Standards - Metadata

There are two ways that metadata accompanies a publication. In the first are digital publication formats that directly embed accessibility metadata (EPUB and PDF). In the second are external metadata record formats (ONIX and MARC) that accompany a digital publication as it moves through the supply chain. In some cases, a digital publication may include both internal and external metadata (e.g., an EPUB could have accessibility metadata in it package document and also be provided to a vendor with an ONIX record).

Source: W3C Metadata Guidelines

Below are some short descriptions of EPUB and PDF metadata, and the main metadata formats of ONIX and MARC, with some additional formats too.

There is a useful guide to metadata available for a more strategic overview: Open Book Collective - Metadata Management. The Publishing CG Accessibility TaskForce at W3C have plans to produce guidance on accessibility metadata in 2026.

EPUB

Accessibility metadata is embedded into your EPUB file's OPF metadata. This MUST include the following accessibility metadata:

- accessMode — a human sensory perceptual system or cognitive faculty necessary to process or perceive the content (e.g., textual, visual, auditory, tactile).
- accessibilityFeature — features and adaptations that contribute to the overall accessibility of the content (e.g., alternative text, extended descriptions, captions).
- accessibilityHazard — any potential hazards that the content presents (e.g., flashing, motion simulation, sound).

EPUB publications SHOULD include the following schema-org accessibility metadata:

- accessibilitySummary — a human-readable summary of the accessibility that complements, but does not duplicate, the other discoverability metadata. The summary also describes any known deficiencies (e.g., lack of extended descriptions, specific hazards).
- accessModeSufficient — a set of one or more access modes sufficient to consume the content without significant loss of information. An EPUB publication can have more than one set of sufficient access modes for its consumption depending on the types of content

it includes (i.e., unlike access modes, this property takes into account any alternatives for content that is not broadly accessible, such as the inclusion of transcripts for audio content).

EPUB creators MAY include additional schema-org accessibility metadata.

There is some companion guidance on Fixed Layout EPUBs and some guidance on techniques for extracting information from EPUB Accessibility Metadata

PDF

The PDF/UA standard defines how to describe accessibility metadata within it. The PDF standard itself can be difficult to understand and translate into actions, and so it can be easier to look at the error checking procedures to determine what is required. Completing this shows that PDFs can be not compliant with the PDF/UA standard through not including standard metdata (such as title and language) rather than metadata about accessibility. The open source tool PAC checks for these:

XMP Metadata

Adobe's Extensible Metadata Platform (XMP) is a file labelling technology that lets you embed metadata into files themselves during the content creation process. Essentially, PAC is checking for the presence of XMP metadata such as title, author, subject, keywords and language.

PDF/UA Identifier

Here, PAC is checking that there is file metadata to confirm compliance with the PDF/UA standard or not.

Title in XMP Metadata

Here, PAC is letting you know that there is nothing within the dc:title field.

The Matterhorn Protocol identifies these metadata errors, which are similar with the addition of a quality check of the document's title and a blank indicator of the document's language.

Checkpoint 06: Metadata

06-001 Document does not contain an XMP metadata stream

06-002 The XMP metadata stream in the Catalog dictionary does not include the PDF/UA identifier.

06-003 XMP metadata stream does not contain dc:title

06-004 dc:title does not clearly identify the document

Checkpoint 11: Declared Natural Language

11-006 Natural language for document metadata cannot be determined.

ONIX

ONIX is an XML-based standard for rich book metadata, providing a consistent way for publishers, retailers and their supply chain partners to communicate a wide range of information about their products. An ONIX record is a separate XML file that is sometimes packaged together with an ebook, and sometimes left separate, but either way, it is distributed alongside an ebook. It contains all kinds of metadata about a book, like title, author, edition, page count, etc. and a set of accessibility metadata.

Most ONIX Accessibility metadata is carried in the data element. This uses ONIX codelist 196 to specify particular accessibility options that are provided by the product, which functions as a granular description of the accessibility features of the e-book, and can also specify the e-book's conformance with accessibility standards and provide links to further detail. Additionally the relevant codes from codelist 81 are important to highlight content types in the e-book (text, images, audio etc) that may require mitigations for potential inaccessibility.

There is some guidance on techniques for extracting information from ONIX Accessibility Metadata for display, and some additional OCLC advice for ONIX providers that includes guidance on accessibility metadata.

MARC

MAchine-Readable Cataloging (MARC) standards are a set of digital formats for the description of items catalogued by libraries, such as books. MARC 21 was designed to redefine the original MARC

record format for the 21st century and to make it more accessible to the international community. MARC 21 has formats for the following five types of data: Bibliographic Format, Authority Format, Holdings Format, Community Format, and Classification Data Format.

Within the Bibliographic Format, there are specific fields to include accessibility metadata.

341 - Accessibility Content

532 - Accessibility Note

There is a Crosswalk between ONIX and MARC available.

BIBFRAME

Bibliographic Framework (BIBFRAME) was designed to replace the MARC standards, and to use linked data principles to make bibliographic data more useful both within and outside the library community. Bibframe includes the property Contentaccessibility

Schema.org

Schema.org is an initiative launched in 2011 by operators of the world's largest search engines at the time to create and support a common set of schemas for structured data markup on web pages. It includes the CreativeWork type 'Book' and includes several standard accessibility tags, including: AccessibilityFeature, AccessibilitySummary and AccessibilityHazard, plus others.

There are no accessibility sections associated with the following metadata standards

- Dublin Core
- BibTex
- DataCite
- CrossRef
- KBART
- OPDS

W3C have recently published a guide to displaying accessibility metadata: [Accessibility Metadata Display Guide for Digital Publications 2.0]

Revision #16
Created 2025-09-11 13:58:30 UTC by Jo Fitzpatrick
Updated 2026-03-03 14:02:46 UTC by Jo Fitzpatrick