



DPC: 25/30471718 DC

BSI Group Headquarters

389 Chiswick High Road London W4 4AL

Tel: +44 (0)20 8996 9000

Fax: +44 (0)20 8996 7400

www.bsigroup.com

Date: 22 April 2025

Origin: International

Latest date for receipt of comments: 15 June 2025

Project No. 2023/00655

Responsible committee: IDT/1 Document Management Applications

Interested committees:

Title: Draft BS ISO 20271-1 BS ISO 20271-2 Document management. Reference model for long-term preservation of textual documents. Part 2: Fundamentals.

Please notify admin.start@bsigroup.com if you are aware of any keywords that might assist in clarifying or identifying the standard or of the content of this standard:

- i) has any issues related to 3rd party IPR, patent or copyright
- ii) affects other national standard(s)
- iii) requires additional national guidance or information

WARNING: THIS IS A DRAFT AND MUST NOT BE REGARDED OR USED AS A BRITISH STANDARD. THIS DRAFT IS NOT CURRENT BEYOND 15 June 2025

This draft is issued to allow comments from interested parties; all comments will be given consideration prior to publication. No acknowledgement will normally be sent. **See overleaf for information on the submission of comments.**

No copying is allowed, in any form, without prior written permission from BSI except as permitted under the Copyright, Designs and Patent Act 1988 or for circulation within a nominating organization for briefing purposes. Electronic circulation is limited to dissemination by e-mail within such an organization by committee members.

Further copies of this draft may be purchased from BSI Shop <http://shop.bsigroup.com> or from BSI Customer Services, Tel: +44(0) 20 8996 9001 or email cservices@bsigroup.com. British, International and foreign standards are also available from BSI Customer Services.

Information on the co-operating organizations represented on the committees referenced above may be obtained from <https://standardsdevelopment.bsigroup.com/>.

E-mail: admin.start@bsigroup.com

Introduction

This draft standard is based on national and international discussions. Your comments on this draft are invited and will assist in the preparation of the consequent standard.

For international standards, comments will be reviewed by the relevant UK national committee before sending the consensus UK vote and comments to the international committee, which will then decide appropriate action. If the international standard is approved, it is usual for the text to be published as a British Standard.

For national standards, comments will be reviewed by the relevant UK national committee and the resulting standards published as a British Standard.

UK Vote

Please indicate whether you consider the UK should submit a negative (with supporting technical reasons) or positive vote on this draft. Please indicate if you are aware of any reason why this draft standard should not be published as a British Standard.

Submission of Comments

- **Annotated drafts are not acceptable and will be rejected.**
- All comments should be submitted online at <https://standardsdevelopment.bsigroup.com/>. You will need to register in order to comment.

Template for comments and secretariat observations

Date: xx/xx/20xx	Document: ISO/DIS xxxx
------------------	------------------------

1	2	(3)	4	5	(6)	7
MB	Clause No./ Subclause No./Annex (e.g. 3.1)	Paragraph/ Figure/Table/Note	Type of comment	Comment (justification for change) by the MB	Proposed change by the MB	Secretariat observations on each comment submitted
	3.1	Definition 1	ed	Definition is ambiguous and needs clarifying.	Amend to read '...so that the mains connector to which no connection...'	
	6.4	Paragraph 2	te	The use of the UV photometer as an alternative cannot be supported as serious problems have been encountered in its use in the UK.	Delete reference to UV photometer.	



DRAFT International Standard

ISO/DIS 20271-2

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

Part 2: Fundamentals

ICS: 37.080; 35.240.30

ISO/TC 171/SC 2

Secretariat: **ANSI**

Voting begins on:
2025-04-22

Voting terminates on:
2025-07-15

This document is circulated as received from the committee secretariat.

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENTS AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

Reference number
ISO/DIS 20271-2:2025(en)

© ISO 2025  **Documents**
Document Sharing Hub



COPYRIGHT PROTECTED DOCUMENT

© ISO 2025

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
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	2
3 Terms and definitions	2
4 Textual documents	4
5 Reference model for textual documents	7
5.1 Purpose	7
5.2 Applicability	7
5.3 Rationale	8
5.4 Multi-Layered reference model	9
5.4.1 Definition of the layers of the reference model.	10
5.4.2 Definition of property types of each layer	12
5.4.3 Recommendations for assessing each layer	13
6 Target documents for applying reference model	16
6.1 Type of document for the reference model	16
6.2 Types of content included in textual document	16
6.2.1 Types of content	16
6.2.2 Text	17
6.2.3 Image	17
6.2.4 Table	19
6.2.5 Domain-specific notations	19
6.2.6 Review and Comment	21
6.2.7 Other Types	22
Bibliography	23

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 documents 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 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).

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 Application*, Subcommittee SC 2, *EDMS systems and authenticity of information*.

A list of all parts in the ISO 20271 series can be found on the ISO website.

ISO 20271 series consists of the following parts, under the general title *Document management — Reference model for long-term preservation of textual documents*:

- *Part 1: Overview*
- *Part 2: Fundamentals*
- *Part 3: Implementation*
- *Part 4: Assessment*

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.

Introduction

Over time, various file formats have been created and eventually phased out, leading to a situation where files stored in obsolete formats become inaccessible. This occurs due to the disappearance of the technologies and standards upon which these formats were based, coupled with inadequate preservation efforts or updates to current technologies. As a result, digital files produced several decades ago are rendered inaccessible, as there is no longer sufficient information about the file structure, making the data unreadable or unanalysable. This issue, which affects files created many decades ago, has led to significant discussions among countries and organisations on the long-term preservation of digital files, underlining its importance as a critical issue in the field of digital archiving.

In the ISO 20271 series, this document defines a reference model for textual documents, which incorporates multiple abstraction layers for technical analysis and quantitative evaluation. It specifies the definitions for these layers and the categorisation of properties contained within them. When storing properties at each layer according to specific file formats, the document establishes technical criteria to ensure that appropriate measures are in place to address potential obsolescence of the preserved files. It defines specific review targets within the file to assess the long-term preservation capability of storage formats for digital documents. Furthermore, it provides guidance on evaluating the long-term preservation of storage standards, designing new file formats for textual documents, and adding new properties to existing textual document standards. This document also presents considerations for referencing and addressing when improving these standards.

This document supports the following activities:

- Format analysis activities for selecting and preparing the evaluation of formats for the long-term preservation of textual document file formats.
- Technical activities for selecting design targets and performing structural design when developing new textual document format specifications.
- Activities related to adding specific properties or making structural improvements to existing textual document format specifications.
- Classification activities concerning textual document formats, including the addition of specific properties or structural improvements to existing specifications.

For information related to other parts of the ISO 20271 series, the ISO 20271-1 document can be referred to. ISO 20271-1 provides an overview of the roles, interrelationships between parts, and the scope of the entire ISO 20271 series.

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

Part 2: Fundamentals

1 Scope

This document specifies the reference model for textual documents and provides detailed recommendations necessary to support long-term preservation from various perspectives, based on the reference model.

ISO 20271-2 defines the fundamental concepts of the reference model for textual documents. This includes the definitions of layers that make up the reference model, elements incorporated within textual documents, property types, classifications of properties by type, and various properties inherent to textual documents.

Additionally, it defines the concepts and structure of a long-term preservation reference model for digital documents, which can be applied to other types of documents beyond textual documents.

ISO 20271-2:

- defines textual documents and outlines major content properties that consist of a textual document.
- provides the concepts of the reference model for textual documents, defines key elements included within the textual documents, and outlines recommendations for enhancing long-term preservation based on the reference model.
- provides guidelines for classifying various properties that can be included in textual documents as outlined in ISO 20271-3 by reference model layers, along with examples of classification and guidelines for enhancing long-term preservation post-classification.

ISO 20271-2 does not specify the following:

- specific technical methods for checking whether the properties exist within a specific textual document or not.
- specific technical methods for analysing a textual document format such as DOC, DOCX, ODT, TXT, PDF, etc.
- specific metadata items for the long-term preservation of textual documents.
- required computer hardware or operating system.
- does not recommend specific textual document file formats as suitable for long-term preservation.
- does not recommend any processes, procedures, or management practices associated with long-term preservation, records management.

This document provides technical recommendations for organizations, individuals, and both public and private entities involved in designing digital textual documents, assessing existing file formats, or enhancing file format specifications. Its primary aim is to ensure that these documents remain technically interpretable and understandable despite potential obsolescence, while accommodating various requirements and levels of information. These recommendations are particularly valuable for users who are not fully acquainted with the technical characteristics of file formats or the core content elements of textual documents.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes recommendations 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.

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

ActiveX

deprecated software framework created by Microsoft that adapts its earlier Component Object Model (COM) and Object Linking and Embedding (OLE) technologies for content downloaded from a network, particularly from the World Wide Web

3.2

ASCII (American Standard Code for Information Interchange)

character encoding standard that is abbreviated from American Standard Code for Information Interchange, is a character encoding standard for electronic communication

3.3

ASMO708

7-bit character encoding standard specifically designed for Arabic text

3.4

Big5

character encoding standard that is a Chinese character encoding method used in Taiwan, Hong Kong, and Macau for traditional Chinese characters

3.5

DOCX

file format, especially for Office Open XML documents

3.6

Elements

components included in a textual document

3.7

EUC (Extended Unix Code)

character encoding standard that is a multibyte character encoding system used primarily for Japanese, Korean, and simplified Chinese (characters)

3.8

EUC-KR (Extended Unix Code for Korean)

character encoding standard that is an 8-bit character encoding that utilizes KS X 1001(Korea Industrial Standards) and KS X 1003, a variant of Extended Unix Code (EUC)

Note 1 to entry: As it is a representative completed Korean encoding, it is commonly referred to as 'Wansung'.

3.9

HWPX

file format, especially for word processing documents based on Open Word Processor Markup Language (OWPML), which is used by most public institutions in Republic of Korea and designated as a permitted format for the long-term preservation of official documents