Namespace Registration for Metadata Identifiers (META)

Namespace ID: META Requested of IANA.

Version: 1

Date: 2020-03-10

Registrant:

Name: Juha Hakala

E-mail: juha.hakala@helsinki.fi

Affiliation: Senior adviser, The National Library of Finland

Address: P.O.Box 15, 00014 Helsinki University, Finland.

Web URL: https://www.kansalliskirjasto.fi/en/

Background:

According to ISO 5127, metadata is data about other data, documents or records that describes their content, context, structure, data format, provenance and/or rights attached to them. Metadata elements and their machine readable codes are specified in (cataloguing) formats. Libraries have been using a metadata format (Machine Readable Cataloguing, MARC) since late 1960s, and starting in 1990s, museums, archives and many other organizations have developed their own metadata formats. There are also metadata formats for administrative metadata (rights, long-term preservation). As a result of this format proliferation and increased availability of high-quality structured metadata as open linked data, it is important to enable human and machine users to make sense of metadata. This requires at least basic understanding of metadata elements in these formats, such as Dublin Core Title or MARC 21 tag 245, Title statement.

Metadata format specifications are usually available in the Web for free. For instance, the Library of Congress has published MARC 21 formats at <https://www.loc.gov/marc/>. The National Library of Finland maintains translations of these formats at (<https://marc21.kansalliskirjasto.fi/>). A list of MARC 21 translations (n = 20) is available at <https://www.loc.gov/marc/translations.html>.

From users’ point of view, using MARC 21 documentation is not as easy as it should be. Information may have been published in a form (e.g. a PDF document) which is not sufficiently machine understandable. And even if information about each metadata element is available in HTML, there are no explicit links between translations. There is no connection between the page describing Title statement in English (<https://www.loc.gov/marc/bibliographic/bd245.html>), the page describing it in Finnish (<https://marc21.kansalliskirjasto.fi/bib/20X-24X.htm#245> and the Swedish version (<https://www.kb.se/katalogisering/Formathandboken/Bibliografiska-formatet/210-249/#245>).

Most metadata format communities have chosen to use location dependent HTTP URLs as identifiers for their metadata formats and their elements. The only exception is the Dublin Core Metadata Initiative, which chose Persistent URLs[[1]](#footnote-1) (PURLs) as identifiers[[2]](#footnote-2). For instance, DC metadata element Creator has two Persistent URLs;

<http://purl.org/dc/terms/creator>

in the /terms/ namespace, and

<http://purl.org/dc/terms/1.1/creator>

in the /elements/1.1/ namespace, because these two creators have different semantics.

NOTE These PURLs identify creator-related metadata elements in the Dublin Core format, not the creators themselves. Creators have other identifiers, such as ORCiDs and ISNIs, which may be expressed as URIs (e.g. <http://www.isni.org/isni/0000000124292982>).

French translation of DCMI Metadata Terms[[3]](#footnote-3) uses the same PURLs, but they resolve to English texts. In order to access French element descriptions, it is necessary to use their URLs, such as <http://www.yoyodesign.org/doc/dcmi/dcmi-terms/index.html#terms-creator> for creator in the Terms namespace. Translations to e.g. Czech, Japanese and Italian are similar in this respect: the identifier of the element is the PURL of the English version (which is OK), but the PURL does not enable the user to retrieve documentation about the element in the appropriate language.

From users’ point of view it is confusing that in MARC 21 and all other formats which rely on HTTP URLs, each translation of the format has its own set of identifiers. In Dublin Core the same identifier is used for several translations, but the identifier provides linking only to the version in English. Therefore users who want to understand the metadata they see, may have difficulties finding this information in an appropriate language.

NOTE Some XML-based metadata formats have XML namespaces:

AudioMD: <http://www.loc.gov/audioMD/>

MARCXML: <http://www.loc.gov/MARC21/slim>

PREMIS: <http://www.loc.gov/premis/rdf/v3/>

VideoMD: <http://www.loc.gov/videoMD/>

Unfortunately namespaces for AudioMD, MARCXML and VideoMD are not resolvable using e.g. HTTP, and functionality supported by the PREMIS namespace is insufficient - links to individual tags such as

|  |  |
| --- | --- |
|  | http://www.loc.gov/premis/rdf/v3/Copyright |

do not provide useful results, since they all take the client only to the beginning of the file containing all element descriptions. When implemented in this manner links to descriptions of metadata elements do not help human and machine users of metadata to understand the metadata provided.

Purpose:

Provide a uniform basis and a tool for identification of elements in metadata formats.

Functional improvements to the current URL-based identifiers of metadata elements.

Benefits:

1. Linking to alternative versions (e.g. full / concise, human readable / machine readable) of element descriptions with one URN.

1. Automatic selection of the appropriate language (see below). For instance, a bibliographic record containing a link to the URI explaining MARC tag 245 in Finnish is only useful for users who understand Finnish, because the HTTP server holding these pages cannot redirect Swedish / English speaking users to MARC tag 245 pages on other HTTP servers that would be appropriate for them.
2. Having a URN namespace dedicated for metadata elements may improve co-ordination on how identifiers for metadata elements are created and what they resolve to. It may also encourage the organizations maintaining metadata formats to provide documentation separately for each metadata element and in a form more suitable to the Web than e.g. PDF.
3. A resolution service for URN:META identifiers can direct users to information in their preferred languages if the resolution method communicates their language preferences to the resolver.

If a URN:META identifier is assigned to the Library of Congress description of MARC 245 tag[[4]](#footnote-4), it is possible for a resolution service to direct a user to the descriptions of this tag in multiple languages[[5]](#footnote-5), depending on the language settings of the user’s Web client. This functionality can be implemented by supporting the HTTP Accept-Language header in the URN resolver. A prerequisite for this functionality is that all element URLs from translated versions are harvested to the resolver’s URN – URL mapping table. Since these URLs should be stable, keeping the links up to date is feasible.

If a network protocol used does not support language negotiation the required functionality may also be implemented with the URN R-component.

NOTE A URN should not be assigned if an element already has a well-managed persistent identifier such as DOI (<https://en.wikipedia.org/wiki/Digital_object_identifier>).

Syntax:

The Namespace-Specific String (NSS) consists of three parts:

o a prefix consisting of a code identifying the metadata format and

optional sub-namespace code(s) separated by a colon(s);

o a hyphen (-) as the delimiting character; and,

* a string assigned under the auspices of the format maintenance agency.

These strings may be constructed according to the local preferences as long as they are aligned with the requirements of RFC 3986 and RFC 8141.

Format maintenance agency is the organization maintaining the original version of the format, such as Dublin Core Metadata Initiative for Dublin Core, or the Library of Congress for MARC 21 formats. A maintenance agency shall specify the NSS syntax for its formats, and it may outsource the URN assignment and maintenance of URN resolver to a third party, such as an organization maintaining a translation of the format.

The following formal definition uses ABNF [[RFC5234](https://tools.ietf.org/html/rfc5234)].

meta-nss = prefix ":" meta-string

prefix = format-code \*( "-" subspc )

; The entire prefix is case insensitive.

format-code = 1\*(ALPHA / DIGIT)

; As assigned by the National Library of Finland

; (identifies the metadata format and the maintenance agency to which the branch is delegated).

subspc = 1\*(ALPHA / DIGIT)

; As assigned by the respective format maintenance agency.

meta-string = path-rootless

; The "path-rootless" rule is defined in [RFC 3986](https://tools.ietf.org/html/rfc3986).

; Syntax requirements specified in [RFC 8141](https://tools.ietf.org/html/rfc8141) MUST be

; taken into account.

The meta-string is case-sensitive unless specified as case-insensitive by the maintenance agency.

The following metadata format codes SHALL be used:

Descriptive metadata

Code Format(s) URL

BF BIBFRAME <http://www.loc.gov/bibframe/>

danMARC2 DANMARC2 <http://www.kat-format.dk/danMARC2/>

DW Darwin Core <https://dwc.tdwg.org/terms/>

DC Dublin Core <https://www.dublincore.org/specifications/dublin-core/>

DDI DDI <https://ddialliance.org/explore-documentation>

EAD EAD <https://www.loc.gov/ead/>

FINMARC FINMARC <https://www.kiwi.fi/display/Marc21/FINMARC>

IMARC INTERMARC <https://www.bnf.fr/fr/intermarc-bibliographique-de-diffusion>

LIDO LIDO <http://network.icom.museum/cidoc/working-groups/lido/>

MARC MARC 21 <https://www.loc.gov/marc/marcdocz.html>

MARCXML MARCXML <http://www.loc.gov/standards/marcxml/>

MIX MIX <http://www.loc.gov/standards/mix/>

MODS MODS <http://www.loc.gov/standards/mods/>

ONIX ONIX <https://www.editeur.org/8/ONIX/>

UKMARC UKMARC <https://www.webarchive.org.uk/wayback/archive/20160000000000/http://www.bl.uk/bibliographic/ukmarc.html>

UNIMARC UNIMARC <https://www.ifla.org/unimarc>

Administrative metadata

Code Format URL

PREMIS PREMIS <http://www.loc.gov/standards/premis/>

TEXTMD textMD <https://www.loc.gov/standards/textMD/>

AUDIOMD audioMD <https://www.loc.gov/standards/amdvmd/>

VIDEOMD videoMD <https://www.loc.gov/standards/amdvmd/>

Cataloguing rules

Code Rules URL

ISBD ISBD <http://iflastandards.info/ns/isbd/elements/>

RDA RDA <https://www.rdaregistry.info>

One code may cover an entire family of formats (e.g. MARC Authority, Bibliographic and Holdings formats). Sub-namespaces may be used to differentiate formats within these format families if necessary.

Since all prefixes with the same format-code are delegated to the same maintenance agency, such families are perforce maintained by the same agency.

National translations of metadata standards and cataloguing rules shall use the codes and URNs of the original specifications. Thus the Finnish translation of MARC 21 shall use the prefix MARC of MARC 21 and URNs assigned to the elements of the English version of the format. If resolution is done via HTTP, the URN resolver can use HTTP language negotiation to direct the client to the correct language version. Other network protocols may support similar functionality in the future.

National variants of metadata formats (e.g. historical FINMARC format, which was based on equally outdated USMARC) shall have their own format codes, since their tags and semantics may differ from the original ones. For instance, UKMARC tag 245 is not the same as USMARC tag 245, since in the former subtitle had its own tag, 248, whereas in the latter, subtitle was included in tag 245.

Metadata application profiles[[6]](#footnote-6) such as Darwin Core, an extension of Dublin Core intended for sharing of information about biological diversity, may have codes of their own if they a) extend the base format substantially and b) are well documented and stable.

The structure (if any) of the meta-string is determined by the authority for the prefix. Within the meta-string, it is recommended that a hyphen is used for separating different sections of the identifier from one another in order to improve the human readability of the string.

Maintenance agencies SHOULD NOT use in meta-strings characters requiring percent-encoding.

Registering format codes:

New codes will be added by the National Library of Finland on request.

Requests should be sent to [meta-request@helsinki.fi](mailto:meta-request@helsinki.fi).

The list of registered format codes will be maintained as a part of this document.

NOTE Formats included above are the ones most commonly used by libraries, archives, museums and in publishing. It is anticipated that the need for adding new formats will not be frequent.

Rules for lexical equivalence:

Whereas the prefix is case insensitive, meta-strings MAY be case sensitive at the preference of the assigning authority; parsers therefore SHALL treat these as case sensitive, and any case mapping needed to introduce case insensitivity is the responsibility of the relevant resolution system.

Case insensitivity of the prefix must be taken into account when URN:META identifiers are compared.

META assignment:

National and international metadata format maintenance agencies may use URN META when they want to assign persistent identifiers for the metadata elements and tags of their formats, and provide URN-based access to machine or human readable descriptions about these metadata elements. For the time being these descriptions are unstructured text on Web pages.

The URN assigned to the element shall not change even if the description of the element is changed. URNs assigned to deleted elements shall not be re-used.

Metadata format maintenance agencies shall have procedures in place to make sure that the assigned URNs are unique and persistent. Since the number of metadata elements on formats is relatively low (at most a few hundreds) such procedures can be simple (e.g. URN can be based on the name of the element).

Security and Privacy:

URN:META identifiers do not have any known security or privacy issues. They are intended to have publicly-known meanings and do not refer to specific individuals, groups, or organizations.

Interoperability:

URN:META identifiers do not have any known interoperability related issues.

Resolution:

General

URNs in the URN:META namespace MUST be resolvable.

Each registrant of a format-code MUST register a base URL (ending in "/") for a resolution service for URN:METAs that have that format-code. A registrant may provide additional base URLs for prefixes composed of that format-code and one or more following subspc’s. The base URL for resolving a particular URN:META is the base URL for the longest registered prefix which is an initial part of the URN's prefix.

A URN:META is resolved by composing a URL by concatenating the base URL with the URN:META and fetching that URL using the normal HTTP/HTTPS GET method. The retrieved resource SHOULD describe the identified metadata element and MAY provide or reference further information about it.

Resolution services MAY respond to GET requests with a redirection response whose Location header field is a URL of a preexisting description of the element. If information about the element is available in multiple languages, a resolution service SHOULD use the HTTP Accept-Language header to select a URL of a resource in the user's requested language.

URN to URL resolution service from the URN to the URL of the page describing the identified metadata element MUST be supported. Other resolution services such as a link to appropriate cataloguing rule may be provided if appropriate.

Example 1

Namespace URN:META:MARC: contains URNs which identify the tags of MARC 21 metadata formats (https://www.loc.gov/marc/) in various languages. Unlike current HTTP URLs, all language versions share the same identifier.

In this example, URNs are expressed as HTTP URIs which use (non-existent) URN resolver located http://example.com.

It is assumed that the language setting of the client is English (and as a default) URNs [http://example.com/urn:meta:marc-<nss](http://example.com/urn:meta:marc-%3cnss)> will be resolved at MARC 21 format specific pages at directories

<https://www.loc.gov/marc/bibliographic/>

<https://www.loc.gov/marc/authority/>

<https://www.loc.gov/marc/holdings/>

on the Library of Congress site.

Case 1

URN of the MARC Bibliographic format tag 245 (Title Statement)

Assuming that the registered resolver base URL for urn:meta:marc is <http://example.com/>, urn:meta:marc-bd245 would be resolved by fetching

<http://example.com/urn:meta:marc-bd245>

which (absent an Accept-Language header in the GET request) would redirect either to

<https://www.loc.gov/marc/bibliographic/bd245.html> (full description of the tag)

or to

<https://www.loc.gov/marc/bibliographic/concise/bd100.html> (concise description of the tag)

depending on the service requested. If full description is set as a default, concise description can be made available via an R-component based service request.

If the GET request contained "Accept-Language: fi" (preferring Finnish), it would always redirect to

<https://marc21.kansalliskirjasto.fi/bib/20X-24X.htm#245>

since there is only one Finnish translation, based on the concise version.

If the GET request contained "Accept-Language: sv" (preferring Swedish), it would always redirect to

<http://www.kb.se/katalogisering/Formathandboken/Bibliografiska-formatet/210-249/#245>

since there is only one Swedish translation, based on the concise version.

File names of pages describing tags in MARC 21 formats have the same syntax (MARC Bibliographic uses bdxxx.html, where xxx is the tag number), no URN – URL mapping table is required in the resolver. Target URLs can be generated from URNs programmatically. Since the target URLs have been stable for decades, the need to modify programmatic mapping should be very infrequent.

Case 2

URN of the MARC Authority format tag 100

Assuming that the registered resolver base URL for urn:meta:marc is <http://example.com/>, urn:meta:marc-ad100 would be resolved by fetching

<http://example.com/urn:meta:marc-ad100>

which (absent an Accept-Language header in the GET request) would redirect to

<https://www.loc.gov/marc/authority/ad100.html>

If the request contained "Accept-Language: fi" (preferring Finnish), it would redirect to

<https://marc21.kansalliskirjasto.fi/aukt/1XX.htm#100>

If the request contained "Accept-Language: sv" (preferring Swedish), it would redirect to

<http://www.kb.se/katalogisering/Formathandboken/Auktoritetsformatet/1XX/#100>

Example 2

URN of the Dublin Core Terms namespace property Title

Assuming that the registered resolver base URL for urn:meta:dc is <http://example.com/>, urn:meta:dc-terms:title would be resolved by fetching

<http://example.com/urn:meta:dc-terms:title>

which (absent an Accept-Language header in the GET request) would redirect to

<http://purl.org/dc/terms/title>

Persistence:

Persistence of URN:META resolution services depends on the persistence of metadata formats.

Metadata about deprecated MARC formats such as USMARC, UKMARC or FINMARC may no longer be available at all, and even if it is, its form and content may not be suitable for URN resolution. See for instance UKMARC documentation at

<https://www.webarchive.org.uk/wayback/archive/20160107133726/http://www.bl.uk/bibliographic/ukmarc.html>

or FINMARC documentation at

<https://www.kiwi.fi/display/Marc21/FINMARC>

Additional documentation / information:

None

Revision Information:

This registration request is applicable to metadata formats and cataloguing rules listed above.

This request has been updated 2020-03-10. The section describing resolution has been revised. Numerous smaller changes have been made, based on most helpful comments from Dale Worley.

1. <https://archive.org/services/purl/help> [↑](#footnote-ref-1)
2. DCMI Metadata Terms (in English) are available at

   <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> [↑](#footnote-ref-2)
3. <http://www.yoyodesign.org/doc/dcmi/dcmi-terms/index.html> [↑](#footnote-ref-3)
4. Such URN can be assigned by the Library of Congress or a third party. [↑](#footnote-ref-4)
5. Provided that they have URLs, which is not always the case. [↑](#footnote-ref-5)
6. https://en.wikipedia.org/wiki/Application\_profile [↑](#footnote-ref-6)