International Standard

ISO 20271-1

First edition
2026-10

Document management — Reference model for long-term preservation of textual documents —

Part 1: Overview

Gestion des documents — Modèle de référence pour une conservation à long terme des documents textuels —

Partie 1: Vue d'ensemble

Reference number
ISO 20271-1:2026(en)

© ISO 2026



COPYRIGHT PROTECTED DOCUMENT

© ISO 2026

All rights reserved.

This ISO publication is protected by copyright and is owned by ISO and/or its licensors.

The content of this ISO publication is provided under licence, not sold. Use is subject to the applicable licence terms issued by ISO, an ISO member body, or an authorized third-party distributor.

Except as required for implementation or expressly permitted by a separate licence, no part of this ISO publication may be reproduced, distributed, modified, used, or made available in any form or by any means – electronic or mechanical, including photocopying, scanning, recording, or posting on internal or external digital platforms.

Any use beyond the scope of the granted rights is prohibited and may result in legal action.

ISO copyright office
CP 401 • CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
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 Summary of the ISO 20271 series **2**
4.1 General 2
4.2 Summary of ISO 20271-2: Fundamental 3
4.3 Summary of the ISO 20271-3: Taxonomy and markup 3
4.4 Preservation assessment methodology 3
5 Applicable document types for ISO 20271 series **3**
6 Exploring applicability and extension to additional document types **4**
7 Audience of the ISO 20271 series **4**
Bibliography **5**

© ISO 2026 – All rights reserved

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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 ISO 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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 Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 2, *Document file formats, EDMS systems and authenticity of information*.

A list of all parts in the ISO 20271 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 2026 – All rights reserved

Introduction

Digital document files vary greatly in type and complexity. They range from simple binary stream files to text-based files readable by humans, machine-interpretable files, and files that, while human-readable, require interpretation and re-expression through complex logical structures and relationships.

Textual document file formats can take various forms, including simple binary or text-based structures and more complex packaged formats that encapsulate multiple components, such as content, metadata and resources, within a single container structure. Many modern document formats adopt such packaged formats.

Ensuring long-term preservation and accessibility of these digital documents can be challenging, primarily due to technical obsolescence, format incompatibilities, or software/hardware dependencies, such as:

— file format compatibility issues:

  — file formats are no longer supported;

  — new software fails to open correctly files created by the older versions of the same software (no backward compatibility);

  — reliance on proprietary formats rather than on standardized open formats can increase the risk of format obsolescence and make preservation efforts more challenging e.g. due to inaccessibility of detailed specifications, unavailability of usable software for handling the formats, or IPR-related issues;

— limited adoption and support of formats:

  — formats are not widely used or supported in the relevant domain;

  — formats lack an active user community or support from multiple vendors;

NOTE 1 The use of formats that are widely adopted, based on open specifications, and supported by multiple vendors can reduce the risk of obsolescence.

— software and hardware dependencies:

  — files can only be opened on specific operating systems or software versions;

  — compression algorithms, encryption technologies or cryptography-based features used (including digital signatures, seals and time stamps) are no longer supported;

— lack of preservation policies:

  — absence of policies or processes for the long-term preservation of document files;

  — insufficient ongoing monitoring of file formats and storage media management.

NOTE 2 Other operational issues such as media ageing, physical corruption, environmental factors, and backup planning are outside the scope of the ISO 20271 series.

The ISO 20271 series defines technical methods for the selection, verification and assessment of the subjects of technical review concerning the obsolescence of textual documents. It provides a reference model and framework for understanding textual documents and the properties that can be present within them. The reference model addresses technical issues and provides assessment criteria to determine which digital components (layers) or metadata are desirable or necessary for long-term preservation, as well as to evaluate the suitability format for this purpose. To support this, the ISO 20271 series is structured as follows (see Figure 1):

— This document (ISO 20271-1) defines the overall framework, scope and principles for the technical review of textual document formats in the context of obsolescence. It also provides an explanation of the structure and scope of the ISO 20271 series to enhance users' understanding.

© ISO 2026 – All rights reserved

— ISO 20271-2 defines a reference model for textual document formats that can be used to analyse a textual document file format. It also provides recommendations for preserving each layer of the format, including content, metadata and presentation, ensuring long-term accessibility and integrity of the document.

— ISO 20271-3specifies properties that can exist in a textual document and provides reference markup for representing those properties.

The ISO 20271 series is intended for use by organizations and practitioners involved in the creation, management and preservation of textual documents. This includes developers of document formats, system architects, archivists, and other stakeholders responsible for ensuring long-term accessibility and sustainability.

The ISO 20271 series supports activities such as the analysis of textual document formats, the identification and classification of document properties, the definition of preservation requirements, and the assessment of long-term preservation risks.

In practice, the ISO 20271 series is intended to be applied as a coherent and iterative process.

ISO 20271-2 provides a reference model for analysing textual document formats by identifying layers and associated properties.

ISO 20271-3 builds upon this model by defining a taxonomy for classifying these properties and by providing non-prescriptive examples of how such properties can be represented.

Together, these parts support a structured approach to the analysis, description, and evaluation of textual document formats in the context of long-term preservation.

Within the reference model, characteristics of textual documents are expressed as properties.

These properties may be associated with one or more layers and form the basis for classification, analysis, and preservation assessment across the ISO 20271 series.

© ISO 2026 – All rights reserved

Document management — Reference model for long-term preservation of textual documents —

Part 1: Overview

1 Scope

This document provides an overview of the ISO 20271 series.

The following are covered in this document:

— the scope and structure of the ISO 20271 series;

— description of textual document files within the scope of the ISO 20271 series;

— a summary of the content of each part in the ISO 20271 series;

— guidelines for the extension of other types of document formats.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

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

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

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

3.1
packaged format
format that encapsulates text, images, tables, and other related objects into a single container file, enabling their combined storage, transfer, and preservation as a single unit

EXAMPLE Office Open XML (.docx), OpenDocument Text (.odt), EPUB.

3.2
semi-structured data
quality indicating that the form of an entity does not follow a rigid or fully predefined structure but exhibits some level of organization

Note 1 to entry: Examples of semi-structured data include the data that contain HTML tags or other markers to separate semantic elements and to represent hierarchies of records and fields within the data.

© ISO 2026 – All rights reserved