TECHNICAL REPORT

ISO/TR 18492

Long-term preservation of electronic document-based information

Conservation à long terme d'information document basée électronique

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/TR 18492:2005
https://standards.iteh.ai/catalog/standards/sist/92cab356-eaaf-48d7-82fc-
0960f0e86b0f/iso-tr-18492-2005

Contents

Page

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

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.

ISO/TR 18492 was prepared by Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 3, *General issues*.

Introduction

Ensuring the long-term preservation of authentic electronic document-based information is a well-documented and identified problem within many fields of expertise, including archival science, document management, e-commerce, e-governance and technology development. As an additional problem, individuals and organizations charged with the responsibility for ensuring long-term access to authentic electronic document-based information have employed a diversity of strategies designed to achieve this goal.

Although there is a clear need to address the problem of long-term access to authentic electronic document-based information, there is a current lack of harmonized international guidance on these issues. This has led to diverse and, sometimes, incompatible approaches that can give rise to potentially mission-critical problems, regarding the accessibility and/or authenticity of the electronic document-based information being retained.

Acknowledging the generic technological obsolescence problem of computer hardware and software as well as the limited life of digital storage media, this Technical Report provides guidance to storage repositories in providing access to and maintaining authentic electronic document-based information that has been retained for future reference.

The purpose of this Technical Report is to provide a clear framework for strategy development and best practices that can be applied to a broad range of public and private sector electronic document-based information to ensure its long-term accessibility and authenticity.

ISO/TR 18492:2005
https://standards.iteh.ai/catalog/standards/sist/92cab356-eaaf-48d7-82fc-
0960f0e86b0f/iso-tr-18492-2005

Long-term preservation of electronic document-based information

1 Scope

This Technical Report provides practical methodological guidance for the long-term preservation and retrieval of authentic electronic document-based information, when the retention period exceeds the expected life of the technology (hardware and software) used to create and maintain the information.

It takes into account the role of technology neutral information technology standards in supporting long-term access.

This guidance also acknowledges that ensuring the long-term preservation and retrieval of authentic electronic document-based information should involve IT specialists, document managers, records managers and archivists.

It does not cover processes for the creation, capture and classification of authentic electronic document-based information.

This Technical Report applies to all forms of information generated by information systems and saved as evidence of business transactions and activities.

NOTE   Electronic document-based information constitutes the "business memory" of daily business actions or events and enables entities to later review, analyse or document these actions and events. As such, this electronic document-based information is evidence of business transactions that enable entities to support current and future management decisions, satisfy customers, achieve regulatory compliance and protect against adverse litigation. To achieve this goal, this electronic document-based information should be retained and appropriately preserved.

2 Normative references

The following referenced documents are indispensable for the application 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 12651:1999, *Electronic imaging — Vocabulary*

ISO 15489-1, *Information and documentation — Records management — Part 1: General*

ISO/TR 15489-2, *Information and documentation — Records management — Part 2: Guidelines*

ISO/TS 23081-1, *Information and documentation — Records management processes — Metadata for records — Part 1: Principles*

3   Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 12651, ISO 15489-1 and ISO/TR 15489-2 and the following apply.

3.1
authentic electronic document-based information
electronic document-based information the accuracy, reliability and integrity of which are maintained over time

3.2
document-based information
substantive information that can be treated as a unit (e.g. an image, text, spreadsheet, database views)

NOTE   Document-based information is inclusive of, but not necessarily limited to: text, images, tabular data (e.g. a spreadsheet), or any combination thereof.

3.3
document-based information content
substantive content contained in document-based information

3.4
document-based information context
information about the circumstances of electronic document-based information creation, control, use, storage and management, and information about its relationship to other similar material

3.5
document-based information structure
logical and physical attributes of document-based information

NOTE   Logical attributes consist of the logical order, e.g. a hierarchy with identifiable subparts, whereas physical attributes comprise elements, e.g. type font, spacing.

3.6
electronic archiving
storage of electronic information in an independent physical or logical space where the information is protected from loss, alteration and deterioration

NOTE   The information may be used as reliable evidence in the future if it has been protected in this manner.

3.7
long-term preservation
period of time that electronic document-based information is maintained as accessible and authentic evidence

NOTE   This period of time can range between a few years to hundreds of years, depending upon the needs and requirements of the organization. For some organizations, this period of time would be determined by regulatory compliance, legal requirements and business needs. For other oranizations, such as archival repositories holding public records, the period of time required to retain electronic document-based information is usually thought to be hundreds of years.

3.8
metadata
data describing the content (including indexing terms for retrieval), context and structure of electronic document-based information and their management over time

3.9
migration
process of transferring electronic document-based information from one software/hardware environment or storage medium to another environment or storage medium with little or no alteration of structure and no alteration in content and context

3.10
storage repository
storage repository organization or entity charged with the storage and maintenance of authentic electronic document-based information

NOTE   It is recognized that this definition is different from technical definitions of "storage repositories".

3.11
technological obsolescence
displacement of an established technical solution in a marketplace as a result of major technological developments or improvements

4   Symbols and abbreviated terms

ASCII   American Standard Code for Information Interchange

CRC   Cyclical Redundancy Code

HTML   Hyper Text Markup Language

JPEG   Joint Photographic Engineers Group

OCR   Optical Character Recognition

PDF/A-1   Portable Document Format — Archive

SHA-1   Standard Hash Algorithm 1

TIFF   Tagged Image File Format

WORM   Write Once Read Many (times)

XML   Extensible Markup Language

5   Long-term preservation

5.1   General

Increasingly, the proliferation of computer technologies that support the creation, use, storage and maintenance of information, results in private and public sector organizations relying on electronic document-based information as the official evidence of their business activities. Consequently, organizations increasingly face the challenge of ensuring the long-term accessibility of authentic electronic information that was created within reliable and trustworthy information systems and stored on electronic media that might be subject to technological obsolescence that if left uncorrected will make the document-based information irretrievable. The importance of this problem is compounded by the fact that organizations are increasingly conducting activities and transactions where no paper evidence exists.

It is essential, therefore, that organizations develop and apply a well-defined strategy for providing long-term preservation and retrieval of authentic electronic document-based information. Subclause 5.2 defines the elements of such a strategy.

5.2 Goals of a long-term preservation strategy

5.2.1 General

This subclause identifies six key issues that storage repositories should consider when they are developing a long-term preservation strategy.

5.2.2 Readable electronic document-based information

A long-term preservation strategy should ensure that electronic document-based information remains readable into the future. To achieve this, the bit stream comprising electronic document-based information should be accessible on the computer system or device that:

— initially created it or

— currently stores it or

— currently accesses it or

— will be used to store the electronic information in the future.

These four processablity options are predicated on the fact that electronic document-based information stored on digital storage media can become unreadable. There are two primary ways in which this can occur.

One is the result of exposure to hostile storage conditions. All of the media currently used for storing electronic document-based information share a common vulnerability to poor environmental conditions, e.g. fluctuations in temperature and humidity. These adverse conditions either damage the media or accelerate the ageing process. Different types of digital storage media require different levels of controlled storage environment to ensure maximum longevity. Some storage technologies are prone to data corruption through magnetic field interference, dust and environmental contaminants (magnetic storage media), while others (optical storage media) are not as prone to these outside factors and less susceptible to media damage outside tightly controlled storage environments. Regardless of which storage technology is in use, it is important to recognize that all forms of storage media can deteriorate and/or degrade through environmental changes.

The second is that non-readability may occur through media obsolescence, which occurs when a storage device (e.g. a tape or disk) is physically incompatible with the available computer hardware (e.g. a tape or disk drive) and therefore cannot be read. Based on past trends, media obsolescence in the future seems inevitable because advances in storage technology continually introduce changes in the way the electronic document-based information is physically stored (e.g. changes in recording technology, changes in disk drive hardware/software interfaces), the form factor of the storage media and in the way the underlying bit stream of document-based information is physically represented (e.g. error correction codes) or the form factor of the storage media. Consequently, over time, older storage media will become incompatible with subsequently used media.

A long-term preservation strategy should specifically address media obsolescence by establishing procedures for periodically transferring document-based information from older to newer media.

NOTE   Data readability is important along with data formatting. Ensuring that the data are formatted in a fashion (i.e. technology neutral formats) that enables users in the future to process the data, should be taken into consideration.

5.2.3 Intelligible electronic document-based information

A long-term preservation strategy should provide intelligible electronic document-based information. Digital information is only intelligible to a computer if the computer also has access to information describing how to interpret the underlying bit stream. The intelligibility of electronic document-based information, therefore, is a function of information about what the bit stream in fact represents and the processing software's capacity to take appropriate action based on this information.

EXAMPLE   The binary code (1s and 0s) comprising a digital Tagged Image File Formatted (TIFF) image carries no intelligibility in its own right. Rather, the image's file header, which contains information such as byte order and the compression algorithm used, enables a computer (through a combination of its operating system and image software) to display and print the image. Similarly, a word processing document carries metadata that makes it intelligible to word processing software.

5.2.4   Identifiable electronic document-based information

A long-term preservation strategy should provide identifiable document-based information. Identifiable document-based information should be organized, classified and described in such a way that it is possible for users and information systems to distinguish between information objects based upon a unique attribute such as name or ID number. Aggregating electronic document-based information into categories based upon shared attributes can facilitate searching and retrieval. Failure to provide such identification can severely limit searching and retrieval.

5.2.5   Retrievable document-based information

A long-term preservation strategy should provide retrievable document-based information, meaning that discrete information objects (or parts of them) can be retrieved and displayed. Retrievability is typically software-dependent in that it requires keys or pointers that link the logical structure of information objects (e.g. data fields or text strings) to their physical storage location.

Generally, this linkage is found in a database record, file system directory structure, file allocation table, header or label that includes the information required to locate the beginning of an object, to indicate the number of bytes of each component or data element and to establish its physical location on the storage medium.

The interpretation of the logical structure of document-based information is a function of an operating system or device driver in conjunction with a particular application system developed to store, manage and access digital information. The retrievability of information objects is therefore inextricably linked to a device driver, software application, file system or operating system.

Newer generations of file formats that support the readability of older file formats help ensure the ability to retrieve electronic document-based information. Backward compatibility however, can be limited because many software vendors support only certain file formats, while others support all versions of various data formats. An example of this would be support for TIFF, JPEG or HTML formatted data, which include backward compatibility.

5.2.6   Understandable document-based information

A long-term preservation strategy should ensure that document-based information is understandable. In order for electronic document-based information to be understandable, it should convey information to both computers and humans. However, the meaning of discrete document-based information is not determined solely by its content. Rather, meaning is derived from the context of both its creation and its use (i.e. metadata). As such, storage repositories should be aware that ensuring the understandability of electronic document-based information differs sharply from ensuring the understandability of paper documentation. Unlike paper documentation, where their physical characteristics typically convey the context of its creation and use, the context of creating and using electronic document-based information is usually linked logically rather than physically.

EXAMPLE   A series of paper documents regarding a particular transaction may be stapled together or placed in a file folder, whereas electronic document-based information of a similar transaction may exist on multiple media in multiple locations and therefore should be electronically tied together. These logical linkages can include identification of both the business process that led to the transaction as well as the participants in the transaction.

The context of creation and use also involves relationships among other document-based information that has been be captured in a variety of ways, including a reference code in a document profile to the other material dealing with the same issue, or a classification code that links each instance of document-based information relating to the same transaction.

5