TECHNICAL REPORT

ISO/TR 17797

First edition
2014-10-01

Electronic archiving — Selection of digital storage media for long term preservation

Archivage électronique — Sélection d'un support de stockage numérique pour une préservation à long terme

Reference number
ISO/TR 17797:2014(E)

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.

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 on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 171, *Document management applications*, Subcommittee SC 1, *Quality*.

Introduction

A significant proportion of digital information generated by different human activities will need to be stored for a long period of time and in some cases for as long as it is possible. Where 'long-term' is used in this Technical Report, a storage period of not less than the anticipated life of the storage media is assumed.

The media currently used to store digital information for the long-term have not been analysed and manufactured for this purpose but mainly developed to maximize transfer rates, density recording, and access time. All these parameters have to be taken in perspective when long-term preservation is the requirement, not just simple backup purposes.

In general, current information management systems might not be conducive to the satisfactory achievement of long-term preservation. For long-term preservation, there needs to be the development of special resources and complex procedures with often increased costs when compared with 'normal' information management systems (duplication of files, refreshing storage, equipment redundancy, monitoring systems, heavy maintenance, frequent and risky migration, high energy consumption, etc.).

Even when a system is designed for long-term preservation, the day-to-day requirements for access and management of the stored digital information needs to be taken into consideration.

When designing systems for long-term preservation, it is necessary to have specific pathways with the objective of providing qualified storage media on criteria such as reliability and stability; this would ensure that the sustainability of digital information leads to optimize the solution for both long-term preservation and access to digital information.

The context of the requirement for long-term digital preservation needs to establish conditions and recommendations for media that is specially manufactured with a guaranteed potential of stability and reliability.

The main criteria involved in the long-term preservation of digital information can be summarized as follows:

a) intrinsic stability of storage media;

b) stability of physical and/or chemical modifications of media produced by record processing;

c) quality and reliability of recording process;

d) preservation of access path to information and metadata;

e) preservation of access tools (i.e. any special software needed to use digital items that have not been migrated to a long-term or standardized format);

f) quality of information;

- — compliance with format specification;
- — data integrity.

Only the first three criteria from the list above are considered as part of this Technical Report.

It is noted that the objective is not to make rules or specifications for use on information management systems as several International Standards, such as ISO 14641-1, ISO 15489-1, and ISO/TR 15489-2, fill this role.

Electronic archiving — Selection of digital storage media for long term preservation

1 Scope

This Technical Report gives guidelines on a selection of the most appropriate storage media for use in long-term electronic storage solutions. It includes a discussion on magnetic, optical, and electronic storage.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 12651-1, *Electronic document management — Vocabulary — Part 1: Electronic document imaging*

3 Terms and definitions

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

3.1
refreshment

data migration where the media is replaced with equivalent media such that all storage hardware and software functionality is unchanged

Note 1 to entry: Refreshment cycles are based on the predicted life span of the medium.

[SOURCE: ISO 13008:2012]

3.2
migration
process of moving digital information, including their existing characteristics, from one hardware or software configuration to another without changing the format

Note 1 to entry: Migration can also include converting to a more current computing environment, involving changes to hardware/software configurations.

[SOURCE: adapted from ISO 13008]

3.3
storage medium
device on which digital information can be recorded

Note 1 to entry: Device can designate a support, a combined support and media player, or a set recorder.

4 Methodology

The characteristics required for storage media should be clearly established regarding the following criteria:

— reliability;

— anticipated longevity;

— sensitivity to environment (operating or storage), internal and external conditions;

— obsolescence of hardware and software.

A part of this Technical Report describes methods or guidelines which lead to the identification of appropriate media based on criteria specified, such as the following:

— control strategies;

— media evaluation process (procedures, monitoring devices);

— tools for monitoring characteristics (ECC, data verification, etc.);

— defining acceptance criteria level;

— analysis of means (existing or desired) for detecting trends.

On this basis, this Technical Report is structured around the following points:

— requirements for the long-term storage media;

— capability of different types of technologies to store digital information in the long term, including aspects of quality, reliability, and durability;

— definition of criteria considered;

— diagnostic elements to be used.

5 Choice of storage system/media

This clause lists some of the issues related to the different types of media that can be used for the storage of digital information. The choice of media should support the long-term preservation strategy and the architecture of the information management system.

All media/recording systems are at risk of a sudden loss of access to digital information, regardless of the technology, so an information management system should be designed to mitigate this risk of loss of this information.

There are various reasons for loss of access, but the most common is the physical failure of media/drivers. This can be the cause in up to 70 % of cases (30 % drive read instability, 38 % drive failure).

The choice of storage system/media is complex because of their often unpredictable behaviour during their life. Reliability models used by manufacturers can provide estimates that are often not achieved in the operational environment. The reliability of storage media is often given by manufacturers in terms of failure rate or in terms of lifespan; for example, the rated lifetime of a particular media type might range from 10 years to several centuries. However, in practical applications it can be found that the actual lifespan can range from a few months to 20 years, as it can be shown that operational life depends upon the operating environment. This disparity needs to be taken into consideration when deciding between different types of media or different models of the same type of media.

A further factor that should be taken into consideration when choosing storage solutions is the issue of obsolescence. Recent developments in storage systems have resulted in a very rapid increase in functional performance, but this has also lead to the rapid obsolescence of implemented solutions. Moreover, reliability and lifespan are typically not transferable from old systems to newer replacements. With each new advance in technology, much of the knowledge gained by tests or various statistical studies on existing systems will need to be replaced by trials and other methods of estimating the reliability of the new systems. This requirement is not supported by research on media degradation processes which makes it very difficult to establish a model for estimating storage media life (see Bibliography).

Storage solutions should be chosen taking into consideration the following:

— results of acceptance tests;

— traceability of the manufacturing processes;

— quality testing by sampling processes;

— longitudinal monitoring of media/drivers;

— environmental conditions of operational use and storage;

— continuous monitoring on the evolution of supply (hardware and software) in relation to the risk of withdrawal of commercial products.

6 Hard disk drive

6.1 General

Hard disk drive (HDD) technology is well established, and over the past 50 years there have been substantial improvements in data transfer rates and capacity along with a marked decrease in product price.

Hard disk drives are electromechanical devices containing generally aluminium platters that are layered with magnetic recording material. Data are written to and read from the disk by moveable read/write heads which float over the surface of the disk.

Two main risks are associated with the use of hard disk drives as physical carriers.

— They have a short life expectancy and should be replaced every five years.

— This technology is also susceptible to data loss from extended use, powering the disk on or off, physical damage of the drive itself, and sudden disk failure, etc.

Configurations

HDD can be used in information management systems under different configurations: either removable or external drives or integrated systems using redundancy and error correction codes to improve performance and reliability.

— **On-line:** the system configuration maintains a continuous or intermittent solicitation through a permanent connection;

— **Off-line:** not under control of processing unit, physically removed or disconnected and can’t be accessed without human intervention:

— idle or inactive;

— RHDD removable hard disk drive;

— iVDR versatile hard disk;

— external HDD, etc.

6.2 Reasons for failure

Storage systems, and in particular those that have moving parts, are prone to mechanical failure. Failure can occur through a number of factors.

6.2.1 Mechanical failures:

— Heads: contamination, disk contact (crash);
— Arm: resonance, damping;
— Platters: scratch, wear, local defect, warping, magnetic layer defect;
— Motor: spindle motor, bearing, defective rotation, lubricant;
— Exploitation processing; data swap.

6.2.2 Electronic failures (motor driver, controller, buffer, connectors, interface, etc.):

— Main causes: power spikes, electrical surge, static electricity;
— Damaged to integrated components;
— Fail connection (power unit, driver, bus, etc.);
— Servo, memory chips.

6.2.3 External factors:

— Magnetic fields;
— Temperature and humidity;
— Water contact;
— Shocks, vibrations.

6.2.4 Disks off-line, the potential factors identified are

— Magnetic thermal decay of recorded bits;
— Media corrosion;
— Media lubricant oil evaporation;
— Fluid dynamic bearing oil evaporation or degradation;
— Electronic corrosion and degradation.

These archival life factors are all functions of temperature and humidity.

See Reference [54].

6.3 Quality indicators

6.3.1 General

Due to the incredibly precise nature of the technology used in hard disks and the fact that mechanical components are used, it is impossible to guarantee the reliability of even the highest quality disks for more than a few years. The failure rate of hard disks follows the so-called bathtub curve model of failures (see Figure 1): they have a relatively high rate of “premature failure” (or early failure), a period of very low failures for several years, and then a gradual reduction in reliability as they reach the end of their useful life

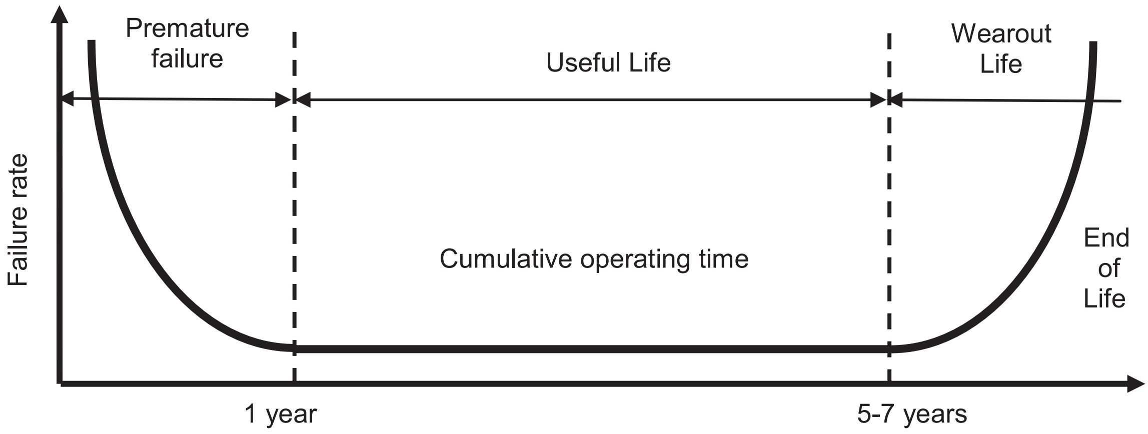

Figure 1 — Failure evolution: the bathtub curve

Most of the time, these are statistics based on systems in operation. However, these results do not give any indication about the new generations because technical similarities do not allow extrapolations. The rapid evolution of hardware provides backward compatibility but cannot predict their behaviour (although we have a good idea of the stability of the supports when they are replaced by the next generation).

6.3.2 Reliability terms

6.3.2.1 General

The disparity of units and values in the specifications provided by the manufacturers concerning the characteristics of reliability should be noted. Examples are:

— MTBF (Mean time before failure);

— MTTF (Mean time to failure);

— AFR (Annualized failure rate);

— MTBDL (Mean time between data loss);

— MTTR (Mean time to repair);

— Useful/Service life.

In addition, reliability data are rarely accompanied by a statement of conditions used for the development of the data.

6.3.2.2 Mean time before failure (MTBF) or Mean time to failure (MTTF)

MTBF is a statistical calculation that projects the average life expectancy of a typical disk in a large population of drives. MTBF does not indicate how long a disk drive will run before it fails, just the probability of failure.

To be interpreted properly, the MTBF figure is intended to be used in conjunction with the useful **service life** of the drive, the typical amount of time before the drive enters the period where failures due to component wear-out increase. MTBF only applies to the aggregate analysis of large numbers of drives; it says nothing about a particular unit.

If MTBF of a model is 500 000 hours and the service life is five years, this means that a drive of that type is supposed to last for five years, and that of a large group of drives operating within this timeframe, on average they will accumulate 500 000 hours of total run time (amongst all the drives) before the first failure of any drive.

There are in fact two different types of MTBF figures:

— ***Theoretical MTBF*** figure for a new drive based primarily upon analysis of historical data of other drives similar to the new one;

— ***Operational MTBF*** derived by analysing field return and comparing them to the installed base.

6.3.2.3 Annualized failure rate (AFR)

Annualized failure rate (AFR) gives the failure probability for a device or component during a full year of use. It is in relation with MTBF.

AFR is calculated from the device failure rate over time observed. Disk resellers are particularly sensitive to this issue since this statistic influences their spare parts strategy.

6.3.2.4 Mean time between data loss (MTBDL)

MTBDL is a statistical figure that attempts to predict how long an array group will operate before suffering a catastrophic failure and data loss. For example data loss occurs when information is written to the failed disk array subsequent to its last backup. RAID technology allows some disk array groups to survive the failure of one or more disks making MTBF less significant. MTBDL is a more meaningful measure of reliability than MTBF or AFR since it considers the performance of the array, not just a sub-component.

6.3.2.5 Mean time to repair (MTTR)

MTTR is the time it takes to repair a failed part. MTTR figures assume that no time elapses between the failure and the beginning of repair operations. That is to say, the appropriate person becomes aware of a bad component the instant it fails, has the necessary replacement parts available, and begins the repair process immediately. Unfortunately, this is not usually a valid assumption. As disk capacity increases, the time to complete the rebuild process grows in importance since it extends the time that the total array is vulnerable to a catastrophic failure.

See Reference [52].

6.3.2.6 Useful/Service life

It is the expected operation time of the HDD given by the manufacturer. This time corresponds to the ending period when failure rate becomes weak and constant. The “probability of survival” R of a HDD model and for a defined Useful life is given by Formula (1):

$$R = e^{-\text{Useful life}/\text{MTBF}} \qquad (1)$$

where

R is the probability of survival, assuming constant failure rate model.

For example, for a model of HDD with a MTBF of 500 000 hours and a useful/service life of five years, the probability of survival of this HDD model is equal to 92 %. If this model is used 10 years the probability of survival decreases to 84 %.