

Namespace Registration for Metadata Identifiers (META)

Namespace ID: META Requested of IANA.

Version: 1

Date: 2020-02-05

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/>

Requesting entity is an organization which maintains Finnish translations of MARC 21 and Dublin Core metadata formats, and would like to replace their current cool URI identifiers with more persistent and versatile URNs.

Purpose:

As a part of general drive towards open linked data, organizations maintaining metadata formats have assigned persistent identifiers (PIDs) or cool URIs to the elements (tags, fields, etc.) in these formats.

Metadata identifiers are format-specific identifier systems used for identification of metadata elements (tags, fields, etc.) in metadata formats and cataloguing rules.

For instance, Dublin Core community users Persistent URLs (PURLs) as identifiers. For example, DC metadata element Title has a Persistent URL (PURL)

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

in the /terms/ namespace, and PURL

<http://purl.org/dc/elements/1.1/title>

in the /elements/1.1/ namespace. Different PIDs are necessary because term definitions in /terms/ namespace may differ from those in /elements/1.1/ namespace.

PURL system has currently maintenance issues and DCMI community has concerns about its long term future. Therefore the community has been looking for alternative PID systems, and have asked if URNs could be used. META namespace will make this feasible. URN for Dublin Core Title element could be e.g. urn:meta:dc:terms:title.

URN:META identifiers can be used to improve the functionality of cool URIs. For instance, Library of Congress has assigned URI

<https://www.loc.gov/marc/bibliographic/bd245.html>

to the description of MARC 21 tag 245 (Title statement), but there is no way to use this same cool URI to access descriptions of this tag in other languages. For instance, URI of the Finnish description of MARC tag 245 is

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

and in Swedish

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

but these three URIs are not interlinked in any way.

If URN:META identifier is assigned to the Library of Congress description of MARC 245 tag, it is possible to direct users to the descriptions of this tag in other languages, depending on the language settings of the users' Web clients. This functionality can be implemented with HTTP Accept-Language header in the URN resolver. Alternatively R-component can be used in the future, if the network protocols used then do not support language negotiation).

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 description. When implemented in this manner providing links to descriptions of metadata elements does not enable human and machine users of metadata to understand the metadata provided.

URIs from registered XML namespaces may be used instead of PIDs like URNs if and when they exist and support element level linking.

A URN should not be assigned if an element already has a (well-managed) PID.

However, URNs should replace "cool" URIs, since using them as identifiers for metadata elements is not ideal. There are several reasons for this:

1. It is not possible to provide links to alternative versions (e.g. human readable / machine readable) versions of element descriptions with one URL.
2. It is not possible to select the appropriate language. 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.
3. There has been no co-ordination on how URIs for metadata elements have been created, so URI may identify an element, a subset of elements or even all elements in the format. In the latter two cases URI fragment is used

to indicate the relevant element, but such approach will no longer work if the document syntax changes so that fragment is no longer supported.

4. Some "cool URIs" have already proven that they are un-cool (e.g. the MARCXML namespace is dead).
5. Link does not resolve to the element description but to something more generic such as beginning of the file containing descriptions of all elements in the format.
6. Some metadata formats (e.g. UNIMARC) do not have element level URIs.

In addition to these generic problems, there are metadata format specific issues. PURL (Persistent URL, <https://archive.org/services/purl/>) which is used for identification of Dublin Core elements, has faced maintenance issues lately. Therefore Dublin Core Metadata Initiative is investigating other PID systems as a potential replacement for PURL.

Syntax:

The Namespace-Specific String (NSS) will consist 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 by the format maintenance agency. Such strings may be constructed according to the local preferences as long as they are aligned with the requirements of RFC 3986 and RFC 8141.

The following formal definition uses ABNF [[RFC5234](#)].

```
meta-nss      = prefix "-" nbn-string

prefix        = format-code *( ":" subspc )
               ; The entire prefix is case insensitive.

format-code   = 1*(ALPHA / DIGIT)
               ; As assigned by the respective format maintenance agency
               ; (identifies the metadata format 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.
               ; Syntax requirements specified in RFC 8141 MUST be
               ; taken into account.
```

The following metadata format codes SHALL be used:

Descriptive metadata

Code	Format(s)	URL
BF	BIBFRAME	http://www.loc.gov/bibframe/
DC	Dublin Core	https://www.dublincore.org/specifications/dublin-core/
DDI	DDI	https://ddialliance.org/explore-documentation
EAD	EAD	https://www.loc.gov/ead/

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/
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://iflstandards.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.

National translations shall use the codes of the original format. 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. HTTP language negotiation will be used by the URN resolver to direct the client to the correct language version.

National variants of metadata formats (e.g. historical FINMARC format, which was based on equally outdated USMARC) should have their own format codes.

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

The structure (if any) of the meta-string is determined by the authority for the prefix. Whereas the prefix is case insensitive, meta-strings can be case sensitive at the preference of the assigning authority; parsers therefore MUST treat these as case sensitive, and any case mapping needed to introduce case insensitivity is the responsibility of the relevant resolution system.

A hyphen SHOULD be used as the delimiting character between the prefix and the meta-string. Within the meta-string, a hyphen MAY be used for separating different sections of the identifier from one another.

A colon SHOULD be used within the prefix only as a delimiting character between the format code and sub-namespace code(s), which splits the format specific namespace into smaller parts.

Maintenance agencies should not use in meta-strings characters which would require percent-encoding.

Rules for lexical equivalence:

None.

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).

If a format has been translated to multiple languages, agencies maintaining the original version and its translations shall agree between themselves who will assign URNs to its elements, default value being the organization responsible of the original version of the format.

Security and Privacy:

URN:META identifiers do not have any known security or privacy issues.

Interoperability:

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

Resolution:

General

URNs in the URN:META namespace shall be resolvable.

URN:META namespace shall support URN to URL resolution service from the identifier to the page describing the identified metadata element, or another service which fulfills a relevant function within this context.

Resolution services may be maintained by the agencies maintaining the formats or a third party. If so, the agency maintaining the format and the agency shall agree on how the URN - URL mappings are maintained.

Locating the appropriate resolver

There is no automated mechanism for locating the correct URN resolver for a URN:META identifier. Once a format maintenance agency has decided to implement URNs, it makes independently the decision of which resolver shall support resolution of its URNs within the URN:META namespace.

Initially URN:META URNs will be represented as HTTP URIs. Eventually URN:META identifiers may become resolvable as such. Each sub-namespace may have its own resolver.

Example

Namespace URN:META:MARC: contains URNs which identify the tags of MARC 21 metadata formats (<https://www.loc.gov/marc/>) in various languages.

In this example, URNs are expressed as HTTP URIs which use the URN resolver of the National Library of Finland, located <http://urn.fi>.

If the language setting of the client is English (and as a default) URNs <http://urn.fi/urn:meta:marc:<nss>> 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. For instance MARC Bibliographic format tag 245 (Title Statement)

<http://urn.fi/urn:meta:marc:bd245>

will resolve to

<https://www.loc.gov/marc/bibliographic/bd245.html>

unless the Accept-Language header setting is used to direct the user to a page in another language.

Since pages for all tags in MARC 21 format for bibliographic data have been named in the Library of Congress Web site using the same logic, no mapping table is required in the resolver. URNs can be mapped to URLs of their descriptive pages programmatically.

Using HTTP language negotiation the client may request this page in other languages. If the language setting of the client is Finnish, URN shall be resolved to

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

and if it is Swedish, to

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

If the language setting is not supported, default (English) shall be used.

Persistence

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

Element specific metadata about deprecated metadata formats such as USMARC, UKMARC or FINMARC may be hard to find and 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. Other formats and rules may be added in the future, with additional examples.