A Guide to the EMMA Metadata

Initial Release—April 18, 2023

Introduction

Each metadata property in EMMA is listed below in the order that it appears in the Bulk Uploads manifest.

The Property Name is how that property is normally referred to; it's followed by a designation of whether it's required or optional when depositing a remediated resource. That's followed by a description of the property, with an explanation of issues that you should be aware of, and a list of any controlled vocabulary terms.

The Field Name is the formal identification of the property in the database and schema. That's followed by the data type of its value or values and its cardinality (how many values it may have).

Here's how these aspects are formatted.

Property Name—*Required or optional*
Description of the property and relevant issues in non-technical language. Includes the controlled vocabulary if applicable.
Controlled vocabulary:
- Term as displayed [data value]
- Term as displayed [data value]

field_name
Data type
Cardinality

Bibliographic Metadata

File Name—*Required*

The exact name of the file that contains the resource, including the file extension. Note that for resources like Word or DAISY, there may be a number of files in the resource (e.g., each chapter in a book), all in a .zip container. In that case, the name of the container, with the extension .zip, is the proper file name. This will enable the system to precisely identify the correct file to deposit or retrieve.

file_data
Data type: json
Cardinality: one

Identifiers—*Optional*

Public identifiers associated with the creative work or the specific resource, like DOI, ISBN, or LCCN. For example, a DOI identifies a journal article in any of its formats; an ISBN is a product identifier, meaning that the hardcover, paperback, audiobook, and ebook each has a unique ISBN. While the alphanumeric identifier is all that's technically required, it's best to include the identifier type, e.g., "ISBN 0-205-30902-X" or DOI: 10.36591/SE-D-4401-8 (or, for a DOI, the actionable form https://doi.org/10.36591/SE-D-4401-8). Avoid using the OCN, which is an identifier of a WorldCat cataloguing record, not a unique identifier of the creative work.

dc_identifier
Data type: string
Cardinality: zero or more

Title—*Required*

The title of the work (e.g., a book or article title). It does not to be the full, formal title, although the more complete it is, the more likely it is to be accurately identified in a search.

dc_title
Data type: string
Cardinality: one

Author/Creator—*Required*

The name of the author(s), editor(s), translators, or other contributors to the creative work. In some cases, for widely known authors and their work, last names may be sufficient (e.g., "Shakespeare" or "Strunk and White"); but in most cases, a full spelled-out name as normally given is best (e.g., "F. Scott Fitzgerald" is better than either "Fitzgerald" or "Francis Scott Key Fitzgerald"). The exception is when sufficient other metadata is provided to disambiguate: "Fitzgerald" would be fine if the title "The Great Gatsby" is also provided. Completeness is more important for depositing than for searching, which can work even with name fragments.

dc_creator
Data type: string
Cardinality: one or more

Version—*Optional*

The particular edition if there's more than one. This can be expressed in a number of ways, for example "Fourth Edition," "4/e," "4th ed.," and even with a date, e.g., "1965 edition."

emma_version
Data type: string
Cardinality: zero or one

Series Type—*Optional*

The type of serial publication of which the resource is a member. This should always be provided if the resource being deposited is a member of a series.
Controlled vocabulary:
- Book [book]
- Journal [journal]
- Magazine [magazine]
- Newspaper [newspaper]

bib_seriesType
Data type: CV term
Cardinality: zero or one

Series Title—*Optional*

The name of the series, such as "Harry Potter," "PLOS," "The Atlantic," "The New York Times." This should always be provided if the resource being deposited is a member of a series.

bib_series
Data type: string
Cardinality: zero or one

Series Position—*Optional*

The sequential position of the resource within the series, such as a journal issue (e.g., "Autumn 2016," "Issue 234," "Vol. X, No. 5"), a book series (e.g., "Book Two," "No. 2," "II"), or a media series (e.g., "Episode Nine," "Second season, episode 5," "Sequel"). Providing this is often necessary to accurately identify the exact resource being required.

bib_seriesPosition
Data type: string
Cardinality: zero or one

Copyright Date—*Optional*

The copyright date as given on the copyright page of a book or on a journal article. If more than one year is listed, use the latest one.

dcterms_dateCopyright
Data type: string
Cardinality: zero or one

Publication Date—*Optional*

The date the specific version of the resource being sought or deposited was published. May not be the same as the copyright date.

emma_publicationDate
Data type: string
Cardinality: zero or one

Publisher—*Optional*

The name of the publisher as listed on the title page of a book or in the header or footer of a journal article. Be as specific as possible. For example, "Alfred A. Knopf" is an imprint of Knopf Doubleday Publishing Group, which is a subsidiary of Random House, which is part of Penguin Random House, often referred to as PRH. "PRH" or "Penguin Random House" would be the least precise; "Knopf" would be the most precise. For journals, it's fine to use an abbreviation if it's widely known and used, such as "NEJM" for "The New England Journal of Medicine" or "PLOS" for "Public Library of Science," but avoid using short journal titles that may be ambiguous, such as "Annals" for "Annals of Internal Medicine."
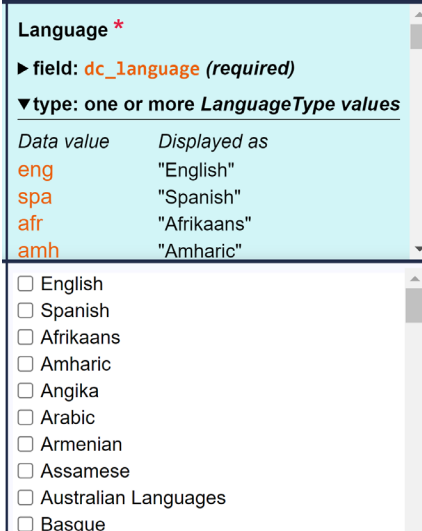
dc_publisher
Data type: string
Cardinality: zero or one

Language—*Required*

The primary language and other languages used in the resource. EMMA provides an extensive drop-down list of languages, using the spelled-out names, and then enters the ISO values in the data:

Language *	
▶ field: dc_language *(required)*	
▼ type: one or more *LanguageType values*	
Data value	*Displayed as*
eng	"English"
spa	"Spanish"
afr	"Afrikaans"
amh	"Amharic"

☐ English
☐ Spanish
☐ Afrikaans
☐ Amharic
☐ Angika
☐ Arabic
☐ Armenian
☐ Assamese
☐ Australian Languages
☐ Basque

dc_language
Data type: ISO LanguageType values
Cardinality: one or more

Rights—*Optional*

Designates the type of rights that are associated with the resource. It does not include a rights statement; it is only intended to inform a user as to what sort of rights protection has been asserted for the creative work. Note that both U.S. and international law have established that provision of a resource that has been remediated to a qualified print-disabled user is an exception to copyright.

Controlled vocabulary:

- Copyright [copyright]
- Creative Commons [creativeCommons]
- Public Domain [publicDomain]
- Other [other]

dc_rights

Data type: Rights value
Cardinality: zero or one

Description—*Optional*

A free-text description of the creative work, such as a synopsis, abstract, or marketing copy. (This is not a technical description of the resource.)

dc_description

Data type: string
Cardinality: zero or one

Subject—*Optional*

Keywords or subject terms describing the creative work.

dc_subject

Data type: string
Cardinality: zero or more

Related Identifiers—*Optional*

Identifiers of related editions or versions of the work, other than the specific edition or version of the resource being sought or deposited. For example, if the primary ISBN is for the hardcover version of a book, having the ISBN of the paperback or EPUB might also be useful in a search.

dc_relation

Data type: string

Cardinality: zero or more

Type—*Required*

Denotes the primary type of content in the resource.

Controlled vocabulary:

- Text [text]
- Sound [sound]
- Collection [collection]
- Dataset [dataset]
- Event [event]
- Image [image]
- Interactive Resource [interactiveResource]
- Service [service]
- Physical Object [physicalObject]
- Still Image [stillImage]
- Moving Image [movingImage]

dc_type

Data type: DcmiType value

Cardinality: one

Format—*Required*

Designates the type of technology that renders the content.

Controlled vocabulary:

- BRF [brf]
- DAISY [daisy]
- DAISY Audio [daisyAudio]
- EPUB [epub]   *(cont'd on next page)*

- Braille [braille]
- PDF [pdf]
- Microsoft Word [word]
- Tactile [tactile]
- Kurzweil [kurzweil]
- RTF [rtf]

dc_format
Data type: DublinCoreFormat value
Cardinality: one

Remediation Metadata

Format Feature—*Optional*
Designates accessibility-specific ways in which the content is conveyed, including specifically accessible file types.
Controlled vocabulary:
- Speech – TTS [tts]
- Speech – Human [human]
- Braille – Grade 1 (uncontracted) [grade1]
- Braille – Grade 2 (contracted) [grade2]
- Braille – Nemeth [nemeth]
- Braille – UEB [ueb]
- Braille – EBAE [ebae]
- Braille – Music [music]
- PDF – Scanned Image [imagePdf]
- PDF – Digital Text [digitalTextPdf]

emma_formatFeature
Data type: FormatFeature values
Cardinality: zero or more

Date Accepted—*Optional*
The date when the remediated resource was added to its repository.

dcterms_dateAccepted
Data type: ISO 8601, YYYYMMDD
Cardinality: zero or one

Source of Content—*Optional*

Where the resource to be remediated was obtained.

Controlled vocabulary:

- ACE / Scholars Portal [ace]
- Bookshare [bookshare]
- HathiTrust [hathiTrust]
- Internet Archive [internetArchive]
- EMMA [emma]
- ATN (Access Text Network) [accessTextNetwork]
- Publisher [publisher]
- Purchased [purchased]
- Faculty [faculty]
- Student [student]
- Library [library]
- Another DSO [otherDso]
- Other [other]

rem_source
Data type: SourceType value
Cardinality: zero or one

Source of Metadata—*Optional*

List of all parties or sources of metadata for the resource, including parties that may have refined existing metadata. Use roles, not names of individuals, e.g., "Librarian," "DSO." If the bibliographic lookup was used for some of the metadata, include "EMMA lookup" or the actual source, i.e., "Google Books," "WorldCat," or "Crossref."

rem_metadataSource
Data type: string
Cardinality: zero or more

Remediated By—*Optional*

The DSO or other service, not the individual, that did the remediation.

rem_remediatedBy
Data type: string
Cardinality: zero or more

Entire Document Remediated?—*Required*

Designates whether the whole resource was remediated or a portion.

Controlled vocabulary:

- True [true]
- False [false]

rem_complete

Data type: True or False value

Cardinality: one

Part(s) Remediated—*Required*

Designates which portions of the resource were remediated, e.g., "Chapters 1–3," "Appendix A," "Pages 27–64," "pp. 17–32," etc.

If the entire resource was remediated, enter "All."

rem_coverage

Data type: string

Cardinality: one or more

Remediated Aspects—*Optional*

Lists the components remediated or the types of remediation done.

Controlled vocabulary:

- Scanned and OCRed [scannedAndOcrd]
- Proofread and Corrected OCR [proofreadCorrectedOcr]
- Added Bookmarks [addedBookmarks]
- Tagged PDF [taggedPdf]
- Added Image Descriptions [addedImageDescriptions]
- Structured Headings [structuredHeadings]
- Linked TOC [linkedToc]
- Fixed Tables [fixedTables]
- Added MathML [addedMathML]
- Foreign Language Markup [foreignLanguageMarkup]
- Transcribers Notes [transcribersNotes]
- Annotations [annotations]

rem_remediatedAspects

Data type: RemediatedAspects values

Cardinality: zero or more

Text Quality—*Optional*

Designates the level of accuracy of the text based on how it was obtained and what was done to improve it, if anything.

Controlled vocabulary:
- Raw OCR [rawOcr]
- Cleaned OCR [cleanedOcr]
- Rekeyed [rekeyed]
- Proofread [proofread]
- Published [published]

rem_textQuality
Data type: TextQuality value
Cardinality: zero or one

Remediation Status—*Required*

Designates whether the resource has been remediated or not, or whether it was considered born accessible (meaning that the standard product available to everyone is sufficiently accessible from the start).

Controlled vocabulary:
- Not Remediated [notRemediated]
- Remediated [remediated]
- Born Accessible [bornAccessible]

rem_status
Data type: RemediationStatus value
Cardinality: one

Remediation Date—*Optional*

Records when the remediation was completed. Note that this is not necessarily (or even often) the date the remediated resource was deposited in EMMA.

rem_remediationDate
Data type: ISO 8601, YYYYMMDD
Cardinality: zero or one

Remediation Comments—*Required*

This is a free-text field where the DSO records information that will be useful to other DSOs to know about the remediated resource.

rem_comments
Data type: string
Cardinality: one

NOTE: The following six properties are specified by schema.org and the W3C EPUB Accessibility 1.1 recommendation, including the controlled vocabularies shown here. Those should be considered authoritative, including companion documents from the W3C.

Accessibility Features—*Optional*

The features in the remediated resource that enhance its accessibility.
Controlled vocabulary:

- Alternative Text [alternativeText]
- Annotations [annotations]
- Audio Description [audioDescription]
- Bookmarks [bookmarks]
- Braille [braille]
- Captions [captions]
- ChemML [ChemML]
- Described Math [describedMath]
- Display Transformability [displayTransformability]
- Display Transformability – background color [displayTransformability/backgroundColor]
- Display Transformability – color [displayTransformability/color]
- Display Transformability – font-height [displayTransformability/font-height]
- Display Transformability – font-size [displayTransformability/font-size]
- Display Transformability – line-height [displayTransformability/line-height]   *(cont'd on next page)*

- Display Transformability – word-spacing [displayTransformability/word-spacing]
- High-contrast Audio [highContrastAudio]
- High-contrast Display [highContrastDisplay]
- Index [index]
- Large Print [largePrint]
- LaTeX [latex]
- Long Description [longDescription]
- MathML [MathML]
- Physical Object [physicalObject]
- Print Page Numbers [printPageNumbers]
- Reading Order [readingOrder]
- Ruby Annotations [rubyAnnotations]
- Sign Language [signLanguage]
- Sound [sound]
- Still Image [stillImage]
- Structural Navigation [structuralNavigation]
- Synchronized Audio Text [synchronizedAudioText]
- Table of Contents [tableOfContents]
- Tactile Graphic [tactileGraphic]
- Tactile Object [tactileObject]
- Tagged PDF [taggedPDF]
- Timing Control [timingControl]
- Transcript [transcript]
- TTS Markup [ttsMarkup]
- Unlocked [unlocked]

s_accessibilityFeatures
Data type: A11yFeature values
Cardinality: zero or more

Accessibility Controls—*Optional*

List the modes in which the entire content of the resource can be consumed.

Controlled vocabulary:
- Full Audio Control [fullAudioControl]
- Full Keyboard Control [fullKeyboardControl]
- Full Mouse Control [fullMouseControl]
- Full Touch Control [fullTouchControl]
- Full Video Control [fullVideoControl]
- Full Switch Control [fullSwitchControl]
- Full Voice Control [fullVoiceControl]

s_accessibilityControl

Data type: A11yControl values

Cardinality: zero or more

Accessibility Hazards—*Optional*

Documents the extent to which the resource presents perceptual hazards.

Controlled vocabulary:
- Flashing Hazard [flashing]
- No Flashing Hazard [noFlashingHazard]
- Motion Simulation Hazard [motionSimulation]
- No Motion Simulation Hazard [noMotionSimulationHazard]
- Sound Hazard [sound]
- No Sound Hazard [noSoundHazard]

s_accessibilityHazard

Data type: A11yHazard values

Cardinality: zero or more

Access Mode—*Optional*

Lists the modes in which the content of the resource, or parts of it, can be consumed.

Controlled vocabulary:
- Auditory [auditory]
- Chart On Visual [chartOnVisual]   *(cont'd on next page)*

- Chem On Visual [chemOnVisual]
- Color Dependent [colorDependent]
- Diagram On Visual [diagramOnVisual]
- Math On Visual [mathOnVisual]
- Music On Visual [musicOnVisual]
- Tactile [tactile]
- Text On Visual [textOnVisual]
- Textual [textual]
- Visual [visual]

s_accessMode

Data type: A11yAccessMode values
Cardinality: zero or more

Sufficient Mode—*Optional*

The access modes (one or more) that enable the entire resource to be consumed in each specified mode.
Controlled vocabulary:
- Auditory [auditory]
- Tactile [tactile]
- Textual [textual]
- Visual [visual]

s_accessModeSufficient

Data type: A11ySufficiennt values
Cardinality: zero or more

Accessibility Summary—*Optional*

Text describing any accessibility issues not covered by the other metadata.

s_accessibilitySummary

Data type: string
Cardinality: zero or more