

The importance of metametadata in MARC 21 - from a German perspective

Introduction

The MARC/RDA Working Group¹ (MRWG) has submitted a follow-up discussion paper on "Recording Data Provenance in the MARC 21 Formats"² for consideration during the Summer 2021 meetings of the MARC Advisory Committee³ (MAC). On behalf of the German part of the MARC community, the German National Library has a member on the MRWG, and significantly contributed to the paper. This was done based on both the impression that the topic of metametadata is an important real-life and forward-looking issue, and on an approach that it is worthwhile to develop options for a syntactical MARC solution.

The reactions from the different constituencies on the MARC listserv were mixed, with a focus on hesitance, to say the least⁴. Amongst a few more positive voices, the German comments⁵ were the most supportive ones. While we are thankful for all the response provided by the community, we see the need to further contribute to the discussion and unfold our approach, here in this paper, and during the MAC meetings.

Use cases

In our context, both for bibliographic data and for authority data metametadata is already in use, with a growing tendency. Some types of information have been identified to be in accordance with the RDA data provenance elements, some of the types seem to have existed before and beyond the RDA elements.

One bibliographic use case is the creation of metadata by a machine: Information about the language of the resource, authors, classification and subject headings, genre data and other types of information are added to a record, accompanied by metametadata about *who* created the information *how* and *when*, and *how reliable* it is.⁶

Another use case is the information which transliteration standard has been used during the creation of a Latin script string out of a non-Latin original script, a case that we have brought to the MARC community in 2016, with no solution so far.⁷

In authority data, mainly for the Gemeinsame Normdatei (GND, Integrated Authority File), we already have metametadata, i.e. information about who added a field to an authority record (and is responsible for its content), about the script and language of a field, about the chronological validity of a field content (e.g. during which years in history has *this* person been living at *this* location, or worked for *this* organization?).

Over the last few years, we have started an intense, inviting and far-reaching process of opening up our GND for a broader range of basically anyone who is interested in entity management. Here, Wikibase is being analyzed and tested to form a secondary home for the GND. In this context, metametadata is key in assisting the use of the metadata, and will form an important layer around the main information. Mirroring this in our internal format (and later on in the MARC Authority format) forms another set of use cases. Quite recently, we have identified the RDA data provenance element "context of use" to be a close equivalent to what we call the "PLUS Code" defined for a specific community and its domain, e.g. the community which describes manuscripts. It has been a promising effort to compare the RDA types of data provenance to what we already identified as candidates for "statements about statements", and we were glad to recognize that there is a large overlap between the two groups.

¹ https://www.loc.gov/marc/mac/MARC-RDA_Working_Group.html

² <https://www.loc.gov/marc/mac/2021/2021-dp10.html>

³ https://www.loc.gov/marc/mac/an2021_age.html

⁴ <https://listserv.loc.gov/cgi-bin/wa?A1=ind2106&L=MARC>

⁵ <https://listserv.loc.gov/cgi-bin/wa?A2=ind2106&L=MARC&P=10762>

⁶ cf. Proposal 2020-01 <https://www.loc.gov/marc/mac/2020/2020-01.html> and its predecessors

⁷ cf. Discussion Paper 2016-DP26 <https://www.loc.gov/marc/mac/2016/2016-dp26.html>

MARC format solutions so far

It has to be kept in mind that large parts of the community in German speaking countries maintain bibliographic and authority data not in the MARC formats, but in different format solutions, often in the format Pica3 / Pica+. In these cases, conversion routines of high complexity transform data from the internal storage format Pica+ to MARC 21 data, and back to Pica+ for reasons of importing data. In defining internal data elements, we are flexible, and for metametadata on the field level we use subfields with uppercase letters as format designation.

Showing this information in MARC led to different solutions: For bibliographic data, field 883⁸ "Metadata provenance" has been created and is widely in use. Its main focus is on automatically created metadata, and it has been quite recently extended to carry information on intellectually created metadata. The technique of this field, the extensions needed to more fully cover metametadata, and its advantages and disadvantages (\$8!) is thoroughly described in the discussion paper (section 2.3.1.). We see this field as a possible way, but not as a good practice in providing metametadata, and prefer to move away from this solution.

For authority data, mainly for the Gemeinsame Normdatei (GND, Integrated Authority File), we are using local coding in MARC: Subfield \$9 is defined, with "sublabels": The same uppercase letter which forms the internal subfield name is the first character of the content of the subfield, a fixed character ":" follows, and then the content itself is provided. Thus, "\$UCyrl" as internal subfield has the MARC equivalent "\$9U:Cyrl", both meaning "The content of *this* field is provided in Cyrillic script". Similarly, "\$9L:" for language and "\$9Z:" for chronological validity of a field content are in use. "\$C" / "\$9C:" are in the pipeline for information about "context of use". In addition, subfield \$5 is used for the creator of the metadata internally mapped to either \$5 where it is defined in MARC, or to "\$95:" where it isn't.

We think that both field 883 and subfields \$9 with sublabels are more or less workaround solutions (although, from the MARC perspective, on different levels). So, it was not only a syntactical challenge to contribute to the discussion paper, but an opportunity to develop better format solutions which have the potential to solve some larger problems.

Relevance to the wider MARC community

One may argue that defining a local solution for the German speaking community might be sufficient here. Out of the characters "\$A" to "\$Z", there still is room available, so that we are able to define additional metametadata subfields and show them in \$9 subfields with sublabels. Alternatively, we may discuss the adoption of the more flexible coded value syntax (explored in section 2.3.5. of the discussion paper, in analogy to subfield \$0), and would significantly benefit from the development so far. We could even think of applying the approach of \$9 and sublabels or \$9 and coded values to bibliographic data, thus getting rid of field 883 in general. A minor disadvantage may be that in a few fields, \$9 is already defined "locally", i.e. for the German part of the MARC community (ISBN with hyphens in field 020, machine-sortable form of volume/sequential designation in fields 800-830). In these cases, the different scopes of \$9 may cause conflicts.

All these are good points. We don't intend to move the MARC format into a direction that wouldn't at least help some other parts of the community in achieving their goals and covering their use cases. On the other hand, we seem to have a good understanding and some experience which we'd like to offer to the community, with a focus on the chances and benefits. We can imagine that now, based on valid analyses by the RDA community, and thoroughly explored by the MARC/RDA Working Group, it is the right point in time to accommodate this approach by the MARC community as a whole. We would not like to miss this opportunity.

⁸ <https://www.loc.gov/marc/bibliographic/bd883.html>

Examples

1. Bibliographic data

1.1. Bibliographic record, with a variety of main and 883 fields (cf. <http://d-nb.info/1216417776>)

```
041 ## $81\p$aeng
082 74 $86\p$a100$a400$qDE-101$223sdb
100 1# $83\p$0 (DE-588)1215943776$0https://d-nb.info/gnd/1215943776
    $0 (DE-101)1215943776$aTolkien, Niels$dl1985-$eVerfasser$4aut
    $2gnd
245 10 $a-The- notion of homonymy, synonymy, multivocity, and pros hen
    in Aristotle$cNiels Tolkien ; Betreuer: Christof Rapp
600 07 $85\p$0 (DE-588)118650130$0https://d-nb.info/gnd/118650130
    $0 (DE-101)118650130$aAristoteles$dv384-v322$2gnd
655 #7 $82\p$0 (DE-588)4113937-9$0https://d-nb.info/gnd/4113937-9
    $0 (DE-101)041139372$aHochschulschrift$2gnd-content
700 1# $84\p$aRapp, Christof$eAkademischer Betreuer$4dgs
883 2# $81\p$aef-lc-pa$d20200909$qDE-101
    $uhttps://d-nb.info/provenance/plan#ef-lc-pa
883 2# $82\p$aie-in-pa$d20200909$qDE-101
    $uhttps://d-nb.info/provenance/plan#ie-in-pa
883 2# $83\p$adnb-pa$d20200909$qDE-101
    $uhttps://d-nb.info/provenance/plan#dnb-pa
883 1# $84\p$aanpi$d20200824$qDE-101
    $uhttps://d-nb.info/provenance/plan#npi
883 0# $85\p$aep-gnd$c0,07143$d20200825$qDE-101
    $uhttps://d-nb.info/provenance/plan#aep-gnd
883 2# $86\p$adnb-pa$d20200909$qDE-101
    $uhttps://d-nb.info/provenance/plan#dnb-pa
```

1.2. Bibliographic record, original title differently translated, using two different norms for transliteration (adapted from <https://www.loc.gov/marc/mac/2016/2016-dp26.html>)

```
245 10 $aAs'ila ḥaula 'l-mar'a wa-'l-masḡid$bfī ḡau' nuṣūṣ aš-šarī'a
    wa-maq ṣidih $cd. sir 'Auda$7(dpetn)DIN 31635:2011
246 3# $aAs'ilah ḥawla al-mar'ah wa-al-masjid$dfī ḡaw' nuṣūṣ al-
    sharī'ah wa-maq ṣiduh$7(dpetn)ISO 233
```

Note:

Here, the option of subfield "\$7" with coded values is chosen for data provenance, fictitious coded value is "dpetn" for "Data Provenance Element: Transliteration Norm".

2. Authority data

2.1. Authority record for Stefan Zweig, with script, language and chronological validity metametadata (cf. <http://d-nb.info/gnd/118637479>)

```
100 1# $aZweig, Stefan$d1881-1942
400 1# $9U:Kore$a[Name in Korean script]$d1881-1942$5DE-576
400 1# $9U:Hebr$a[Name in Hebrew script]$d1881-1942$5DE-576
400 0# $9U:Jpan$a[Name in Japanese script]$d1881-1942$5DE-576
400 0# $9U:Cyrl$9L:srp$a[Name in Cyrillic script and Serbian
    language]$d1881-1942
400 1# $9U:Cyrl$9L:rus$a[Name in Cyrillic script and Russian
    language]$d1881-1942
400 0# $9U:Cyrl$9L:ukr$a[Name in Cyrillic script and Ukrainian
    language]$d1881-1942
400 0# $9U:Cyrl$9L:bul$a[Name in Cyrillic script and Bulgarian
    language]$d1881-1942
400 0# $9U:Grek$9L:gre$a[Name in Greek script and language]$d1881-1942
551 ## $0(DE-101)040221539$0(DE-588)4022153-2$aGroßbritannien
    $4ortx$wr$iExil$9Z:1934
551 ## $0(DE-101)040787044$0(DE-588)4078704-7$aUSA
    $4ortx$wr$iExil$9Z:1939
551 ## $0(DE-101)04008003x$0(DE-588)4008003-1$aBrasilien
    $4ortx$wr$iExil$9Z:1940
```

2.2. Authority record for Codex Manesse, second "see from" title with additional information (cf. <http://d-nb.info/gnd/4135049-2>)

```
130 #0 $aManessische Handschrift
430 #0 $aCodex Manesse
430 #0 $aHandschrift$gUniversitätsbibliothek Heidelberg$nCod. Pal.
    germ. 848$5DE-2489$9C:[PLUS Code for "Manuscript cataloging"]
```

Note:

"DE-2489" is the MARC Organization Code / ISIL of the Handschriftencensus, see <https://sigel.staatsbibliothek-berlin.de/suche/?isil=DE-2489> ,

cf. <https://en.wikipedia.org/wiki/Handschriftencensus> .