

Video Language Coding: Best Practices

Created by the
OLAC Cataloging Policy Committee
Video Language Coding Best Practices Task Force

2006-2007 task force members:

Kelley McGrath, Chair
Cindy Badilla-Melendez
Susan Leister
Katia Strieck
Carolyn Walden

2012 task force members:

Kelley McGrath, Chair
Karen Gorss Benko
Irina Stanishevskaya
Carolyn Walden

Contents

Introduction	3
Fields and Subfields Used for Video Language Coding	3
Spoken and Written Language Information	4
Standard Examples.....	4
No Spoken Content	5
Multiple Works with Different Language Information	5
Mixed-Language Soundtracks	6
Brief Sequences in Language(s) Other Than the Main Language(s)	7
Original Language	8
Types of Information Not to Include in 008/lang and 041	8
Accompanying Material	9
Packaging and Credits	9
Appendix 1: Changes Made to the MARC21 Format as a Result of the Work of This Task Force	11
Appendix 2: Captions, Intertitles, and Subtitles	12
Functional Definitions of Terms Used in This Document	12
Function vs. Encoding Method for Subtitles and Captions on VHS and DVD	13
Resources for Further Information on Subtitles and Captions on DVDs	14

Introduction

The 2006-2007 task force was charged with creating a set of best practices for coding MARC 008/lang and 041 language information for videos, especially DVDs, and with using that exercise to examine whether any changes could be made to the MARC21 format (coding or directions) that would improve access to the multiple types of language information found on videos. The work of that task force resulted in a number of changes to the MARC format, which are described in appendix 1. This current document, completed by the 2012 task force, provides guidance for coding video language information using the current MARC documentation.

Our recommendations are based on the following premises:

1. Coded language data is intended for use in retrieval, limiting, and sorting.
2. Coded language data does not need to describe all language-related information about an item that might be of interest to users. Coded language information can be expanded on and complemented by information in 546 free text language notes.
3. Coded language data should support the retrieval of the language(s) of the main work(s) on the item rather than the language(s) of supplementary and bonus materials.
4. Coded language data should support retrieval based on language(s) in which the item is usable rather than all language(s) that might be found in the item.
5. For moving image materials, patrons are most interested in retrieving, limiting, and sorting by the following types of language information:
 - Spoken, sung, or signed language of the main content
 - Written language of the main content (including subtitles, captions, and intertitles)
 - The original language of the work

Fields and Subfields Used for Video Language Coding

Fields and subfields recommended for routine use for videos are marked with an asterisk.

- * 008 35-37/lang: Spoken, sung and signed languages
- * 041 \$a: Spoken, sung and signed languages
- 041 \$b: Languages of summaries on containers
- 041 \$e: Languages of librettos
- 041 \$g: Languages of accompanying material
- * 041 \$h: Original languages of main work(s)
- * 041 \$j: Written languages, including subtitles, captions and intertitles

Note that 041 \$d, which is labeled as “sung or spoken text,” is not used from moving image materials. It is primarily used for sound recordings.

Spoken and Written Language Information

Although current systems do not support granular searching of 041 subfields, we believe that coding spoken and written languages separately is important. By coding the data more explicitly now, it will be there to support future search systems. We recommend following the updated MARC21 documentation for moving images by coding 008/lang and 041 \$a only for spoken, sung or signed languages and using the MARC code zxx (no linguistic content) for videos with no spoken language. Use 041 \$j to code for written language, including intertitles, subtitles and captions. This will enable the separate retrieval of spoken and written languages when desired, as well as the creation of an index for “accessible in” languages that would include both 008/lang and 041 \$a plus 041 \$j.

Standard Examples

Original English dialogue; English packaging, menus, and credits

```
008/lang eng  
041 0# $a eng $h eng
```

Original English dialogue; closed-captioned in English; English packaging, menus, and credits

```
008/lang eng  
041 0# $a eng $j eng $h eng  
546 ## Closed-captioned.
```

Japanese language film; English subtitles; English packaging and menus

```
008/lang jpn  
041 1# $a jpn $j eng $h jpn  
546 ## In Japanese with English subtitles.
```

Japanese language film; optional dubbed English soundtrack; optional English subtitles; English packaging and menus

```
008/lang jpn  
041 1# $a jpn $a eng $j eng $h jpn  
546 ## Japanese or dubbed English soundtrack; optional English subtitles.
```

German language opera; optional English, French, German, or Italian subtitles; English packaging and menus

008/lang ger
041 1# \$a ger \$j eng \$j fre \$j ger \$j ita \$h ger
546 ## Sung in German; optional English, French, German, or Italian subtitles.

English language film with English, French, or Spanish soundtracks; closed-captioned in English; optional subtitles in English, French, Spanish, Portuguese, Chinese, or Thai. English packaging and menus

008/lang eng
041 1# \$a eng \$a fre \$a spa \$j eng \$j chi \$f fre \$j por \$j spa \$j tha \$h eng
546 ## Closed-captioned; English or dubbed French or Spanish soundtrack; optional English, French, Spanish, Portuguese, Chinese, or Thai subtitles.

Recording of *The Bridge*, an opera performed in sign language, simultaneously sung in English

008/lang sgn
041 0# \$a sgn \$a eng \$h sgn
546 ## Performed with gestures, American Sign Language, a musical soundtrack, and in English.

No Spoken Content

Symphony performance; no spoken/sung language; credits in German; disc menu and packaging in English.

008/lang zxx

A Chaplin silent film on DVD with multiple subtitle tracks

008/lang zxx
041 1# \$a zxx \$j eng \$j chi \$j fre \$j kor \$j por \$j spa \$j tha \$h eng
546 ## Silent film with English intertitles and musical acc.; optional French, Spanish, Portuguese, Chinese, Thai, or Korean subtitles. Optional audio commentary track in English. Menus in English, Spanish or Portuguese.

Multiple Works with Different Language Information

Since 041 is a repeatable field, use separate 041 fields when needed for different works on one record.

Shorts! Volume One : 15 Award-Winning Film Festival Shorts

008/lang eng
041 0# \$a eng \$j eng \$h eng
041 1# \$a dut \$j eng \$h dut
041 1# \$a dan \$j eng \$h dan
546 ## Selected movies are closed-captioned; soundtracks principally in English; The Chinese wall in Dutch with optional English subtitles; John and Mia in Danish with optional English subtitles.

Disc 2 of *The Wild West*. Includes the Italian film *C'è Sartana (Fistful of Lead)*, dubbed in English, and the English language television feature *The Gunfighters*.

008/lang eng
041 1# \$a eng \$h ita
041 0# \$a eng \$h eng
546 ## *Fistful of lead* dubbed in English. The *gunfighters* in English.

Mixed-Language Soundtracks

When substantial portions of a video are in more than one language, code for all substantial languages present.

An Algerian DVD that is a clear mixture of French and Arabic. The characters often switch between the two languages within a sentence and depending on whom they are talking to use either French or Arabic. No subtitles.

008/lang ara [since the package and credits were in Arabic]
041 0# \$a ara \$a fre \$h ara \$h fre
546 ## Dialogue consists of a mixture of Arabic and French.

Europa Europa is primarily in German and Russian, but has some parts in Hebrew and Polish. This example illustrates both handling more than one primary language and omitting those languages that occur only briefly.

008/lang ger
041 1# \$a ger \$a rus \$j eng \$j fre \$j spa \$h ger \$h rus
546 ## Filmed in German and Russian, with brief sequences in Hebrew and Polish. Optional subtitles in English, French, or Spanish. Closed-captioned in English.
(noting the Hebrew and Polish sequences in 546 is optional)

Because of the inability to mark a mixed-language soundtrack in MARC, sometimes users searching by spoken language will retrieve videos that are not usable to monolingual speakers of

languages coded in 008/lang and 041 \$a. In the examples of *Joyeux Noel* and a hypothetical DVD below, both have the same information in 008/lang and 041 \$a, but only the second example is useful to a monolingual speaker of French or German. There is no alternative that would compensate for this problem in the existing MARC language coding structure nor are there existing user interfaces that are sophisticated enough to deal with these distinctions.

Joyeux Noel. Soundtrack of DVD and original film in English, French, and German; optional English, Spanish, or Portuguese subtitles.

008/lang fre
041 1# \$a fre \$a eng \$a ger \$j eng \$j por \$j spa \$h fre \$h eng \$h ger
546 ## Soundtrack in a mixture of French, English and German; optional English,
Spanish, or Portuguese subtitles.

A hypothetical DVD of a French film that has optional English, French, or German soundtracks and English, Spanish, or Portuguese subtitles.

008/lang fre
041 1# \$a fre \$a eng \$a ger \$j eng \$j por \$j spa \$h fre
546 ## French, English, or German soundtracks; optional English, Spanish, or Portuguese
subtitles.

Brief Sequences in Language(s) Other Than the Main Language(s)

Code lang/041 for languages which are substantial and which the intended audience needs to be able to understand to use the item. Include brief or subsidiary languages in 546 if thought important.

The Internet Movie Database often lists all languages (see their record for *The Godfather* at <http://www.imdb.com/title/tt0068646/>), even from brief sequences; DVD packages tend to list soundtracks by the language(s) of the intended listener (except in the case of originally multi-lingual films described in the previous section). The latter approach is more useful for most users of library catalogs.

The Godfather is primarily in English, but has a few Italian and Latin sequences.

008/lang eng
041 0# \$a eng \$j eng \$h eng (*ignore the Italian and Latin for coding purposes*)
546 ## In English with brief sequences in Italian and Latin with English subtitles.
(*noting the Italian and Latin sequences in 546 is optional*)

NOT: 041 0# \$a eng \$a ita \$a lat \$j eng \$h eng \$h ita \$h lat

Original Language

Users are often interested in films that were originally in French, Spanish, Arabic or some other language. The usage of 041 \$h was broadened in 2011 to allow its optional use whether or not the resource is or contains a translation. This enables catalogers to record 041 \$h for a monolingual video issued in its original language, such as the Spanish language video of a movie originally filmed in Spanish shown below. This would also enable libraries to develop a separate index for 041 \$h that would search only for the original language of films and television programs.

```
008/lang spa  
041 0# spa $h spa
```

Use \$h for the original language regardless of whether that language content was spoken, sung, signed or written. The example on page 5 of a Chaplin silent film includes 041 \$h eng for the original English intertitles. The *Bridge* example on the same page shows the use of 041 \$h sgn for the original sign language.

Types of Information Not to Include in 008/lang and 041

- Packaging language(s) (disc or tape label, container, disc menu)
- Special feature language information (audio commentary tracks on DVDs, spoken and written languages on special features)
- Credits
- Accompanying material (e.g. guides, booklets)

Although the language of summaries on containers can be recorded in 041 \$b, there is rarely any utility in doing so.

In an ideal world, one might code for special feature language options, but we do not recommend coding these languages for two reasons. (1) Due to the lack of sufficiently granular parsing in the 041 field, these languages cannot be distinguished from the language(s) of the main content. (2) Situations in which one could imagine patrons using this information for searching or limiting rather than for informational purposes while looking at a record are rare and do not justify the effort to create and code separate subfields for the language(s) of special features.

The language(s) of accompanying material can be explicitly coded in 041 \$g (defined as language code of accompanying material other than librettos, or scripts or accompanying sound for visual materials) when the material is considered significant. It is not clear when this would be useful for searching or retrieval. Unlike musical recordings, few videos have significant accompanying material and, in practice, this subfield generally does not seem to be recorded if both the accompanying material and the video are in English. It seems unlikely that most users would want to search separately for videos with accompanying material in a certain language (although it might be useful contextual information once they are looking at an individual record) nor that we would have enough data in such an index to support a useful search option. We are also concerned about the historical use of \$h (language code of original and/or intermediate translations of text) after \$g, although it appears to be rarely used for moving

image materials, because it decreases the ability to use \$h to help determine the original language of the main material.

If deemed important, include the language of accompanying material in a note, but do not record it in coded form in 041. However, coding \$g is unlikely to produce any negative impacts. Optionally, at the discretion of the cataloging agency, use \$g for accompanying material.

Accompanying Material

Hypothetical Japanese film from Criterion Collection; Japanese language film with English subtitles and substantial booklet of criticism about the film in English translated from German

Best practice recommendation:

008/lang jpn
041 1# \$a jpn \$j eng \$h jpn
546 ## Japanese soundtrack; optional English subtitles.
Include the booklet in 300 \$e as accompanying material or describe it in a note.
Optionally, mention language information about the booklet in a note.

Libraries wishing to explicitly code the language of the booklet should use

008/lang jpn
041 1# \$a jpn \$j eng \$h jpn \$g eng
546 ## Japanese soundtrack; optional English subtitles. Booklet in English.
Include the booklet in 300 \$e as accompanying material or describe it in a note. Mention language information about the booklet in a note.

Packaging and Credits

Swedish film that has been dubbed into English; credits (except for title) still in Swedish. Packaging and menus in English.

008/lang eng
041 1# \$a eng \$h swe
546 ## Dubbed in English; credits in Swedish.

Our Daily Bread. No spoken dialogue, no intertitles, no subtitles; English credits, menus, and packaging; originally produced in Germany.

008/lang zxx
546 ## English language-credits version of a German film without dialogue.

Il Cerchio = The Circle. Edizione italiana. Farsi soundtrack (original Farsi); English subtitles; Italian credits on screen.

008/lang per

041 1# \$a per \$j eng \$h per

546 ## Farsi soundtrack with English subtitles. Onscreen credits in Italian.

Appendix 1: Changes Made to the MARC21 Format as a Result of the Work of This Task Force

From 2008 to 2011, OLAC proposed several changes to the MARC21 format to enable accurate coding of the types of language information that we have identified as important. The changes are summarized here.

1. The instructions for 008/lang and 041 \$a for moving images were revised to limit their use to spoken and signed language. This supports the ability to search spoken and written languages separately. The former definition of 008/lang and 041 \$a for moving images created ambiguous data because it did not accurately distinguish between spoken and written language. The old definition seemed to be intended to code the “main” language of the item, reverting to written language if there is no spoken language.
2. A new subfield \$j was created for written languages on moving image materials, including intertitles, subtitles and captions.
3. The definition of 041 \$h was narrowed so that subfield is used only for the original language(s) of the main work and loosened so that these original language(s) can optionally be recorded here whether or not a translation is involved.

In order to narrow the definition, several other types of language information that were previously recorded in 041 \$h have been moved to new subfields.

The new definition of 041 \$h excludes languages of intermediate translations. These rarely, if ever, occur in moving image formats, but if they do occur are now to be recorded in the new subfield 041 \$k. 041 \$k is defined as containing “Language code(s) for an intermediate language between the original and the current translation, where the resource was translated from an intermediate language other than the original.”

The definition was also narrowed to exclude original languages for things other than the main content. Two new subfields were defined for other types of original languages:

\$m - Language code of original accompanying materials other than librettos (R)

The subfield follows the related subfield \$b or \$g.

\$n - Language code of original libretto (R)

The subfield follows the related subfield \$e.

These are also unlikely to be used in most moving image cataloging. They are mostly used by the music cataloging community and could potentially be applied to videos of musical performances. For example:

Video of an opera sung in French, original opera in Italian; libretto in English, French, German, and Italian; liner notes in English, French, German, and Italian but liner notes known to be originally written in German.

```
008/lang fre
041 1# $a fre $h ita $e eng $e fre $e ger $e ita $n ita $g eng $g fre $g ger $g ita
    $m ger
```

For most purposes, the simpler version below with the information about the libretto and liner explained in notes to the extent thought important suffices.

```
008/lang fre
041 1# $a fre $h ita
```

Optionally, at the discretion of the cataloging agency, record the languages of accompanying material, but omit \$n and \$m.

```
008/lang fre
041 1# $a fre $h ita $e eng $e fre $e ger $e ita $g eng $g fre $g ger $g ita
```

We recognize that the reuse of 041 \$h is not an optimal solution since some existing data in 041 \$h represents languages other than the original language of the work. However, our initial proposal for a new subfield in 041 for original language was rejected. It is also possible to bring out original language with local genre-form terms as recommended by OLAC's best practices for moving images and Library of Congress Genre-Form Thesaurus (LCGFT) (http://www.olacinc.org/drupal/capc_files/LCGFTbestpractices.pdf), but the Library of Congress will not include non-genre information, such as language information, in its official genre-form terms. In the future, the most appropriate place for original language information may be in a FRBR-based Work-Primary Expression record as discussed by OLAC's Moving Image Work-Level Records Task Force in part one of their report (<http://www.olacinc.org/drupal/?q=node/27>). In current systems, it is necessary to record original language in the bibliographic record in order for it to be useful.

Appendix 2: Captions, Intertitles, and Subtitles

Functional Definitions of Terms Used in This Document

Intertitles: Generally associated with silent films, intertitles usually appear as separate frames containing written dialogue or other information to aid in comprehension.

Subtitles: Subtitles are text, usually appearing at the bottom of the screen, that provides a translation or transcription of spoken dialogue. Intended for viewers who can hear the soundtrack, subtitles are usually used for translations of foreign language films.

Captions: Captions are similar to subtitles, but also include contextual clues for viewers who cannot hear the soundtrack, such as identification of the speaker when it's not clear from the action on screen, and sounds, such as explosions or phones ringing.

Function vs. Encoding Method for Subtitles and Captions on VHS and DVD

The encoding format refers to the technical means of communicating textual information to the viewer. For VHS videos there is generally a 1:1 correspondence between the function and the encoding format of captions and subtitles.

VHS	Function	Encoding format
Closed-captions	Caption	Embedded in video signal. Requires hardware (line-23 decoder in VCR and in TV; included in all U.S. TVs 13" and over manufactured since July 1993). Disappear when tape is fast-forwarded
Open-captions	Caption	Imprinted on the tape; usually look blocky like closed-captions, but cannot be turned off and do not disappear when tape is fast-forwarded
Subtitles	Subtitle	Imprinted on the tape; cannot be turned off

Unfortunately, for DVDs there is not a 1:1 correspondence. Sometimes there is no practical way to tell if optional subtitles are functioning purely as subtitles or have the additional information necessary to function as captions.

DVD	Function	Encoding format
Closed-captions	Caption	Embedded in video signal. Requires line-23 decoder in DVD player or drive and in TV (included in all U.S. TVs 13" and over manufactured since July 1993). Not all DVD players or drives, especially older models, include the necessary decoder. In addition, the way the DVD has been encoded can interact with particular hardware and software configurations, such that the line-23 captions do not work (some captions will play on a stand-alone DVD player with TV, but not on a computer DVD drive or vice versa)
Open-captions	Caption	Cannot be turned off; may be encoded as part of the film image
Subtitles	Subtitle	Cannot be turned off; may be encoded as part of the film image
Optional subtitles	Subtitle	Digital subpicture bitmap overlay (not possible on VHS; usually turned on or off from the disc menu or remote, but sometimes hard-coded by the DVD producer so that they cannot be turned off)
Optional subtitles for the deaf and hard of hearing/captions	Caption (Sometimes referred to by publishers as SDH, "subtitles for the deaf and hard of hearing," or "English captions." Sometimes called "English subtitles" even though captioning information is also included)	

Resources for Further Information on Subtitles and Captions on DVDs

DVD Demystified

What's the Difference Between Closed Captions and Subtitles?

<http://www.dvddemystified.com/dvdfaq.html#1.45>

Joe Clark

Basic DVD Accessibility Capabilities

<http://joelclark.org/access/dvd/capabilities.html>

Federal Communications Commission

Closed Captioning: Guide

<http://www.fcc.gov/guides/closed-captioning>

Described and Captioned Media Program

An Introduction to DVD and Web Captioning

<http://www.dcmp.org/caai/nadh77.pdf>

Wikipedia definitions:

http://en.wikipedia.org/wiki/Closed_captioning

<http://en.wikipedia.org/wiki/Intertitle>

http://en.wikipedia.org/wiki/Subtitle_%28captioning%29