TECHNICAL SPECIFICATION

ISO/IEC TS 22424-2

First edition
2020-01

Digital publishing — EPUB3 preservation —

Part 2: Metadata requirements



Reference number
ISO/IEC TS 22424-2:2020(E)

© ISO/IEC 2020



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2020

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Abbreviated terms ... **2**
5 Syntax ... **2**
6 Packaging metadata ... **4**
6.1 General ... 4
6.2 Package creator / submitter information ... 4
6.3 Package status ... 5
6.4 Package identifier ... 5
6.5 Work and publication identifiers ... 6
6.6 Core media type resource identifiers ... 8
6.7 Foreign resource identifiers ... 9
6.8 Identifiers for metadata records ... 10
6.9 Dates ... 11
6.9.1 General ... 11
6.9.2 Creation date of a submission information package ... 12
6.9.3 Modification date of a submission information package ... 12
6.9.4 Creation/modification date of an EPUB publication ... 12
6.9.5 Creation/modification of a metadata record ... 13
6.10 Metadata format and its versions ... 13
7 Administrative metadata ... **15**
7.1 General ... 15
7.2 Technical metadata ... 16
7.2.1 File formats and their versions ... 16
7.2.2 Digital signatures and checksums ... 19
7.3 Rights metadata ... 20
7.3.1 General ... 20
7.3.2 Preservation related rights ... 21
7.4 Structural metadata ... 22
7.5 Preservation metadata ... 24
8 Structure of submission information packages ... **26**
9 Content of submission information packages ... **27**
Annex A (informative) **Digital signature** ... **29**
Annex B (informative) **Events** ... **31**
Bibliography ... **35**

© ISO/IEC 2020 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 34, *Document description and processing languages*.

A list of all parts in the ISO/IEC TS 22424 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

© ISO/IEC 2020 – All rights reserved

Introduction

This document facilitates the long-term preservation of EPUB publications by specifying metadata elements which are required or recommended for long-term preservation (such as identifiers) and the ways in which the EPUB publication and related metadata can be packaged. EPUB versions 3 and 3.0.1 are covered; if necessary, the EPUB version applicable is specified.

Long-term preservation in general requires two things:

— making the object such as EPUB publication fit for preservation – including features to be used and feature to avoid;

— packaging the object (and any metadata related to it) together with any additional data such as other versions of the object and other documentation into an Open Archival Information System (OAIS) submission information package (SIP).

ISO/IEC TS 22424-1 concentrates on the archivability of EPUB documents.

The background to this document comes from the Open Archival Information System, which is described in ISO/IEC TS 22424-1.

When a submission information package (SIP) is formed, mandatory preservation metadata need to be present in the package. Depending on the agreements made between the producer and the archive, metadata elements are stored either in the container document or the EPUB publication itself, or both. Usually an archive would expect to find all relevant metadata in the container, unless the submission agreement allows embedding of metadata into EPUB publications.

This document does not require any changes to be made to the current of future EPUB standards. However, when an EPUB publication is created or modified for submission to an archive, there are some EPUB features that should be used and others that should be avoided. ISO/IEC TS 22424-1 describes how the EPUB format should be applied. This document concentrates on mandatory and recommended metadata elements needed for the long-term preservation of EPUB publications and their METS encoding. ISO/IEC TS 22424-1 recommends the usage of METS but allows also other container standards; this document concentrates on preservation metadata and its METS encoding in SIPs. Future editions of these documents may specify other encodings such as BITS (Book Interchange Tag Suite)[1)].

In order to guarantee access to documents, OAIS archives may migrate documents into new file formats when the original formats are no longer supported by commonly used rendering tools. If the document to be migrated is an e-book in an outdated EPUB format, migration can be made to a more modern version of EPUB or, at least in principle, to another e-book format.

Generally, migration into another file format should be straightforward if the current and new format are compatible and there are efficient and reliable migration tools available. If the target format is a more modern version of the current format, compatibility should not be a problem. But if a format is rich, migration tools may not be able to render all the properties of a resource.

This document applies to EPUB versions 3 and 3.0.1. Earlier versions (EPUB 2 and 2.0.1) are not covered. Since there are no implementations of version 3.1, it is not covered in this document either. EPUB 3.2 was published in May 2019[2)]. It will be taken into account in the next edition of this document.

This document does not cover issues related to migration between EPUB versions or from EPUB to other e-book formats. Migration to other formats is often lossy; this applies to e-book formats as well, since there are EPUB features which are not supported in other e-book formats, and vice versa. Moreover, even if the same feature is supported, technical implementations can be incompatible. For instance, if an EPUB 3 publication using fixed layout is migrated to Amazon's KF8 format, preserving fixed layout properties requires special attention since there are significant technical differences between these formats in how this feature has been implemented.

1) https://www.loc.gov/preservation/digital/formats/fdd/fdd000453.shtml

2) https://w3c.github.io/publ-epub-revision/epub32/spec/epub-spec.html

© ISO/IEC 2020 – All rights reserved

Sometimes migration cannot be applied at all; programs cannot be migrated without access to and good understanding of the source code. In such cases long-term preservation is possible only if the OAIS archive responsible is able to emulate either the program's original hardware or software environment.

Within the preservation community, emulation is considered to be a viable option for some content. For the time being there is no full understanding on how emulation will function in the long-term, but this may change with emulation as a service approach coming to the market.

Metadata requirements in this document are based on the migration of file formats. Emulation is not covered (just a single example of emulation-related preservation metadata is given), although emulation is likely to be the best preservation method for fixed layout EPUB publications and interactive EPUB publications. Preservation metadata requirements for emulation-based preservation strategy may be added into a future version of this document.

Supporting emulation might require just information about appropriate tools in the submission agreement or in the related documentation. A more sustainable approach is to include a description of the emulation environment (hardware and/or software) in the premis:object section of the PREMIS metadata record in the SIP. During ingest this information is copied into the archival information package (AIP). If migration is used, hardware and software environments needed for rendering the versions of the document in the AIP can be specified separately as access environments.

Ambition level of migration may vary. Usually it is to preserve the intellectual content, since retaining also the original look and feel of preserved documents is considered to be too demanding. If semantics and layout are interlinked, it is important to keep also the original EPUB publication in order to facilitate preservation of the semantics via emulation-based access to the original content.

Migration both requires and produces preservation metadata. For instance, staff in the archives has to figure out which tools can be used to carry out the migration, and what weak points they may have. The intention of the preservation community is to maintain this information in format libraries such as PRONOM[3)] . When a new AIP is created after a migration, the package should contain both the old and the new representation of the migrated document and preservation metadata describing the migration event and the possible differences between the document versions[4)]. Depending on their needs and archived resources archive users can then make a choice between the original, which is authentic but possibly difficult to render, and the migrated document, which should be easy to use but less authentic. In practice, finding access software to outdated versions of preserved documents may be difficult. The OAIS archive, on the other hand, can migrate the original document again when better tools can be used, or if there are significant issues in migrated documents.

Metadata elements that need to be included in SIPs are a priori essential for digital preservation. For instance, if there is no digital signature present and a secure transfer channel has not been used, it is impossible to guarantee the information entering the archive has not changed during transfer or that it is coming from a correct source. Moreover, if the data has already been tampered with before it enters the archive, all subsequent preservation actions may be useless.

This document does not specify generic conformance requirements for EPUB publications, but may make some restrictions to the use of EPUB specifications. The generic conformance requirements made in the EPUB Contents Documents Specification apply to EPUB publications in SIPs as well.

ISO/IEC TS 22424-1 defined a set of requirements for archivable EPUB publications. Please consult ISO/IEC TS 22424-1 for more information.

---

3) http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

4) This document is only concerned with those metadata elements which are to be included in SIPs. Preservation metadata needed in AIPs (which describes the preservation related events such as migration) is beyond the scope.

 © ISO/IEC 2020 – All rights reserved

Digital publishing — EPUB3 preservation —

Part 2:
Metadata requirements

1 Scope

The ISO/IEC TS 22424 series supports long-term preservation of EPUB publications via a dual strategy. This document makes EPUB compliant with current practices of Open Archival Information Systems (OAIS) archives and technical requirements of repository systems. The former tend to rely on OAIS in their operations; the latter prefer to ingest electronic documents only in containers conforming to standards such as METS (Metadata Encoding and Transmission Standard).

ISO/IEC TS 22424-1 considers EPUB features from a long-term preservation point of view.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8601 (all parts), *Date and time — Representations for information interchange*

ISO/IEC TS 22424-1, *Digital publishing — EPUB3 preservation — Part 1: Principles*

METS *Metadata Encoding & Transmission Standard. Version 1.12.1.* [online]. Library of Congress, 2019. Available from: https://www.loc.gov/standards/mets/

PREMIS *PREMIS Data Dictionary for Preservation Metadata. Version 3.0.* [online]. Library of Congress, 2015. Available from http://www.loc.gov/standards/premis/

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC TS 22424-1 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
data dictionary
organized and constructed (electronic data base) compilation of descriptions of data concepts that provides a consistent means for documenting, storing and retrieving the syntactical form (i.e. representational form) and the meaning and connotation of each data concept

Note 1 to entry: PREMIS is a data dictionary. PREMIS Data Dictionary for Preservation Metadata (https://www.loc.gov/standards/premis/) is a leading metadata specification for metadata needed for long-term preservation.

[SOURCE: ISO 24531:2013, 4.14, modified — Note 1 to entry has been added.]

© ISO/IEC 2020 – All rights reserved