

“Modern” MARC (03/24/2026)

The Library of Congress (LC) will begin focusing its cataloging efforts on creating more “modern” MARC bibliographic records. Among other changes, discussed below, these records will include less coded data in fixed fields in favor of textual data in variable fields (to align with original Resource Description & Access (RDA), LC’s current cataloging code) and more identifiers. There will be no large scale retrospective conversion of records; these changes will be introduced organically, as LC catalogers perform their duties and themselves alter their cataloging practices to reduce data redundancy.

While LC cataloging practice will start to align better with the changes described below, it is important to note upfront that these changes – these practices – can already be observed in LC distributed records and in the broader library community. Vendor records, which LC loads and distributes before and after LC modification, routinely include numerous identifiers and have implemented the modelling changes discussed herein to accommodate RDA. Records imported into LC’s system via copy cataloging are similar – they are replete with identifiers and also contain modelling changes to accommodate RDA. LC practice is to do as little (or as much) as needed to these imported records in support of its collections. Since the modelling variations and identifiers in these records are entirely valid, and to remove them would result in a lesser description and take cataloger effort, these differences and additions have been retained. Also members of the Program for Cooperative Cataloging (PCC) enhanced bibliographic records with URIs as part of the [PCC URIs in MARC Pilot](#), starting in 2019. PCC catalogers are now permitted and encouraged to add URIs to both MARC bibliographic and authority records. (Although this document is specific to MARC bibliographic records, the name authority records LC has distributed for a decade now have long supported the features under discussion here pertaining to bibliographic records.)

What is notable – and which this document serves notice about – is that LC practice will now actively create, and therefore distribute, records that contain additional identifiers and feature more RDA-specific and linked data modelling choices.

MARC over time

MARC has a long history and also a long history of supporting changing cataloging practices. The MARC format came into production for exchange of bibliographic data in 1968 with the MARC II format that covered textual monographic resources (books). Over the next 15 years it was expanded to encompass formats for other resource groups: serials, maps, music and recorded sound, films and visual material, computer files, and manuscripts. And in the following years each media was expanded to cover special subgroups like video disc, globes, electronic text, etc. Up until 1990 each resource group had its own format with an attempt made to treat overlapping data elements the same cross the media. During this early period new cataloging rules also brought on adjustments to the formats. Various special chapters of Anglo-American Cataloguing Rules (AACR) came out in the 1960s followed by AACR2 in the early

1970s. In the early 1990s a major project called “format integration” brought the different media-based MARC specifications together, introducing many MARC changes to bring the various elements that had been introduced over time to a better consistency in a single specification, USMARC.

However, with the introduction of RDA in 2010 – the current cataloging rules -- many new redundancies and differences were introduced to MARC. They reflected new ways of modeling bibliographic data and new terminology for describing it.

MARC coded data with RDA

One of the major changes introduced by RDA was a preference for textual data in newly established fields over coded data in existing fixed fields. An example can be illustrated by some of the coded data fields in MARC, especially the 007 and 006 fields. Textual subfields in the MARC 3XX range were established to redundantly carry in textual form much of the coded data in the 007 fields. Much, if not all, of the coded information in these fields can be recorded in (new) variable fields.

For example data from the 007:

007 kd#bc|

is duplicated with the 340:

340 ## \$a cardboard \$d collotype \$g black and white

Catalogers are now trained to use these new 3XX fields. But the prior practice of recording this (same) information in fixed fields continues. This is not only producing duplicative and redundant data in MARC bibliographic records, but also requires additional, arguably redundant, effort on the part of catalogers.

In addition to textual description (versus coded), RDA strongly encourages the use of identifiers in the data and they are essential for linked data. Many of these identifiers take the form of Uniform Resource Identifiers, or URIs, which typically look like (and are) a web address, or URL. An example:

<http://id.loc.gov/vocabulary/mmaterial/crd>

The use of URIs, or identifiers generally, not only introduce additional data to the records, but also require an alteration to how the data are modelled in MARC. Where before (see the 340 example above) a single field may have been used for multiple values, in order to clearly associate identifiers with labels, the field must be repeated:

340 ## \$a cardboard \$0 <http://id.loc.gov/vocabulary/mmaterial/crd>

340 ## \$d collotype \$0 <http://id.loc.gov/vocabulary/mproduction/collo>

340 ## \$g black and white \$0 <http://id.loc.gov/vocabulary/mcolor/blw>

While this repetition of fields is sometimes done today to relate URIs and values, it will be used more consistently in LC records to avoid ambiguity.

Data duplication

As indicated above, most of the data carried in coded form in an 007 will be carried in the 3XX fields established for that data when RDA was adopted. However, for a transition

period the first two character positions of 007s (007/0, 007/1) will continue to be carried in records, even though they duplicate information in the new fields 337 and 338 established for RDA.

Minimal ISBD punctuation

In line with the Program for Cooperative Cataloging's (PCC) policies on reducing the amount of ISBD punctuation in bibliographic records and allowing flexibility in the application of ISBD punctuation, records will continue to retain full ISBD punctuation in only the 245 and 264 fields, partial ISBD punctuation in 250 and 490, but will have minimal ISBD punctuation in other parts of the bibliographic record. MARC Descriptive cataloging form (Ldr/18) will be coded "i" (ISBD punctuation included).

Non-Latin scripts

LC MARC bibliographic records will eventually start to expand the non-Latin scripts used. They will have the following characteristics:

- Access points – e.g. 130, 240, 700, 710, 711, 730 – will continue to hold transliteration of non-Latin data, with paired 880s containing the non-Latin data.
- Transliteration will generally be supplied in the 245, 250, 264, and 490 fields with paired 880s containing any non-Latin data.
- Transliteration of non-Latin data will be in regular fields (not-880 fields), as now.
- Transliteration will follow the ALA-LC transliteration tables as now.
- The 880 fields will be linked to the transliterated fields in the usual manner using the \$6 subfield technique.
- Non-Latin script data that is not paired with a Latin field may be in regular fields rather than 880 fields. This will occur mostly in the 5XX range. Some unpaired non-Latin will continue to be in 880 fields with an indicator clarifying the data does not pair with a Latin field.
- BCP47 language and script codes will be included in MARC records as described below.

BCP47 codes in data provenance subfields (usually \$7)

BCP47 is a shorthand reference to a specification that encodes human languages and scripts in contemporary data structures, especially those designed for the web. The specification is published by the Internet Engineering Task Force (IETF).

BCP47 components are separated by hyphens and are identifiable by length and position in the overall code. The first two or three characters are the language: "ko" or "en" or "sgn". If the second component (after a hyphen) is four characters, it is a script reference: "ko-hang" or "en-latn." Typically, if the script can be inferred from the language, it is not included since it is considered unnecessary information. For example, the "latn" in "en-latn" or "hebr" in "yi-hebr" is unnecessary because if the text is in English, Latin script is assumed and if the text is in Yiddish, Hebrew script can be assumed.

Language and script can be derived from existing MARC records in most cases. Adding the BCP47 encoding and capturing this information in the BIBFRAME and MARC

environments allows for greater precision and sets the stage for more language-specific descriptions in bibliographic data.

- In Modern MARC records, BCP47 language/script codes will be included in Data Provenance subfields of 880 fields and, usually, their Romanized pairs.
- BCP47 language/script codes may be included in other fields as deemed appropriate by a cataloger, provided capturing the language and/or script is deemed material.
- Although the Data Provenance subfield design permits specific subfield identification, the BCP47 codes added in Modern MARC will reference the entire MARC field. Over time, we will pursue ways to make this more precise. Some may deem this is an imperfect design as it will invariably identify a string with both non-Latin and Latin text as being entirely non-Latin. For example:
880 1 \$6100-01/\$1 \$a 김 동찬\$c(Cartographer)\$e cartographer \$7(bcp47)ko
- The script component of the BCP47 code will not be included when the script can be inferred from the language component.
- BCP47 codes will be all lowercase.

It is early days for this feature and it is therefore subject to modification.

Provision statement

To support RDA, a new provision statement field, 264, was established in MARC in order to separate and label statements for production, publication, distribution, and manufacture. Those statements had been gathered into one MARC 260 statement for many years since 260 was not repeatable until 2011. LC will build on that separation of provision statements. MARC records that formerly contained several provision statements in one 260 field will use the 264 field and separate the provision statements into one 264 for each statement.

Earlier practice:

260 ## \$a London, England : \$b Penguin Books ; \$a New York, N.Y. : \$b Penguin Putnam, \$c 1999.

Description converted to:

264 #1 \$a London, England : \$b Penguin Books.
264 #1 \$a New York, N.Y. : \$b Penguin Putnam, \$c 1999.

Series transcription

MARC field 490 is used to record a transcribed series statement. The 490 \$a is repeatable for a main series that has a subseries or when a series has a parallel title. In order to disambiguate the combined transcribed series statement, MARC records that formerly contained several series statements in one 490 field will separate the series statements into one 490 for each statement: main series title and subseries title or series title and parallel series title.

Earlier practice:

490 1#\$aDepartment of State publication ;\$v7846.\$aDepartment and Foreign Service series ;\$v128

Description converted to:

490 1#\$aDepartment of State publication ;\$v7846.

490 1#\$aDepartment and Foreign Service series ;\$v128

Making LCSH more linked data compliant

Library of Congress Subject Headings (LCSH) policies on genre/form subdivisions are not linked data compliant, since for many LCSH heading + genre/form subdivision(s) strings, there is no corresponding comprehensive URI. BIBFRAME is an opportunity to make LCSH more linked data compliant. Modern MARC records will not always include LCSH form subdivisions. In lieu of LCSH form subdivisions, a faceted term from Library of Congress Genre/Form terms (LCGFT) will be assigned in the 655 field to identify the genre/form of the resource.