceted vocabularies provide an alternative means to represent such concepts by post-coordination. If the entire concept is to be represented as a single semantic unit within the LCSH vocabulary, in our view that becomes an issue for the maintenance agency rather than for implementers. Indeed, LC is well aware of this issue and is taking steps to address it.[4]

The repeatability of $0 raises a number of issues for implementation. Nothing in MARC prohibits the inclusion of multiple occurrences of $0 pointing to different objects. The Task Group noted the German National Library's practice of including multiple $0s only if they designate the same entity. Fields in which $a and other significant subfields are repeatable also create difficulties, since this raises the issue of which subfield occurrence each $0 corresponds to. The Task Group believes that these issues can largely be addressed through best practices. In many cases, ambiguity can be avoided by the simple expedient of repeating the field.

Documentation

During the course of the Task Group's discussions we identified a number of critical gaps in the documentation that was available to metadata practitioners. A particularly important need was for a set of guidelines for creating well-formed URIs from a range of widely used data sources. In response to this need Task Group members produced a Formulating URIs document which will shortly be ready for publication. A second area of need that we identified was for clarity concerning which subfields the $0 (or $1) pertains to. In existing MARC practice this could seem ambiguous, not only to catalogers but perhaps even more crucially to application developers not versed in MARC assumptions. The Task Group adopted the principle that the $0 could refer to only one entity within each MARC field, and we have created a set of tables to document which subfields correspond to that entity within each field. (A sample is included above.) Our MARC proposals also reflect the same underlying principle. These tables are still a work in progress, with release planned for next year. Also in preparation is an FAQ providing information on basic concepts, available resources, and the current state of our work. In addition to these task group outputs, OCLC hopes to issue a paper in the near future on the issues surrounding real world object URIs.

These task group documents will require ongoing maintenance if they are to remain of use to practitioners, and some group within PCC will need to be assigned responsibility for them once the URI group concludes its work. In addition, the Task Group has encountered issues surrounding infrastructure for collaborative work and public sharing of outputs, and will raise them with the PCC leadership.

Engagement with service providers

OCLC representatives were deeply involved in the development of several of our MARC proposals, notably the successful proposals to introduce $1 for real world object URIs and 758 for work entities. We have held discussions with OCLC about support for URIs in MARC outputs, and OCLC's statement about its plans is included here as Appendix 1. Among other tool providers, Terry Reese has already built support for the new MARC provisions into his MARCEdit suite of tools.

---

[4] See, for example, Kevin Ford, "When URIs become Authority", http://www.loc.gov/bibframe/pdf/ALAmw2013-sac_Ford.pdf.

The outgoing chair recently invited Lihong Zhu, a member of the Ex Libris Linked Open Data Special Interest Working Group, to join the Task Group as a liaison. In October several members of the Task Group attended a workshop to develop use cases for Casalini Libri's SHARE-VDE service, and we will remain in contact with them about Task Group recommendations for best practices and their potential use in Casalini data.

YEAR 3 OBJECTIVES

Developing best practices for inclusion of URIs in PCC MARC records. The Task Group's efforts so far should be viewed only as preparatory to the real business of populating our records with URIs and making their use and exchange an integral part of cataloging workflows. Setting down guidelines for use of MARC URI fields and subfields will give both libraries and data providers (including cataloging and authority vendors) a clear set of objectives to work towards.

- A particular challenge, and an area where guidance will especially be needed, is implementation of the new MARC provisions that introduce practices that may not yet be familiar to many practitioners: the distinction between authorities and real world objects, the use of work and other entities identified by URIs rather than authorized access points, and the use of URIs to designate predicates unambiguously.
- Authority data are an area which is ripe for the introduction of URIs and where their routine use could have a large impact. Some issues surrounding current practice and possible next steps are laid out in Appendix 2.

Providing training and continuing education opportunities for metadata practitioners. Anecdotal evidence suggests that there is an increasing awareness among metadata practitioners of the benefits of using identifiers in their workflows. However, there appears still to be a need for wider dissemination of information about issues that arise for their implementation in MARC, as well as about the specific solutions proposed by the Task Group. As best practices are developed, training materials will be needed and the Task Group hopes to work with the PCC standing committees to produce them. In the coming year Task Group members will also be looking for opportunities to present at community forums, including webinars.

Engaging with data providers on provision and handling of URIs. Much of the effort of populating existing records with URIs will be undertaken by cataloging and systems vendors. The group has recently begun to reestablish providers who are active in this area. An important stakeholder is OCLC, who serves as the main data exchange hub for most PCC members. OCLC has undertaken to provide (optional) output of URIs derived from controlled headings in WorldCat, and this represents a very significant step forward. Handling of URIs in incoming records from contributing libraries presents a different set of challenges. OCLC already has arrangements in place for accepting URIs from some national libraries and from NLM for MeSH, and these may serve as a partial model for their handling of URIs from other member libraries. However, the handling of URIs supplied in fields that are traditionally controlled using LC authorities presents more complicated questions. The Task Group hopes to work through a set of use cases with OCLC that may include situations such as the following:

- A library supplies a record that matches an existing one in WorldCat, but has a 100 giving a $1 for an ISNI rather than the LCNAF controlled heading in the WorldCat record.
- A library supplies local as well as standard URIs in 1XX/7XX $0 and $1.

- One library supplies a textual identifier in $0/$1, while another supplies the corresponding URI.

Assessing related activities and impact. The work of the URI Task Group has close affinities with other efforts currently under way in the library community, including several within the PCC itself. The management of identities via ISNI, for example, points to an increased role for $0/$1 data in PCC bibliographic and authority records. The PCC's interest in faceted vocabularies from LC and elsewhere also presents an opportunity to incorporate URIs into the strategy for their implementation. The forthcoming PoCo strategic planning sessions will provide an opportunity to align the Task Group's work with these efforts. Towards the end of its second year the Task Group conducted a survey in order to form a more complete picture of the current range of activities (as well as to identify unmet needs and concerns) among libraries interested in the use of identifiers. The Task Group will continue to analyze the results of this survey and hopes to repeat the exercise toward the latter part of its third year.

Handing off ongoing responsibility to standing committees and other PCC groups. It is clear that practice will continue to evolve after the Task Group concludes its third year. At a minimum, the Formulating URIs document will need to be updated regularly. Vocabulary sources and tools and systems alike can also be expected to continue to develop, and PCC will need to evaluate them for member use on an ongoing basis. Above all, PCC cataloging practices will continue to evolve in the light of further experience with linked data principles. It will be one of the Task Group's essential responsibilities to work with the PCC leadership to ensure that these responsibilities have appropriate ownership within PCC after the life of the Task Group.

URI SURVEY OBSERVATIONS

To help inform its third year strategy, the Task Group shared a survey on August 17, 2017 on several listservs, e.g. BIBFRAME, OCLC-CAT, METADATALIBRARIAN, etc., to collect information regarding the following:

1. Context, Goals, Staffing among libraries that have planned or are in the process of planning for implementing URIs in $0 in MARC.
2. Awareness of recent MARC standard updates related to identifiers that the Task Group has worked on and libraries' plans for implementation.
3. Platforms, Tooling, Vendors that libraries are using or considering.
4. URI for non-MARC data.
5. The kinds of help that libraries are seeking from an organization such as PCC.

Several themes emerged from the survey data which spoke to initial survey objectives. These themes included:

1. There was a strong desire to work with URI in MARC, but respondents were not sure about best practices to pursue this work. This was highlighted by mixed response to whether consideration would be given to use of $0 and $1 separately as well as the new 758 field.
2. While respondents generally noted expected benefits of URI in MARC, uncertainty remains about how discovery will improve and there was a general lack of clarity on how and when return on investment for this work might be realized.

3. Staffing of these initiatives mainly involves cataloguing and metadata librarians and system librarians
4. Survey results indicated a strong desire to have further communication about best practices and additional documentation.

Another finding was that many were making use of URIs in MARC in conjunction with experiments with transforming MARC data to BIBFRAME. Several data modeling issues emerged, including, but not limited to:

1. Handling of parallel fields for original scripts (see appendix 3)
2. How to record predicates that can express unambiguous relationships between objects and/or agents and objects.

This initial survey encountered a number of logistical difficulties, including some unforeseen problems with the survey form itself. The Task Group will continue to analyze the survey results and will consider a follow-up survey to answer outstanding questions.

RELATED SOURCES

URI Task Group Charge and reports 2015-2017. See Task Group wiki
https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html
Linked data advisory group report June 2017
https://www.loc.gov/aba/pcc/documents/LinkedDataInfrastructureModels.pdf

Appendix 1

OCLC's support for $0 and $1
Jean Godby
13 October 2017

Support for $0

As currently implemented in WorldCat, $0s are used to store identifiers for selected authority files, including the Dutch NTA Names file and the German GND headings.

A different approach is used to store identifiers for LCNAF, LCSH, and MeSH.
- For these three schemes, OCLC stores record identifiers outside the MARC record structure, which are converted to clickable links when controlled headings are displayed in Connexion and WorldShare Record Manager.
- If users supply a $0 for controlled fields, it will be removed when the heading is controlled by OCLC through nightly automated processes.
- $0s will be retained if they are supplied in other situations in which the subfield is allowed. This includes (but is not limited to) name and subject headings belonging to vocabularies other than LCSH, LCNAF, and MeSH.

Starting on December 31, 2017, the WorldShare Record Manager application will have an option to export records that will include $0s containing a URI for the authority records in GND, LCNAF, and MeSH. This feature will be expanded to support more authority files, and will be available in more OCLC products, including the Metadata API, in the future.

Support for $1

URIs for real world objects that are broadly understandable across domains represent a key feature that distinguishes legacy standards from linked data, and OCLC's researchers have pioneered the transformation in the library community.

For example, VIAF descriptions were amended in 2011 to be 'about' real-world people, places, organizations, and creative works. At the same time, the VIAF server was upgraded to implement the Cool URIs web protocols for delivering information about real world objects. These changes brought VIAF into full compliance with linked data conventions. In earlier versions of the VIAF data model, the primary object was a cluster of headings, which was of interest primarily inside the library community.

In WorldCat catalog records, real world object URIs have been visible in the "Linked Data" tab since 2012. For example, note the RDF description for an audiobook version of *Zen and the Art of Motorcycle Maintenance* excerpted below. The author, Robert Pirsig, is represented by a VIAF real world object URI <http://viaf.org/viaf/78757182>. The same URI is reproduced in a subject statement (expressed as schema:about), enabling a machine process to make the reasonable inference that this book is both by and about the same person.

<http://www.worldcat.org/oclc/828862621> # Zen and the art of motorcycle maintenance : an inquiry into values
    a schema:CreativeWork, bgn:CD, bgn:SoundRecording, schema:Book ;
        library:oclcnum "828862621" ;

library:placeOfPublication <http://id.loc.gov/vocabulary/countries/riu> ;
schema:contributor <http://viaf.org/viaf/78757182> ; # Robert M. Pirsig
schema:about <http://id.loc.gov/authorities/subjects/sh85119708>;
schema:about <http://viaf.org/viaf/78757182> ; # Robert M. Pirsig
schema:about <http://id.worldcat.org/fast/1111441> ; # Self
schema:bookFormat bgn:AudioBook ;

The real-world object URIs in this description were automatically generated from OCLC's internal data structure, but are only available in the linked data markup, not the associated MARC record. This is because no MARC field with the appropriate meaning was available until the $0/$1 proposal (MARC 2017-08) was approved in June 2017.

OCLC is actively involved in the development of this new standard. Jean Godby, OCLC Senior Research Scientist, is chairing the PCC-URI RWO subgroup, which is now working to make the $0/$1 standard mature and actionable; open issues are listed in the RWO subgroup section of this report. The major impediment to full implementation is the identification of criteria that make the distinction clear to catalogers and machine processes. But as our linked data published on WorldCat already demonstrates, OCLC researchers believe that the task can be at least partially automated, and we are working on a more robust solution.

Appendix 2

URIs in Authority Data

Subfields $0, $1, and $4 are valid in numerous authority fields. The Task Group recommends that PCC take steps to permit their inclusion in LC/NACO authority records. The *LC Guidelines Supplement to the MARC 21 Format for Authority Data* (https://www.loc.gov/catdir/cpso/lcmarcsuppl.pdf) will need to be revised, as it currently proscribes the use of these subfields in most fields where they are valid. Subfield $1 is too new to have been mentioned at all in the guidelines supplement.

0XX Fields

The *LC Guidelines Supplement* does not bar the use of $0 in two 0XX fields: 034, 043.

In other 0XX fields, $0 is prohibited. For example:

    050   Do not use subfield (implementation decision not yet made):
            $0

    053   Do not use subfield (implementation decision not yet made):
            $0

    060   Do not use subfield (implementation decision not yet made):
            $0

As a test of the fields where it is currently not prohibited, $0 was added to the 034 and 043 fields of the following authority records in OCLC Connexion, and the records successfully validated:

034 ## $d W1221955 $e W1221955 $f N0473622 $g N0473622 $0 http://sws.geonames.org/5809844 $2 geonames
034 ## $d -122.33207 $e -122.33207 $f 47.60621 $g 47.60621 $0 http://sws.geonames.org/5809844 $2 geonames
151 ## Seattle (Wash.)

043 ## $a e-gx--- $0 http://id.loc.gov/vocabulary/geographicAreas/e-gx
151 ## $a Germany

3XX Fields

Numerous guidelines for 3XX fields say:

        Do not use subfields:
            $0, $6 or $8

Others (e.g., 370) say:

        Do not use subfields:
            $0, $6, $8
        Do not use subfields (implementation decision not yet made):

$i, $3, $4

The guideline for field 375 does *not* prohibit the use of $4, but an attempt to include it in an authority record in OCLC Connexion did not pass validation:

375 ## $a Males $0 http://id.loc.gov/authorities/demographicTerms/dg2015060003 $2 lcdgt

5XX Fields

The 500 field guidelines prohibit using subfields $0 and $4:

> Do not use subfields:
> $e, $h, $j, $v, $x, $y, $z, $0, $4, $5, $6, $8

The 510, 511, 530, and 551 field guidelines, on the other hand, do not prohibit using $0:

> 510 Do not use subfields:
> $e, $h, $v, $x, $y, $z, $4, $5, $6, $8
>
> Consult LC's Cooperative Programs Section before using subfield $0.
>
> 511 Do not use subfields:
> $h, $j, $v, $x, $y, $z, $4, $5, $6, $8
>
> Consult LC's Cooperative Programs Section before using subfield $0.
>
> 530 Do not use subfields:
> $g, $h, $v, $x, $y, $z, $4, $5, $6, $8
>
> Consult LC's Cooperative Programs Section before using subfield $0.
>
> 551 Do not use subfields:
> $g, $v, $x, $y, $z, $4, $5, $6, $8
>
> Consult LC's Cooperative Programs Section before using subfield $0.

It is unclear why the text is different for 500. In any case, we believe that the time is right to start allowing NACO libraries to include URIs in 5XX fields. As it turns out, in OCLC Connexion, $0 actually already does validate in 500, 510, 511, 530, 551. For example, when $0 was included as a test in 5XX fields of the authority records shown below, the records all validated in Connexion:

100 1# $a Leão, Ricardo
500 1# $i Real identity: $a Martins, Ricardo André Ferreira $0
http://id.loc.gov/authorities/names/n2015210032 $w r

110 2# $a American Society for Abrasive Methods
410 2# $a A.S.A.M.
410 2# $a ASAM

510 2# $i Predecessor: $a American Society for Abrasives $0
http://id.loc.gov/authorities/names/n82136819 $w r
510 2# $i Successor: ‡a Abrasive Engineering Society (U.S.) $0
http://id.loc.gov/authorities/names/n82136820 $w r

151 ## $a Sri Lanka
551 ## $a Ceylon $0 http://id.loc.gov/authorities/names/n80061038 $w a

However, when $4 was added to the same fields in these records, the records did not validate. For
example:

110 2# $a American Society for Abrasive Methods
410 2# $a A.S.A.M.
410 2# $a ASAM
510 2# $4 http://rdaregistry.info/Elements/a/P50012 $i Predecessor: $a American Society for Abrasives
$0 http://id.loc.gov/authorities/names/n82136819 $w r
510 2# $4 http://rdaregistry.info/Elements/a/P50016 $i Successor: ‡a Abrasive Engineering Society
(U.S.) $0 http://id.loc.gov/authorities/names/n82136820 $w r

4XX Fields

Subfields $i and $4 are also established in the see from tracing fields. The *LC Guidelines Supplement* for
Tracings and References – General Information – 4XX Fields says "Do not use subfield $i or subfield $w
code r in 4XX fields." The guidelines for each individual 4XX field also prohibit the use of subfield $4:

> 400 Do not use subfields:
> $e, $h, $i, $j, $v, $x, $y, $z, $4, $5, $6, $8
>
> 410 Do not use subfields:
> $e, $h, $i, $v, $x, $y, $z, $4, $5, $6, $8
>
> 411 Do not use subfields:
> $h, $i, $j, $v, $x, $y, $z, $4, $5, $6, $8
>
> 430 Do not use subfields:
> $h, $i, $v, $x, $y, $z, $4, $5, $6, $8
>
> 451 Do not use subfields:
> $i, $v, $x, $y, $z, $4, $5, $6, $8

While at first it might not be obvious why one would have a need for either $i or $4 in the 4XX fields,
there can be a relationship between a name recorded as a variant access point and the authorized
access point. While RDA does not currently have relationship designators for these kinds of
relationships, it does provide specific instructions for some kinds of these relationships. For example,
RDA 9.2.3 gives instructions for recording the following kinds of variant personal names: Real Name;
Secular Name; Name in Religion; Earlier Name of Person; Later Name of Person; Alternate Linguistic
Form of Name; Other Variant Name. Within the last two categories, further subtypes are shown in
examples, such as: Different Language Form; Different Script; Different Transliteration; Name as Saint;

Phrase Used to Name a Person; Full Form of Name Consisting of Initials. RDA 11.2.3 lists a number of kinds of variant corporate body names that may be recorded: Expanded Name; Acronym/Initialism/ Abbreviated Form; Alternative Linguistic Form of Name; Other Variant Name. Further subtypes of some of these are also enumerated in examples.

The American Library Association proposal 6JSC/ALA/43 to the Joint Steering Committee for RDA on *Revision and Expansion of RDA Appendix K: Relationship Designators: Relationships Between Persons, Families, and Corporate Bodies* (http://www.rda-jsc.org/sites/all/files/6JSC-ALA-43.pdf) proposed including a new section in Appendix K for relationship designators to relate different names of a person (page 14 of the proposal) and listed a number of other possible designators that could be used for different names of persons, families, and corporate bodies (page 21). Here is a sample:

earlier name A name that the person bore previous to assuming another name. *Reciprocal relationship:* later name

> **name before gender change** A name borne by the person previous to changing gender. *Reciprocal relationship:* name after gender change

> **name before marriage** A name borne by the person previous to marrying. *Reciprocal relationship:* name after marriage

Should the RSC choose to establish these kinds of relationship designators, they would be used in subfield $i of 4XX fields in LC/NACO authority records. They will also have URIs established when they are added as properties in the RDA Registry. A hypothetical simple example:

100 1# $a Clinton, Hillary Rodham
400 1# $4 [URI for relationship] $i Name before marriage: $a Rodham, Hillary Diane $w r

In the meantime, other agencies have already begun to create and use URIs for relationships in 4XX fields. For example, a URI in subfield $4 of 410 fields to identify that the form of name found in the field is an abbreviated form is already in use in the Gemeinsame Normdatei (GND), the authority file use by libraries in German-speaking Europe. In the GND ontology (https://d-nb.info/standards/elementset/gnd) these are called "Annotation Properties." A simple example:

110 2# $a Deutsche Nationalbibliothek
410 2# $4 http://d-nb.info/standards/elementset/gnd#abbreviatedNameForTheCorporateBody $i Abkuerzung $a DNB

Recommendations

We recommend that the *LC Guidelines Supplement* be revised to permit the inclusion of $0, $1, and $4 in authority fields in which they are authorized in MARC 21. Any technical problems that still exist that prevent the implementation of these subfields should be resolved. Instructions/best practices with ample examples should then be added to the DCM Z1 and/or LC-PCC Policy Statements as needed. The Standing Committee on Standards and Standing Committee on Training should be tasked with making the necessary changes to documentation and developing training materials for inclusion of $0, $1, and $4 in authority records.

Appendix 3

MARC Issues for Longer-Term Consideration

During its first two years the Task Group focused on MARC proposals and best practices that it judged would have the greatest impact in the current environment. However, in the course of its discussions the Task Group identified a number of areas where further work on the MARC format could significantly improve its support for URIs. It is unlikely that the Task Group will have sufficient capacity to pursue all of these issues in its remaining year, but we document them here for consideration by other stakeholders.

In deciding how far to take any further attempts to restructure the MARC format, costs will need be weighed against benefits. Some MARC enhancements (such as additional support for URIs in authorities) have the potential to offer worthwhile gains in the near future, while others (such as dealing with structurally more complex MARC data like 041 subfields or 880 paired fields) are arguably better addressed through other strategies, such as providing tables as an aid to conversion.

These are complex issues and they often raise questions that the MARC community has limited experience with. In developing its MARC discussion papers and proposals, the Task Group and its collaborators benefitted from extensive consultation with a wide range of stakeholders. The Task Group recommends that PCC look for ways to continue that process for future MARC proposals.

- *Authority format support for URIs.* Following the approval of the Task Group's proposal to redefine $0 and $1 to recognize the distinction between real world objects and authorities, a reappraisal of some provisions within the authority format may be advisable, particularly in the treatment of standard identifiers in 024. The Task Group also identified the authority 7XX block as holding promise for a more sophisticated treatment of mapping relationships among vocabularies, with $4 utilized for mapping predicates such as skos:closeMatch.

- *Providing for a way to indicate the subject of an RDF statement.* This presents major difficulties for the bibliographic format, since it is indeterminate which entity in the FRBR stack a bib record represents. Task Group members brainstormed some possible approaches to this problem but did not develop these ideas to sufficient maturity to present in a discussion paper.

- *Provisioning for URIs in additional MARC variable fields that currently carry a MARC code.* MARC Proposal 2017-01, developed by the British Library in partnership with the Task Group, established the principle that a URI can substitute for a MARC code in $4. The same principle could be extended to other MARC subfields. For example, $5 could in principle be populated with a URI rather than a MARC organization code. However, some of these fields present difficulties stemming from limitations present in the MARC definitions. The definition of $5 is itself ambiguous.

- *Multiple relationships in a single MARC field.* In some MARC fields, such as 041, subfield codes indicate a relationship to the resource being described (e.g. $h indicates "language code of original", $b indicates "Language code of summary or abstracts", etc.). If there is only one such subfield, it can be transformed in RDF to a predicate indicating a single relationship between the resource and the value(s) of the subfield; however, multiple relationships and object values indicated by multiple subfields in single MARC field creates a more complicated RDF conversion.

A conversion program would not be able to differentiate which object value belongs to each relationship predicate as it will not interpret any ordering of $0. This problem might be resolved by putting each subfield in a new MARC field or by using indicators for different relationships (this also would require parsing the subfields into multiple MARC fields) or doing nothing in MARC and relying on an RDF conversion program to parse the data into separate relationships and supply appropriate predicate and object URIs.

- *MARC 33X fields.* A special case of the MARC code issue is the MARC 336-338 block, where provision exists for a term, a code, *and* a URI. The term in $a and the code in $b are not necessarily derived from the same vocabulary source, but $2 for vocabulary source is not repeatable. In order to facilitate programmatic addition of $0 to the 33X fields, and subsequently facilitate MARC to RDF conversion, the Task Group would defer to the Standing Committee on Standards to provide best practice from the reports that the Task Group provided in April 15, 2017.[5] A longer-term solution would be for the maintenance agencies to align their vocabularies. To pursue this solution, however, is beyond the scope of the Task Group.

MARC data field	MARC subfields (singly or combined) equating to RDF Object
336 - Content Type (R)	a or b
337 - Media Type (R)	a or b
338 - Carrier Type (R)	a or b

- *Indicating source vocabulary.* The Task Group considers it to be a weakness of MARC that it provides for source vocabulary to be indicated in some fields (notably 6XX) but not others. The lack of provision for $2 in 1XX/7XX was much discussed within the Task Group but ultimately not pursued, in part because in principle an indication of source vocabulary is not strictly necessary if a dereferenceable URI is provided. Nevertheless, for some workflows a $2 in 1XX/7XX would be a distinct benefit.

- *880 parallel fields for vernacular scripts.* Knowing the script of a resource is an important factor in determining its usability for a given audience. The language of a resource can be determined in MARC from language codes. But this is not the same as being able to determine either the language or the script of data in a given field. MARC has very limited provisions for indicating script or language at the field level. Furthermore, while vernacular scripts are found in 880 fields, automated SPARQL queries currently rely on the transliterated Latin forms. The inability to easily and reliably identify the script in MARC data likewise inhibits the ability to assign the script correctly upon conversion to RDF, as seen in the examples below.[6]

---

[5] The Task Group on URIs in MARC. April 15, 2017 Report. Section II.c. MARC Objects/Work Reconciliation.
https://www.loc.gov/aba/pcc/bibframe/TaskGroups/PCC_URI_TG_20170415_Report.pdf#page=6
[6] LC has documented the issue here: https://github.com/lcnetdev/marc2bibframe2/issues/33

○ Authority data: http://id.loc.gov/authorities/names/n80150350

```
100 1_ |6880-01|a Xu, Zhimo,|d1896-1931.
880 1_ |6100-01/$1|a徐志摩,|d1896-1931.
```

100 1_ |6880-01|a Xu, Zhimo,|d1896-1931.$0http://id.loc.gov/authorities/names/n80150350
880 1_ |6100-01/$1|a徐志摩,|d1896-1931.$0http://id.loc.gov/authorities/names/n80150350

```
-<madsrdf:PersonalName rdf:about="http://id.loc.gov/authorities/names/n80150350">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Authority"/>
    <madsrdf:authoritativeLabel xml:lang="en">Xu, Zhimo, 1897-1931</madsrdf:authoritativeLabel>
  -<madsrdf:PersonalName>
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Variant"/>
    <madsrdf:variantLabel xml:lang="en">徐志摩, 1897-1931</madsrdf:variantLabel>
```

○ Bibliographic data: http://bibframe.org/resources/JRp1504786898/bibframe.rdf

```
-<bf:Person rdf:about="http://bibframe.org/resources/JRp1504786898/7259900person6">
    <bf:label>Aihara, Shigeru, 1948-</bf:label>
    <bf:authorizedAccessPoint>Aihara, Shigeru, 1948-</bf:authorizedAccessPoint>
    <bf:authorizedAccessPoint xml:lang="ja">相原茂, 1948-</bf:authorizedAccessPoint>
    <bf:hasAuthority rdf:resource="http://id.loc.gov/authorities/names/n82115867"/>
  </bf:Person>
```

● *Field for record graph URI.* UC Davis made a request to define a field to store a URI for the RDF graph representative of the entire MARC record. The Task Group discussed adapting field 884 for this purpose. However, as currently defined 884 records information about how data with a non-MARC origin was converted into MARC, not where to find data that was generated from the MARC record in hand. 884 seems therefore not to be the appropriate field to serve this purpose.