Program for Cooperative Cataloging
PCC BIBCO / CONSER / NACO / SACO

URI FAQs
PCC URI Task Group on URIs in MARC
September 26, 2018

1. What is a URI?

Wikipedia defines a uniform resource identifier (URI) as follows: A string of characters that identify a resource. A URI can be specified in the form of a URL or a URN. More information from Wikipedia: https://en.wikipedia.org/wiki/Uniform_Resource_Identifier.

W3C defines a URI simply as an ASCII string used to identify things on the Semantic Web.

For cataloging professionals, a URI tends to be an HTTP uniform resource identifier. A resource is not necessarily accessible via the Internet; e.g., human beings, corporations, buildings, works of art, rivers, and books in a library can all be resources. Abstract concepts can also be resources. Other terms used for resources are entity and thing.

An entity may have an identifier established in an authority database, such as the Library of Congress/NACO Authority File via the LC Linked Data Service (http://id.loc.gov), or by a service that creates identifiers such as Wikidata (http://wikidata.org) or BBC Things (http://www.bbc.co.uk/things/).

An identifier, constructed with a Web service protocol as a prefix, e.g. http://, is referred to as an HTTP URI. In the Resource Description Framework environment, an HTTP URI is a dereferenceable URI that facilitates operations from machine to machine. (http://www.bbc.co.uk/things/).

2. What forms can a URI take?

URIs can be classified either as Uniform Resource Locators (URLs) or Uniform Resource Names (URNs). In addition to identifying a resource, URLs provide a means of locating the resource by describing its primary access mechanism, e.g. **http://** or **ftp://** or **mailto:**, etc. URNs uniquely identify a resource, but do not necessarily specify its location or how to access it.

Each URI begins with a scheme name that refers to a specification for assigning identifiers within that scheme. There are dozens of schemes, but the most common for library applications are http, https, ftp, and mailto. The scheme name is always followed by a colon. Some URI schemes, such as http and ftp, are associated with network protocols.

Examples of URIs include:

http://isni.org/isni/0000000034980992
http://viaf.org/viaf/130909670
https://doi.org/10.1037/arc0000014

ftp://ftp.fao.org/docrep/fao/005/H5365E/H5365E00.pdf
mailto:John.Doe@example.com
telnet://192.0.2.16:80/
urn:oasis:names:specification:docbook:dtd:xml:4.1.2
urn:isbn:0-679-73669-7

3. What is an IRI?

IRI stands for Internationalized Resource Identifier; it is an extension of the URI scheme and is defined in RDF 3987 [1] .

Whilst URIs contain characters from a subset of the ASCII character set, IRIs may contain the full range of Unicode characters. IRIs are of benefit to institutions wishing to mint persistent identifiers in a variety of scripts. However, they are more susceptible to IDN homograph attack [2]. Additionally, support for IRIs among the various Semantic Web technology tools is still uneven. [3]

Further information about IRIs is available here: http://www.w3.org/TR/rdf11-concepts/#section-IRIs

[1] http://www.ietf.org/rfc/rfc3987.txt
[2] https://en.wikipedia.org/wiki/IDN_homograph_attack
[3] http://svn.aksw.org/papers/2010/ISWC_I18n/public.pdf

4. What is a real world object?

A real world object is an entity, such as a person, place, etc. It can be actual or conceptual. It is often referred to as a Thing.

W3C's document, Cool URIs https://www.w3.org/TR/cooluris/, states a convention for distinguishing a Thing and documents about the Thing (e.g. a webpage or authority data).

5. Which URIs apply to linked data?

When considering the use of URIs in the context of linked data, the question isn't so much what URIs apply to linked data, but what is the function of a URI in the context of linked data. Very broadly, URIs in a linked data context help to establish knowledge about an object. This

knowledge may be in the form of relationships, or concepts, or facts -- though, the most important aspect of the URI is that this information is developed to be consumed by machines, not people. One of the common misconceptions when assigning data to a $0 of a MARC record is that any URI that points to information about the object is a valid one; that this field enables catalogers to provide users with more information about a term, a concept, a person, etc. And indirectly, it does... but not in a way that the public directly consumes. Linked Data URIs create the bridges that allow systems to share and understand information, and in this context, only URIs that point to machine actionable data should be utilized.

6. Why when I use a URI in a browser, does it send me to a different link?

This is called dereferencing, i.e. retrieving a representation of the resource identified by the URI. If the semantic web data is published according to best linked data practices, the URI identifying the Thing is different from the URI identifying the Web document describing the Thing.

For example, http://sws.geonames.org/6252001/ identifies the United States; once in a browser, this redirects to http://www.geonames.org/6252001/united-states.html, the URL for the Web document describing the United States.

For this reason, you should not assume that the URL you see in the browser address window is the one that you should use in your bibliographic or authority record.

For further information on dereferencing see the W3C document Cool URIs for the Semantic Web: http://www.w3.org/TR/cooluris/

7. What if I just want to add a link to a web page (e.g., an author's website)?

To point to a web address (as opposed to an identifier) that provides further information about an entity, such as the website of an author, use the 856 field with first indicator 4 (for http address), and second indicator 2 (for related resource). The URL goes in the $u. Subfield $3 or $z may be used to describe the resource being pointed to. Examples:

856 42 $3 Author's website $u http://stephenking.com/
856 42 $u http://margaretatwood.ca/ $z Connect to author's website

8. Will my ILS accept URIs?

In many cases ILS will accept URI, but it would be prudent to exercise caution and carefully work through configuration and testing. In some cases ILS may make little or no use of URI and in

others their use may cause issues for display and/or indexing of headings. Impact on other services through cataloguing workflow should also be considered. For instance, at present records with a $1 cannot be uploaded to OCLC since it is not yet configured as a valid field in their systems.

9. Do any vendors provide URIs?

Vendors such as MARCIVE and Backstage Library Works provide URI for MARC bibliographic and authority data alongside work on authorities. Casalini is also working on the provision of URI for MARC data through its SHARE VDE project.

10. What MARC fields and subfields can URIs be added to?

Subfields $0 and $1 are established in numerous fields in all of the MARC formats. $0 contains an "Authority record control number or standard number." This number may be a URI. The recently approved $1 has been designed to hold the URIs of RWOs (Real World Objects). Please see question 11 for further information about the difference between URIs in $0 and $1. The relator code in $4 was redefined to host a URI for relationship.

These subfields should not be confused with $u (Uniform Resource Identifier), which should only be used to record document web addresses or URLs. Subfield $4 in numerous fields of the bibliographic and authority formats can hold a URI for relationships between agents and works, expressions, manifestations, and items or for relationships between works, expressions, manifestations, and items. In the authority format field 024 can also hold URIs.

Some examples:

Bibliographic Format

100 1# $a Stravinsky, Igor, $d 1882-1971, $e composer. $4 http://id.loc.gov/vocabulary/relators/cmp $0 http://id.loc.gov/authorities/names/n79070061

257 ## $a Korea (South) $2 naf $0 http://id.loc.gov/authorities/names/n79126802 $1 http://vocab.getty.edu/tgn/7000299-place

336 ## $a text $b txt $2 rdacontent $0 http://id.loc.gov/vocabulary/contentTypes/txt

336 ## $a text $2 rdaco $0 http://rdaregistry.info/termList/RDAContentType/1020

370 ## $g London (England) $2 naf $1 http://www.wikidata.org/entity/Q84

380 ## $a Novels $2 lcgft $0 http://id.loc.gov/authorities/genreForms/gf2015026020 $1 http://www.wikidata.org/entity/Q8261

385 ## $a Children $2 lcdgt $0 http://id.loc.gov/authorities/demographicTerms/dg2015060010

610 20 $a Harvard University $x Students $v Yearbooks. $0 http://id.loc.gov/authorities/subjects/sh85059205

650 12 $a Arthritis $x diagnosis. $1 http://id.nlm.nih.gov/mesh/D001168Q000175

655 #7 $a Picture books. $2 lcgft $0 http://id.loc.gov/authorities/genreForms/gf2016026096 $1 http://dbpedia.org/resource/Picture_book

700 1# $4 http://rdaregistry.info/Elements/w/P10129 $i Motion picture adaptation of (work): $a Austen, Jane, $d 1775-1817. $t Lady Susan. $1 http://viaf.org/viaf/183486135

711 2# $a Olympic Winter Games $n (21st : $d 2010 : $c Vancouver, B.C.) $0 http://id.loc.gov/authorities/names/n2006017550 $1 http://dbpedia.org/resource/2010_Winter_Olympics

Authority Format

024 7# $a http://isni.org/isni/0000000122802598 2 uri

024 7# $a http://id.worldcat.org/fast/1789938 $2 uri

024 7# $a http://www.wikidata.org/entity/Q913 $2 uri

370 ## $e Zagreb (Croatia) $2 naf $1 http://sws.geonames.org/3186886/ $1 http://vocab.getty.edu/tgn/7015558-place

372 ## $a Figure skating $2 lcsh $0 http://id.loc.gov/authorities/subjects/sh2005002252 $1 http://dbpedia.org/resource/Figure_skating

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

377 ## $a fre $0 http://id.loc.gov/vocabulary/languages/fre

380 ## $a Novels $2 lcgft $0 http://id.loc.gov/authorities/genreForms/gf2015026020 $1 http://www.wikidata.org/entity/Q8261

500 1# $4 http://rdaregistry.info/Elements/w/P10203 $i Screenwriter: $a Kushner, Tony $w r $0 https://www.idref.fr/034453245 $1 http://dbpedia.org/resource/Tony_Kushner $1 http://www.wikidata.org/entity/Q704433 $1 http://www.bbc.co.uk/things/68b04078-0443-4a61-96f2-4bdab1cdc163

530 #0 $4 http://rdaregistry.info/Elements/w/P10226 $i Continuation of (work): $a Environmental and natural resources series $w r $0 http://id.loc.gov/authorities/names/no2003004696

11. What is the relationship of a URI in $0 to its MARC field and component subfields?

The subfields that correspond to the object designated by the URI in $0 (or $1) vary from one field to another. Because of its complex history MARC is simply not consistent about this. The PCC URI group is drafting a set of tables to spell out the significant subfields for each of the more commonly used MARC fields. (*See* Task Group's April 15, 2017 report, page 5: https://www.loc.gov/aba/pcc/bibframe/TaskGroups/PCC_URI_TG_20170415_Report.pdf#page=5)

12. What is the difference between URIs in $0 and $1?

Differences in Definition:

$0 reflects the library community's longstanding commitment to controlled headings and the sources that have established them, while $1 points to factual descriptions of entities.

• According to Library of Congress documentation, $0 contains "the system control number of the related authority or classification record, or a standard identifier such as an International Standard Name Identifier (ISNI)." The control number can appear as an identifier, or as a token in a URI that resolves to a description, whose purpose is to cite the source where an authoritative heading used elsewhere in the field has been established. The description accesses features information about the heading, such as its provenance, revision history, or representations in multiple languages or scripts.

• The newly defined $1 is defined as a place for catalogers or automated processes to insert URIs that identify real-world objects that the field is about and resolve to machine-

understandable RDF descriptions, such as persons, places, and organizations that have names such as New York, Albert Einstein, or Microsoft. The descriptions feature biographical details, geospatial coordinates, domains of influence, and photos or other images. $1 defines a class of URIs that are specialized for linked data applications such as clustering, fact extraction, disambiguation, and identity resolution.

Differences in Usage:
In linked-data terms, $1 is an open-world solution to identity management, while $0 is primarily about the library community's management of headings and is relatively closed.

• $0 contains a pointer to the source of authority control for a heading, while $1 points to the real-world-object described in the field. Otherwise, there is no special dependency between $1 and other subfields in the same field. In particular, $1 is not defined as a source for an authority-controlled heading.

• $0 typically contains information published or endorsed by standards bodies in the library community, while the contents of $1 carry no such presumption. When adding a URI to $1, the cataloger is adding a crucial source of identifying information that may come from a library-community resource such as VIAF, or a third-party resource known only to specialized domains, such as performing arts or scientific sub-specialties.

Differences in structure:
$0 contains a mixture of legacy and semantic-web encodings, while $1 contains URIs that have been formulated according to linked-data conventions.

• The $0 may either contain a control number, or a URI containing the control number as a token. For example, a $a field containing the string "Lennon, John 1940-1980" may contain a $0 subfield with the LCNAF control number n 80017868. Alternatively, the $0 may contain the URI http://id.loc.gov/authorities/names/n80017868. Both point to essentially the same information in a variety of forms, such as a human-readable HTML page or a machine-understandable RDF encoding. Of course, the identifier predates the semantic web, so it may also identify a unique record in a database or paper copy of the authority file.

• The $1 contains a URI that conforms to linked data conventions. It is globally unique, persistent, and resolves to an RDF-encoded description of a real-world object. Examples include http://www.wikidata.org/entity/Q1203 and http://viaf.org/viaf/196844.

• In theory, the differences between $0 and $1 URIs are detectable by URI validators such as Vapour. But in practice, automatic detection is challenging because the Web protocols and the implementation of URIs have changed over time. To address this problem, the PCC-URI task group has published the Formulating URIs document cited above, which identifies the relevant

syntax patterns for resources most likely to be consulted by the library community. Longer term, we anticipate that improved data models and software tools will automate much of the task of constructing the appropriate URIs for a given MARC or other resource-description context.

• Link to Vapour: http://linkeddata.uriburner.com:8000/

13. Why are skos:Concepts not considered Real World Objects (RWOs) with respect to $0 and $1?

Simple Knowledge Organization System (SKOS) "is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web." https://www.w3.org/2004/02/skos/intro. skos:Concepts, the central class in SKOS, are used to build entries within a particular Knowledge Organization Scheme. The concept works as a proxy for a thing in the real world, and it can have statements about it that do not apply to the RWO, e.g. versioning information for the term, or what scheme the concept is in - neither of which is true about the RWO.

See also section 3.1 *Mapping Concept Schemes* in the SKOS Primer (ttps://www.w3.org/TR/skos-primer/), specifically the language on skos:exactMatch and owl:sameAs. The semantics of skos:Concept are that they exist within a particular vocabulary, and they have assertions within that particular vocabulary. We would not say that two skos:Concepts are the owl:sameAs each other. They are not, in the same way that the skos:Concept http://rdaregistry.info/termList/RDAColourContent/1002 is not owl:sameAs http://vocab.getty.edu/aat/300137660. They may be a skos:exactMatch or skos:closeMatch, or they might have the same foaf:focus, but they are not themselves the same thing.

14. Shouldn't the URIs in $0 and $1 be coordinated? Doesn't that create extra work?

One comment on the MARC Proposal 2017-08 expressed concern that $1 introduces maintenance problems because it must be kept in sync with $0. But this is primarily a consequence of the Library of Congress implementation of 'Authority' and 'RWO' URIs, which are derived by partitioning a source authority record into two sets of statements that must be reassembled in some circumstances. But when the contents of $1 is a URI maintained outside the library community, there is no formal dependency between the data models of the library authority file mentioned in $0 and the $1 resource.

15. Which URI sources should I use in my cataloging?

Please watch for an upcoming best practice document for common URI vocabularies in MARC.

16. Is there a limit to the number of URIs I can use in one field?

Strictly speaking the MARC definitions place no constraints on the number of URIs or their source. But using URIs from different sources creates conceptual and practical problems it would be best to avoid. Best practices will be provided by PCC during 2018.

17. Can I put URIs in name authority records?

Currently URIs can be given in note fields (e.g. 670) and in the 024 field. See the NACO 024 Best Practices Guidelines.

URIs cannot as yet be given elsewhere in the authority record, for example not in 5XX or 3XX fields. The PCC URIs in MARC Pilot is exploring and identifying best practices for the recording of URIs in other MARC authority fields.

18. Can I put URIs in bibliographic records in Connexion?

URIs can be added to bibliographic records in Connexion, but there are issues surrounding it, e.g., how the use of URIs relates to functionality for controlling headings. OCLC's handling of $0 and $1 for controlled headings is described below.

OCLC-MARC Format Update 2017, described in Technical Bulletin 267 (https://www.oclc.org/support/services/worldcat/documentation/tb/267.en.html), includes the redefinition of subfields $0 and $4 to include URIs that are in the form of a Web retrieval protocol. Such URIs may now be included in any bibliographic field for which $0 or $4 is authorized.

In OCLC-MARC Format Update 2018, described in Technical Bulletin 268 (https://help.oclc.org/WorldCat/Cataloging_documentation/Technical_Bulletins/268) OCLC announced the availability of $1 to accommodate RWO URIs.

Note that in fields that can be controlled by OCLC Connexion software (1XX, 6XX, 7XX, 8XX), $0s is removed when the field is controlled. However, $1 is retained.

There is at present no simple automated way within OCLC Connexion to add URIs, so care must be taken when they are added by hand. Examples of some OCLC bibliographic records that include URIs: #742510466, #820632069, #870305395, #924738796, #992709092, #1004632218.

Examples of some fields with URIs:

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

257 ## United States $a Great Britain $2 naf $0 http://id.loc.gov/authorities/names/n78095330 $0 http://id.loc.gov/authorities/names/n79023147

336 ## text $b txt $2 rdacontent $0 http://id.loc.gov/vocabulary/contentTypes/txt
337 ## computer $b c $2 rdamedia $0 http://id.loc.gov/vocabulary/mediaTypes/c
338 ## online resource $b cr $2 rdacarrier $0 http://id.loc.gov/vocabulary/carriers/cr

344 ## digital $2 rdatr $0 http://rdaregistry.info/termList/typeRec/1002
344 ## $b optical $2 rdarm $0 http://rdaregistry.info/termList/recMedium/1003
344 ## $g surround $2 rdacpc $0 http://rdaregistry.info/termList/configPlayback/1004
346 ## Laser optical $2 rdavf $0 http://rdaregistry.info/termList/videoFormat/1009
346 ## $b NTSC $2 rdabs $0 http://rdaregistry.info/termList/broadcastStand/1002
347 ## video file $2 rdaft $0 http://rdaregistry.info/termList/fileType/1006
347 ## $e region 1 $2 rdare $0 http://rdaregistry.info/termList/RDARegionalEncoding/1002

347 ## text file $2 rdaft $0 http://rdaregistry.info/termList/fileType/1002

382 01 piano $0 http://id.loc.gov/authorities/performanceMediums/mp2013015550 $n 1 $s 1 $2 lcmpt

386 ## $4 http://id.loc.gov/vocabulary/relators/fmd $i Film director: $a Mexicans $2 lcdgt $0 http://id.loc.gov/authorities/demographicTerms/dg2015060329
386 ## $4 http://id.loc.gov/vocabulary/relators/fmd $i Film director: $a Men $2 lcdgt $0 http://id.loc.gov/authorities/demographicTerms/dg2015060359

700 1# Cuarón, Alfonso, $e film director, $e screenwriter, $e film producer, $e film editor. $4 http://id.loc.gov/vocabulary/relators/fmd $4 http://id.loc.gov/vocabulary/relators/aus $4 http://id.loc.gov/vocabulary/relators/fmp $4 http://id.loc.gov/vocabulary/relators/flm
700 1# Cuarón, Jonás, $e screenwriter. $4 http://id.loc.gov/vocabulary/relators/aus
700 1# Heyman, David, $d 1961- $e film producer. $4 http://id.loc.gov/vocabulary/relators/fmp

710 2# IFLA FRBR Review Group. $b Consolidation Editorial Group, $e editor. $4 http://rdaregistry.info/Elements/e/P20048 $4 http://id.loc.gov/vocabulary/relators/edt

780 00 $4 http://rdaregistry.info/Elements/w/P10226 $t Environment (Rivonia, South Africa) $x 2219-8199 $w (DLC) 2010252067 $w (OCoLC)530178144

19. If I have the choice, is it preferable to put URIs in bibliographic or authority records?

URIs can have value in both bibliographic and authority records. There are some kinds of data - and not only URIs - that can more logically and non-redundantly be provided in authorities, but the fact is that we operate in a mixed environment where a clean separation is not made. Currently URIs are for the most part not approved for use in NACO authorities. The Task Group hopes to address this restriction.

20. When can/should I use the new field 758?

The 758 field became part of the MARC specification in December 2017 and was implemented by OCLC in September 2018. PCC has not yet issued best practices for use of this field; these can be expected in late 2018 or early 2019.

21. Why were linking entry fields (76X-78X) not included in the task force proposals?

76X-78X linking entry fields tend to be associated with instance or manifestation data. While there is no reason instances or manifestations should not have RDF representations that could be linked in $0 (or $1), it is not clear that stable sources exist yet for these data. The task group therefore gave enhancements to these fields a lower priority than its other proposals. In addition, the 758 field has been defined to accommodate instance entities and predicates if required. Although the task group expects to give less emphasis to MARC proposals in its third year, it is open to use cases that may justify proposals affecting the 76X-78X fields.

22. How to formulate and obtain a linked data URI for a resource?

Ideally it is best to acquire URIs through automated processes such as SPARQL queries or via lookup tools built into metadata editors. This is not always possible at present. Catalogers wanting to add properly formed and coded URIs to their records should consult the PCC Formulating URIs document (https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf).

23. Are validators available for dereferenceable URIs?

There are validators which check whether semantic web data is correctly published according to current best linked data practices; in particular they check whether the URI tested identifies an entity, i.e. a RWO or a Web document describing the entity. At the time of writing, we are aware of the following validators: the Vapour validator http://linkeddata.uriburner.com:8000/vapour and Vafu http://vafu.redlink.io/

24. What tools are available?

Libraries adding URIs to their catalogs have used tools such as MARCEdit MARCNext, LOD/OpenRefine, and custom scripts (utilizing SPARQL), as well as working directly in SPARQL for querying endpoints to enrich data with a URI (or an IRI).

25. Where can I find training resources on URIs and linked data?

The Linked Data Exploratorium (http://explore.dublincore.net/explore-learning-resources-by-competency/) contains a great number of useful training resources related to linked data in general and URIs specifically.

26. Are RDF URIs sensitive to use of http versus https?

The Web community has made a push toward more secure delivery of Web documents in the last decade using the HTTPS protocol. For human readability an RDF URI may resolve or redirect to a web page/document displaying information about the resource it identifies, but the RDF URI itself does not represent that web page/document. For example, an RDF URI from Wikidata, e.g. http://www.wikidata.org/entity/Q36322 may trigger an entity 303 redirect from server to various outputs:

1) A generic document after machine content-negotiation (can default to json syntax)
https://www.wikidata.org/wiki/Special:EntityData/Q36322
2) An RDF turtle
https://www.wikidata.org/wiki/Special:EntityData/Q36322.ttl
3) An HTML
https://www.wikidata.org/wiki/Q36322

Notice the returned URIs are all in secure protocol format, https://, because they represent web

documents, while the RDF URI is http://. Additionally, the paths in respective URIs are not exactly the same. These differences are subtle.

Since an RDF URI is not a Web address, but rather an identifier, it should not need to change from HTTP protocol to HTTPS. Ideally, an RDF URI published as http://www.wikidata.org/entity/Q36322 should not be re-used or re-stated elsewhere as https://www.wikidata.org/entity/Q36322. That said, whether or not it makes a difference to use http: or https: in an RDF URI may ultimately depend on the host server. The host server may be set up to seamlessly resolve http: to https: and vice versa, in which case it may not make a difference in how it resolves, but it may make a difference to a SPARQL query. In addition, if the server is not set up to resolve one to the other, then using http: or https: will make a difference in how the RDF URI resolves, as well as to machine querying. Therefore, it's best to re-state an RDF URI exactly as it is published by the host and for the host not to change RDF URIs from http: to https:.

27. Is permalink different from a canonical URI?

A persistent URL that takes a user to a Web document is called a permalink. A host may declare a URI to be canonical that is, the URI preferred by the host and tagged as canonical for content negotiation. Canonicalization of a URI by a host allows content negotiation between machines and search engine optimization (SEO) to index the link preferred by the host for displaying Web content.