Digital information is volatile, so we need its preservation

April 2022

Stefano Cariolato

Today the digital revolution is almost completed and information of any kind (texts, images, video-clips and TV broadcasts, music and songs, WEB pages) is now recorded and disseminated in digital format rather than with a traditional media (paper, film, magnetic tape), with a change that involved all human activities of any type, both collective and individual.
Moreover, also archived printed records are increasingly transformed into digital format, both for diminishing their volume and faster search procedures, while the new information is only directly generated in the electronic form.
Besides other recordings are also volatile by their very nature,such as e-mails or WEB pages, but they could host information of value in the future and equally can deserve conservation.
Starting with paleolithic graffiti engraved in caves, followed by inscriptions on stones, cuneiform scripts on clay tablets, texts on vellum or silk and papyrus rolls, ending with paper, ever information has been registered in a persistent form visible and immediately readable by a human being knowing that type of writing.
But while a book or a letter can immediately be read even centuries after their writing, if its physical support has withstood the time passing, digital information has a shorter life, even in the absence of deterioration of the used media, because of the same technological development that makes quickly obsolete any recording by irreversibly mutating both the reading hardware and software.
Finally digital recordings are carried out in a great variety of different formats, sometimes incompatible with each other or subject themselves to obsolescence, thus unnecessarily complicating the task of preserving their content.
If humanity does not find the way and the will to preserve the content of digital documents, similarly to what was done in the past with paper documents, gradually the culture generated in each period will decrease until it shall disappear.
A story without a future would create a future without history. As for all content of this volume chapter, much more specific information can be found on the ebook "World without history" (ISBN 9788831627078) of the same author.

Digital Documents obsolescence

From now on, the extended document definition will be used, meaning any form of information recorded on any physical medium. Digital documents, regardless of their specific content (texts, images, movies, music, etc.) have now become the new recording standard: in fact most of the new documents produced are directly recorded in digital form (text files, image and TV formats, WEB pages, music formats) and in most organizations, both private and public, the digitalization of paper documents is underway.
All this brings to the fore the problem of the conservation and archiving of these documents, whose consultation depends forcibly on the availability of hardware and software suitable for their viewing and listening. If this is not possible then the document content is virtually lost, which is defined as digital obsolescence, and in the case of digital documents, it is added to the material obsolescence of the physical support (partial destruction) and the immaterial obsolescence due to the incapacity of understanding it (language now abandoned or unknown, as it was initially in the case of very ancient recording like Egyptian hieroglyphics or Etruscan writings).

The term digital obsolescence therefore refers to the possibility of losing the information entrusted o a digital resource, since the latter is no longer legible for a number of reasons: practically it would be a huge string of bites without any mean to interpret them.
To clarify the issue, let's recall the step needed for reading a digital document, irrespective of it actual nature (text, sound, image, video, graphic rendering, etc.):

- reading the support, that is transforming the physical printed marks of the registration,whatever their nature (mechanic, burned, chemical spot, magnetic, electric), into a stream of bits reproducing the original file. To this aim both a suitable hardware for detecting the physical marks on the support and a software for autocorrecting and rendering the detected signals as bits are needed.

- reading the file, that is transforming the stream of bits in a new form, normally of electric nature and apt to pilot a device (video screen, sound amplifier, paper or 3D printer, actuators, etc.), by which the recorded document can be reproduced for final use or enjoyment by a human being.
  To this task users also have to know which is the meaning and function of every bits of the file stream o they could be the document real content, or control bits for a correct reproduction, or else redundant bits for auto-correction in case of errors o and this descriptive and essential information constitutes what is named the format of the file. There are many different file formats for every type of document (text, image, sound, video, etc.) and operating System, simple or fairly complex, and they sum up now to several hundreds, both private and public. Without the knowledge of the right format the task of reproduction of a file could be like inspecting a black box in a laboratory.

Hardware obsolescence

When a technological innovation carries to the substitution of one type of support with another more efficient (e.g. cassette tape >> CDs> solid state memory), the industry ceases to produce the relative reader, which consequently becomes increasingly rare and eventually it disappears altogether.

The displaying hardware becomes unavailable for the disappearance from industrial production of the player, such as happened for the 8 " floppy disks or ZIP disks and their readers, now unavailable, or for the replacement of music cassettes with music CDs before and today for the replacement of CD players with flash memory readers in the listening equipment currently produced. In addition to this, also player supporting software, needed for reading stored data, is subject to very rapid obsolescence.

Software obsolescence

Entrusting the documents to be conserved in the long run to proprietary software (think of a trivial Microsoft Word file), ties the use of the data stored to the fate of the software itself and the interoperability between versions, often absent for marketing reasons: in fact if the new version of the software is incompatible with the previous one the customer is forced to replace it to keep up-to-date, with a consequent benefit for the sales of the producer, but problems with access to old documents by the user. Another example of software obsolescence is, for example, the abandonment of a particular recording format and the consequent unavailability of programs able to directly decipher the document content, such as Visicalc electronic sheets which can now only be read with the use of an emulator.

Unlike the reading of a paper document, for which only human eyes are needed, in case of a digital document three more components have to be available:

1. a suitable hardware player
2. a software for transforming physical recording in a correct result bits string
3. a reading software for interpreting that string and produce a human visible or audible result

The great change is the fact that our eyes are not any more enough. As for points 1 and 2 nothing can really be done right now other than preserve both document and its original format in a new form, constantly readable through the years with the evolving devices produced by the industry. Players and their internal software will anyway disappear from the market, other than perhaps in private collections or museums.
As for the reading software instead the users can right now begin to correct a dangerous situation and simplify the preservation task, because digital documentation can be divided into:

1) Documents produced with **proprietary** programs and formats, whose destiny in terms of permanence and compatibility depends entirely on the sales policies of a single producer.
They, on the other hand, correspond to some of the most durable and famous software on the market. Two of the most widespread problems they present are:

- the lack of public specifications that describe their structure and format in detail, which prevents others from reconstructing in the future the document content if the producer has abandoned the product with which it has been written,
- the fact that often the successive program versions do not guarantee compatibility with documents produced with their previous versions (backward compatibility). In fact, they tend to evolve rapidly and to be declined in numerous versions for different computer environments, with a sometime deliberate limited backward compatibility.

2) Documents produced with programs and in proprietary formats but with **public specifications**, which allow other suppliers or organizations to develop software that can use them.

3) Documents recorded using formats with **specific open standards**, such as those of international standardization bodies (ISO), whose adoption by users and producers is though hindered by the needs of commercial market protection by large companies and also by States for security reasons in the case of "sensitive" documents.
It is immediate to choose the ISO format as the best for digital preservation, for its standardization and maintenance by an international body that guarantees its market neutrality and longevity.
The same problem arises also for file formats, that desirably should converge toward less variants and above all not covered by proprietary confidentiality. As a matter of fact there are different types of format from the point of view of their internal specification knowability, generally categorized as follows:

- proprietary formats with confidential specifications
- proprietary formats with public specifications
- standard public formats

An increase in the adoption of public formats for new files production is then one of the best choice for the future preservation purpose of digital documents, or at least the use of formats with public specifications.

The preservation of digital documents must therefore take into account all these considerations to defend their content over time, independently of the inevitable technological and market developments.

Support obsolescence

A last hint to the original support obsolescence is worthwhile, because tapes, CDs, DVDs and solid state memory do not preserve their content for ever, but are limited to a certain number of years, depending on:

- type of digital document
- type and quality of the recording
- type and quality of the support
- physical conditions of conservation (temperature, humidity, level of electric and magnetic fields)

Moreover it has to be considered that, contrary to paper text which are partially readable when damaged, for digital document we have not always the same possibility even with the availability of suitable devices and software: if the damage is localized in some particular indexing area of the support (for instance a TOC in a file or a key-frame in a video) reading will be possible only with a physical scanning and the use of some reconstructing software apt to interpret the original document format, that obviously has to be known.
Accordingly the preservation will have to be performed using very efficient support for the document conservation, otherwise this task might uselessly become complicated.
Currently the more probably long lasting support is the Arch disk, that was launched into space by the SpaceX Falcon Heavy with the red Tesla Roadster of Elon Musk, on which the entire scify Galactic trilogy by Isaac Asimov was optically recorded. In fact this small crystalline quartz disk can contain up to 360 terabytes of data (as 76600 DVDs), resist up to a temperature of 1000 degrees and keep the data intact virtually forever. This new technology, developed by Peter Kazansky at the University of Southampton's optoelectronic research centre, uses a five-dimensional laser recording process on crystal or glass; it is the result of one of the current laboratory research aimed at solving the problem of material obsolescence of digital media.
Aside to this also less expensive supports are studied and under development, both in the field of solid state and biological memories. Solid state memories, especially phase change ones, promise noticeable performance in terms of data lifespan, but are still in the development phase and will only reach the market in the future. Even the use of organic memories is promising, but for using their advantages it will be necessary to wait. DNA based data storage can hold hundreds of terabytes per gram, but the durability is still questionable and the costs enormous, beside the problem of indexing them. The major current challenge is the lack of an appropriate combination of storage technology and medium possessing the advantages of both high capacity and long lifetime.
If within a generation, as it seems reasonable to expect, technology and industry will have solved the problem of the duration of the support, the only remaining causes of digital obsolescence will be those related to the availability of adequate devices, format knowledge and reading programs.

Digital Preservation

Preserving digital documents does not only mean conserve them in some form in an apt archive, but also give the final user the capacity of searching, filtering and retrieving them efficiently. This is why the researchers are addressed to build huge Relational Databases for preservation purposes.
The **Relational DBMS** is also the best choice for hosting documents of different type that in the meantime could be put in interrelation each other, like text with images, or music, video etc.
There are two main methods to avoid obsolescence of digital documents, based on the following two fundamental choices:

- **Migration** is an approach that involves periodically moving files from one file encoding format to another that is usable in a more modern computing environment. Migration of the original contents and formats to the current ones of every future epoch, in order to adapt the

documents to the technology of that time, requires a recurrent commitment of resources for the transfer of content that retains all their original characteristics. It has the merit of allowing a gradual convergence of the formats used towards shared public standards. it allows only the preservation of migrated documents, those neglected in the past or rediscovered in the future with all probability will be lost, if not so important to be examined in a IT laboratory.

- **Emulation** of original software environments for reading documents stored in their original format. Emulation backers want to build software that mimics every type of application that has ever been written for every type of file format, and make them run on whatever the current computing environment is. over time it requires the accumulation and maintenance of software emulators able to recreate the original IT environment and allow the document to be viewed.
  It allows you to view not only the documents already saved in the database, but also to read newly found or past neglected documents in case you have the appropriate emulator. After all file format migration is not always the solution. Some CAD and CAM file formats cannot easily be migrated, for example. The aerospace industry has found that migration of older CAD files to a newer format requires a lot of validation, mainly because they are required by a regulatory framework to demonstrate that their data is sound and meets very strict standards.
  In short, the cost of migration and validation is for them much higher than an emulation solution, an approach which involves keeping the CAD software and maintaining it.
  An additional advantage of emulation is the relative platform independency, with the production of a Packet containing all what is needed for reproduction of the original. It consist of retaining the document in its original format encapsulated in a virtual "envelope", with embedded software instructions for retrieval, display, and processing of its content. The envelopes would contain contextual information and the transformational history of each object as additional metadata. Execution of the instructions would rely on an archive of hardware and software emulators or on instructions in the envelope with specifications to construct emulators.
  on the other hand Emulation requires the acquisition of all dependent pieces, contextual understanding of how the software and Operating Systems (OS) were used and expertise to build and maintain the emulated environments. Such an environment is challenging to recreate, and requires parallel preservation efforts of the digital files, the software, the OS and sometimes knowledge of how the creators used or modified both software and hardware. It does not favour the convergence of formats towards shared standards, nor compels software producers to agree and converge toward common architecture. With the passing of time the emulators needed will accumulate, increasing their number, and maintenance costs could soar.

As everybody can see, the two represented strategies are antithetical and possess respective merits and defects, which will probably prevent a definitive and integral choice in favour of one or the other. The conservation sites that adopt an immediate conversion of the document upon inclusion in the database will prefer the migration strategy, as they find it easier and less expensive to migrate with a format converter than to set up a software factory; those instead that have collected recordings in the original format of not standard documents, like for instance the case of digital art, will perhaps be more attracted by the emulation technique, that initially is more rapid to implement and less expensive also if it entails a much greater volume of used memory.

These are seen as alternatives to one another, but both approaches are supposed to be used in conjunction with refreshing. The concept of **Refreshing** involves periodically moving a file from one physical storage medium to another to avoid the physical decay or the obsolescence of that

medium. Because physical storage devices decay, and because technological changes make older storage devices inaccessible to new computers, some ongoing form of refreshing is likely to be necessary, however raising the issue of assuring authenticity of the document.
As the Emulation solution is highly dependent on the specific type of document and original software environment, it is not possible to treat this issue in general terms, so the issue will be restricted only to the discussion of the Migration method, to be implemented on Relational Databases, that is also the currently more shared and preferred solution. It has also the advantage of allowing its generalisation, due to the fact that the original document is immediately migrated in another form inside a database, erasing all direct relations with and constraint of the original environment, which will instead documented beside it.

The generalisation of the problem can then lead to standard solutions, which can be implemented by private companies and public organizations, which have to retain certain type of documents not only for some year, but decades. Think for instance to libraries, State bodies, legal documents, maintenance documents of long lasting products, recording of publications, meetings minutes, police and trial records, building and public works projects, musical and theatrical performances, movies and TV recordings, and so on, the list is incredibly long. In this way Database Migration is becoming a standard solution for digital preservation, with the recording of a packet, containing the document in a standard format that could be maintained for a long period, enclosed with all the necessary information both on its content and possibly its original format or interpreting programs, which are named **metadata**.
Metadata are obviously extremely important both for the search task of the wanted document and future correct reproduction of the original document, so their creation has to be complete and thorough on the part of Database curators; this basic task on the other hand open a complex problem entailing personnel skills, completeness of original documentation, language used for the descriptions, and time demanded for their completion. It has also to be considered that the number of documents to be preserved will be huge, so the human work to this aim has to be limited because too slow. Perhaps an aid will come from Artificial Intelligence and Machine Learning applications, which could perform the task of examining both the original document and its format, producing rapidly a semi-finished product to be then controlled by a curator.
First examples of this use of AI technologies are the Ca 'Foscari University Venice Time Machine, and the American Chronicling America Project at the Library of Congress.

Within **Preservation Archives**, structured as Relational Databases, the Migration solution is used with a method, named Encapsulation, which makes the preserved objects self-describing, virtually "linking content with all of the information required for it to be deciphered and understood".
As a matter of fact the original document, converted if needed to a standard format for that type of information, will be linked to its metadata, with both embedded in a packet. Those packets would be realised by using "logical structures called containers or wrappers" to provide a relationship between all information components, that could be used in future development of emulators, viewers or converters; these containers could also have specifications such as to be automatically computer readable. The method of encapsulation is usually applied to collections that will go unused but preserved for long periods of time. A widely implemented model for this purpose is OAIS, that will be examined later.

Digital Formats

As seen in previous paragraph, digital document formats play an important role in the issue of long-term document preservation, as they have a decisive impact on the question of software obsolescence. The formats of the digital documents, used both in the past and currently, obviously depend on the type of document (text, image, video, music, web page, etc.) and originate both as a commercial initiative of various content producers and as a result of international formal agreements between official standardization authorities.
Digital information is produced in a variety of standard and proprietary formats, including ASCII, common image formats, word processing, spreadsheets, database documents, formulae, charts, multimedia files, sound and video. As a result of such a heterogeneous nature of storable information, a high number of file formats are now spread, and many of them often need specific software to view or edit the file. Sometimes they also have different internal versions for the same format, as well as different formats exist for the same type of content depending on the operating system used, totalling several thousands of them. These formats are continuously evolving and becoming more complex due to new features and functionalities. Hence there are today several file formats which are available but are also incredibly complex, making the binary code of the file meaningless to a human observer if the required software is not available to interpret it.

The continuous development of new applications has produced a huge number of different formats, besides also different versions for some of them; the number of file formats is thus incredibly high. The File Extension collection has indexed over 15,000 file name extensions, each corresponding to a different format. Hence formats represent a possible challenge in document preservation because:

1. Many file formats become obsolete due to several reasons such as:

    - In case of proprietary formats with confidential specifications, the developer of that file format goes out of business, stops supporting that format due to technological changes, or its market share declines; no specification is made public.

    - Supporting programs using that format change significantly.

    - In case of proprietary formats with public specifications no third party intervenes to support them, because of lack of interest or because the available documentation is not complete and accurate.

2. Format depends on obsolete hardware or operating system.

3. New versions of application software may not support earlier format versions (no backward compatibility).

4. There are many public domain format sites such as My File Formats, Wotsit’s Format, File Formats Encyclopaedia etc., but they lack any vision or plan to sustain those formats on Internet over long period.

5. Many application programs that dominate the market create documents with complex structure and layout, and the software for such uses typically models and stores document structure and layout in proprietary formats. Although the software may provide mechanisms for converting documents to common interchange formats, use of such mechanisms often results in the loss or inadequate rendering of content. This is obviously a key factor to consider when proceeding to a file format migration, as some characteristics of the original document could be lost in the new version.

All that produces great confusion and consequently a serious problem to be solved. That's why there are many initiatives taking place to read and convert old file formats.
For this reason the number of formats have to decrease substantially and in the meantime standardized, in contrast to the recurring attempt by companies to impose private formats for customer-retention objectives. The main road for solving the problem would be an international agreement that should:

- Distinguish the files according to the **type of content**, possibly further catalogued in sub-categories if the document has specific characteristics that require special recording capacity.

- For each file type, identify the specifications of the **format** required for preservation purposes, for example by evaluating the factors that must characterize it for efficient use. An example of this has been proposed by the Library of the United States Congress.

- Define a standard retention format for each file type, shared and used by all the actors responsible for archiving digital documents. It would then constitute the standard form for that type of document within the Database or to be recorded embedded in the document's metadata.

Hence:
Generally speaking, open standard formats should be used whenever possible, available formats with public specification should the next to be considered and proprietary close formats should only be considered as a last resort. As for the latter, formats that are in widespread use are more likely to have ongoing and extensive support from software suppliers and user communities, and tools for migration and emulation are more likely to emerge from industry. As for standard open formats versus proprietary ones the choice is not that simple and needs to be looked at closely.
Proprietary formats, such as TIFF of Adobe Incorporated, are seen as being very robust and affordable; however, these formats will ultimately be susceptible to upgrade issues and obsolescence if the owner goes out of business or develops a new alternative. on the other hand also standard format have their flaws, that is the slowness of their development, with the subsequent endorsement by the users community that risks to arrive too late compared to preservation needs. Moreover for many new areas and applications, e.g. AutoCAD computer-aided design (CAD) and drafting software or Virtual Reality, only proprietary formats are available. In such cases a crucial factor will be the export formats supported to allow data to be moved out of these proprietary environments.

- Formats that support the inclusion of metadata are highly recommended as the metadata provides vital information on the provenance and technical characteristics of the preserved document.

- The ability to exchange electronic records with other users and IT systems is an important consideration, hence formats which are supported by a wide range of software or are platform-independent are the best choice.

- Formats that provide error-detection facilities to allow detection of file corruption which may have occurred during transmission are recommended. One such example is PNG (Portable Network Graphic), an image format which includes multiple methods for checking its integrity built into the file itself.

An effort to be made when choosing archiving formats is to limit their number, reducing the range of file formats to support, in order to reduce complexity. A sound approach to preservation planning

is to normalise, rather than add multiple migration formats to the collection. The smaller the range of formats, the lower the overheads. This objective, together with the need to share reference standard formats with other preservation organizations, leads to the choice of a common format for each document type. For some kinds of content, there is consensus around the choice of preservation format.
For example audio archiving where WAV is commonly used. In other areas consensus is much more difficult to achieve. The preservation of digital video is a complex area where progress has been stymied by a lack of agreement, and an uncontrolled proliferation of wrapper formats, delivery methods, and encoding methods. The choice of image file formats is slightly clearer, with a limited choice of formats for archiving and others for delivery. It has been generally agreed that the TIFF format is the correct format for archiving master files (or the RAW or DNG), but this is now being challenged by the JPEG 2000 format which provides a far much smaller level of lossy compression compared to TIFF and is open source, or the last introduced webp for Internet images.
So the uninterrupted technological development of applications acts continuously as a game-changer, thwarting the search for a final solution to the selection of permanent standard preservation formats. But the format issue is not over here, because it has also to be defined, at the top Database level, a format both for the containing packet and the metadata enclosed in it, which will have to be valid within the entire Database structure and possibly to be advantageous for sharing information among a set of preservation databases to be interconnected.
Format of metadata are named **meta-metadata**, and although this information is difficult to codify, it usually refers to metadata that describes the metadata record itself, or to high-level information about metadata “policy” and procedures, most often on the project level.
Meta-metadata information such as who records the metadata, when and how it gets recorded, where it is located, what standards are followed, and who is responsible for modification of metadata and under what circumstances, hence constitute a second level of metadata, which need to be defined and valid for all the preserved objects beside the format of the packets themselves (XML for OAIS).

Standard Formats

As already discussed, one of key digital preservation risks is that a file format should become obsolete, and files in that format could not be opened by future computer systems. If this happens the information in that file is effectively lost, less than a long and uncertain work of reverse engineering on the physical registration of the file to reconstruct its format specifications. Therefore we have to manage this risk by requiring that files to be saved in long term preservation formats are unlikely to become obsolete for a very long period of time, and when they do become obsolete they are expected to be easily migrated to newer formats. These formats are the standard open formats, that will be presented in this section.
It is necessary to consider the following principles while choosing a format for creating or storing any digital document with the aim to make it available for the long term :

- The format should be simple to describe, understand and implement.
- The format should not depend on specific hardware.
- The format should not depend on specific operating systems.
- The format should not depend on proprietary software.
- The format should be resilient against errors or tampering.

In a perfect world, for the reasons listed above, one would always use only open standard formats certified by international authorities, like the ISO ones; in the real world instead an extremely vast set of different formats are used for registering document content, composed of :

- ISO validated open standards ;
- proprietary formats, both open and confidential, rarely used or on the contrary so widely adopted to have become market standards;
- personalized and versatile formats used for specific complex purposes, like some wrapped artworks formats, created and maintained within the context of a users community.

The aim is to reduce the number of different formats to simplify the task of long run preservation, hopefully toward a set of largely adopted and permanent solutions. However, standardization is a very long process, constantly lagging behind developments in technology and market, and in any case must be adopted in practice to be actually useful. It is therefore in the user interest, be it a firm or an organization, to choose these open formats at the very least for all those documents they need or want to keep for the long run.
Libraries and archives with large collections of complex and diverse digital materials are only beginning to test strategies that normalize various types of digital records by converting them from the large variety of formats into a smaller and more manageable number of standard formats like the ISO ones. The same should do the digital documents producers, that is all those private companies and organizations that have an interest in long lasting digital retention.

ISO Standards

The International organization for Standardization (abbreviation ISO) is the most important organization in the world for the definition of technical standards. Founded on February 23, 1947, it has its headquarters in Geneva, Switzerland, and its members are the national standardization bodies of 162 countries in the world. ISO creates documents that provide requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose.
ISO has published 22362 International Standards and related documents, covering almost every industry, from technology, to food safety, to agriculture and healthcare. ISO International Standards impact everyone, everywhere. ISO issues standards regarding both IT security, management systems and the preservation of digital documents, such as the open Archival Information System, ISO 14721: 2002.
Moreover ISO cooperates closely with the IEC (International Electrotechnical Commission), responsible for the standardization of electrical equipment, together with which it publishes the rules concerning electronics and information technology.
The ISO standards guarantee the following features:

- OPENING
  A format is called "open" when it complies with public specifications, i.e. available to any one interested in using that format. The availability of format specifications always makes it possible to decode the documents represented, even without suitable programs.
- SAFETY
  The security of a format depends on two elements:
  o the degree of modifiability of the contents of the file;
  o the ability to be immune from virus insertion.

- PORTABILITY
  Portability refers to the ease with which formats can be used on different platforms, both from a hardware and software point of view.

- FUNCTIONALITY
  Functionality of a format refers to the possibility of processing the file with different computer products, able to ensure a wide variety of functions for the creation and management of the document.

- DEVELOPMENT SUPPORT
  Development support consists in making available the resources necessary both for the maintenance and development of the format and the IT products that manage it (the bodies responsible for defining technical specifications and standards, companies, developer communities, etc.).

- SPREAD
  The spread is the extension of the use of a specific format for the training and management of IT documents. This element, together with the explicit commitment of the standardisation bodies, ensures that the format is supported over time, through the availability of several IT products suitable for its use.

ISO releases open standard file formats of every type, used in all sort of software environment, and also a number of metadata standards which can be used by Preservation Databases.
Employing ISO standard formats both for files and metadata trough various Preservation Archives allow the possibility of easier interoperability between them, the possibility of common search procedures and the realisation of a network of Archives available through the WEB.

The OAIS model

Originally created as part of a broader effort to develop formal standards for the long-term storage of digital data generated from space missions, the open Archival Information System (OAIS) has since formed the foundation of numerous architectures, standards and protocols, influencing system design, metadata requirements, certification, and other issues central to digital preservation.
An OAIS is an Archive, consisting of an organization of people and systems that has the responsibility to preserve information and make it available for a user community.
Realisation of such an archive can be achieved following rules and standards expected by its Reference Model. Let´s remark that this model is not limited to the HW and SW functionalities and performances constituting parts of the Archive, but dictates also organizational structures and methods as well as specializations and duties.
So it is proposed as an architectural design that, while it does not specify a particular technical design in implementing HW and SW elements realising the model, it is an indispensable guide to creating an archive with an accepted set of rules about roles, responsibilities and methods that encourage safe, long-term archival and easy access to preserved information.
A key element of the OAIS design is the concept of package, the unit of information to be archived. In practice, each package will represent either a single digital document or a logically independent documents set. Packaging information can be thought as wrapped information encapsulating metadata and essence components. Packages must be fully self descriptive (all information which describes the package – metadata, presentation renditions, inventory of contents, etc. – must be inside the package itself) and self-validating (the package must contain an inventory of its contents along with self-checks such as digital signatures). In real archival operations OAIS mainly uses ISO standard formats for archiving and preservation purposes, based substantially on XML format for

the documents registration; as a matter of fact in preserving databases and information systems, the XML standard plays a major role, as this is considered to be the most appropriate preservation format for structured textual information.

Since its adoption the OAIS Reference Model has been welcomed and widely preferred by many digital preservation communities. It should also provide a basis for more standardisation and, therefore, a larger market that vendors can support in meeting archival requirements.
Most modern digital preservation initiatives reference the OAIS Model standard, now adopted in the western world as the “de facto” standard for building digital archives. As a matter of fact it has been chosen by manifold countries such as USA, UK and EU.
There is no doubt that in the field of digital preservation and in particular in the adoption of suitable standards, in the development of implementation models and generally in the practical realisation of adequate archives, the United States and the Anglo-Saxon world seem to be more advanced than the European Union, where it seems to predominate chatters, directives always destined to someone else and divergent realisation. But the situation is not so bad, there is a delay, it is true, but some initiatives have actually been launched. Clearly the U.S. have the advantage of sharing common language, laws and regulations, economic interests, which in Europe are an obstacle because they are different between one country and another.

Preservation Archives

The Economist has proclaimed that "the most precious commodity in the world is no longer oil but data", as a simple analysis of stock market trends demonstrates. For instance, shares in Facebook were first sold to the public on 2012. At that time, the company had fixed assets and cash of 6,3billions USD but the share valuation was 104 billions USD, meaning that the company had intangible assets valued at 97,7 billions USD, that is the net real value of the data. Think also that the cost of production to Facebook has been almost nil. From what comes the data value ?
The value lies in re-use potential, as Facebook had smartly discovered, and this fact makes it clear one of the aspects of digital preservation. Because if data are not preserved they can’t be reused: they are a kind of renewable resource, whose real value could last over time for many years regardless of their mere historical value.
The value of data is in their reuse. If they will not be secured for access and reuse in the long term they simply will not produce any long term value. Data is rapidly becoming the lifeblood of the global economy. It represents a key new type of economic asset. Those who know how to use them have a decisive competitive advantage in this interconnected world, both by increasing productivity and by promoting innovation, or offering users more profitable products and services, so often leaving behind long-standing competitors. The problem is whether economic resources have to be spent now to transfer in the future the corresponding real value of the re-use of the preserved data. For answering to this seemingly simple question organizations and private companies have to examine the type of data they produce or anyway possess, choosing which are eligible to long-term conservation. For some of them the choice will be fairly simple, because their preservation is mandatory by law or recommended by agreements or standard use, for other ones it could be clear that their retention is completely useless. The problem stands in between. Some kind of data could possibly be eligible for conservation and re-use depending on future developments, which however are currently unknown.
Let's take a famous example coming from a renown organization, the NASA. Not long ago, NASA recovered the availability of a scientific satellite, launched in 2000, which had ceased to function. This satellite, called IMAGE, was designed to visualize the terrestrial magnetosphere and produce the first complete global images of the plasma present in this region of space. After completing its initial two-years mission in 2002, the satellite was no longer able to contact the base on December 18, 2005 for unknown reasons.

It was considered lost. On January 20, 2018, an amateur radio operator again intercepted signals from this satellite, and it was later confirmed by NASA, which however immediately found itself in difficulty: IMAGE was now obsolete. NASA could not decode the data contained in its signals. The types of hardware and operating systems used in the IMAGE Mission operations Center from 2000 to 2005 no longer existed, and other systems had been upgraded to different versions subsequent to those operating at that time. The recovery of information and the ability to control the devices on the satellite now would have required a significant and costly engineering effort.
The NASA team was able to read some satellite maintenance data, confirming that at least the main control system were operational, but the scientific payload of the spacecraft, still impressive and of great scientific importance, could not be decoded due to the obsolescence of hardware and software systems installed on board. The satellite was so lost another time, and this is too bad, because its data could have been useful for studying near Earth plasma storms that disturb both satellites and communication systems.
This is an example of digital obsolescence, both of hardware and software, which occurred in just13 years, from which an important scientific damage has resulted. Another point of view when examining data in relation to their possible preservation is the foreseeable time for their re-use.
So we can distinguish two categories of them:

Data with short-term re-use
Public bodies hold a very wide array of information and content, ranging from demographic, economic and meteorological data to art works, historical documents and books. Given the pervasive availability of such information and content in digital form, and the widespread use of information and communication technologies by secondary users, public sector information and content are an increasingly valuable resource for the production of innovative goods and services and a major source of educational and cultural knowledge for the wider population.
Thus the information contents and data held by public administrations, thanks to the continuous technological evolution, represent an extraordinary opportunity to provide more efficient services but also, encouraging the re-use by other public entities or private companies, to be used in areas different from those for which they were produced or collected initially.

These kinds of data, also named **open Data**, can generally be listed as referring to:

o **Geodata**: data used to create maps, for example the location of roads and buildings, the topography, visualization of borders, georeferencing of commercial establishments etc .;

o **Culture**: data referring to cultural works and products (for example: titles, authors, etc.), and generally preserved by libraries, galleries, archives, museums;

o **Sciences and Technology**: data produced as part of scientific research, from astronomy to zoology;

o **Economics and Finance**: data on public accounts (income and expenses), information on financial markets (securities, shares, bonds, etc.);

o **Statistics**: data produced by offices and statistical services, social, economic, demographic indicators, etc.

o **Weather**: the various types of data used to understand climate change;

o **Environment and Health**: information on the environment (presence and level of pollutants, water,quality, waste), rates and causes of mortality, incidence of diseases in certain areas, frequency and localisation of bushfires, mapping and cadastral information, etc.

o **Transport**: timetables, routes, travel time statistics etc.

o **Agriculture**: satellites' data about crops, water and plants' diseases. Satellite data uses high-precision microwave technology to measure soil moisture and surface temperatures in individual field zones, at a spatial resolution of 100 x 100 meters and without negative interference due to cloudiness. As a result, an archive of data that goes back several years is now available, and that with the passing of time will give the users also a picture of possible microclimate changes.

Obviously they represent a true mine of gold for all those private entrepreneurs that will be capable of inventing new products and services based on them, and above all this kind of raw material is free of charge, so capable of generating high profits. Seemingly also each private firm can wonder about its own data and their possible use as a new asset to exploit.

Data with long-term re-use

They belong to several categories, such for instance the following:

- Legislation
- Cadastral data
- Archival content
- Economy
- Science
- Technology
- Environment and Climate
- Medicine
- Literature
- Publications
- Art
- Music
- Photography
- Movies
- Media
- Television
- Professional studios
- New digital products

Some of them are compulsory, other obvious about their possible useful re-use, for other still, eventually, some doubt is possible, like Literature and Art. It may seem at first glance that all digital documents about these two fields have no other reuse than the continuous reproduction of the works themselves and the publication of their historical critique. The former does not need any special preservation measures because it is constantly carried out by publishers, both in paper and digital formats, using techniques and devices of their own time. The second belongs to our cultural background and, as for all the other mentioned disciplines, does not in itself shows the indispensability of its conservation beyond its mere historical value.
On a closer examination it appears instead that things are not so simple. First of all there are two distinct kinds of literature and art works, written on paper (like books and sheet-musics) or painted or sculpted, and original directly produced in digital form, both for texts, images, video and digital art pieces.
As for the former they need a digital rendition and subsequent preservation simply for being re-used by new digital publishing editors as ebook in the future, or for allowing future critique to access both of them and preceding historical critique works for comparative examinations. So the

documents of the past must be preserved because they are susceptible to a new examination or because they prove indispensable for new research and new publications.
As for the latter finally, if humanity continues to think that literature and art deserve protection for future enjoyment, she absolutely has to retain them in the best manner, and mandatory deposit has to be extended also to electronic publications and artworks.
As a very last hint it should be remarked the efforts currently under way for the preservation of new digital products that did not exist with pre-digital technologies, like games or virtual reality.
Implementing the conservation of long-term digital documents will be a difficult and costly task, as it will only be possibly achieved with a full implementation of adequate computerized archival structures; alternatively, a large part of the existing digital documents, scattered through myriads of servers all over the world, will be lost forever, while the remainder will still be dispersed and difficult to consult, as if it would be stored in a labyrinthine library similar to the one described in the "Name of the Rose".

The OAIS (ISO 14721), which has already been mentioned, is a reference model for the long-term preservation of digital resources, and above all an indispensable guide to create an archive with an accepted set of roles, responsibilities and methods. Its users have created an "OAIS community" on the WEB to learn about the news and consult each other when issues of implementation emerge.
OAIS is currently the benchmark standard for the construction of data archiving preservation environments. This is acknowledged in European and US projects like for example:

- CASPAR
- PARSE-Insight
- DRIVER
- SHAMAN
- ERPANET
- PLANETS
- DigCurV and APARSEN
- aDoRe (Los Alamos National Laboratory)
- OCLC Digital Archive Service
- Stanford Digital Repository
- MathArc (Cornell UL and SUB Göttingen)

It is also the baseline for several audit and assessment tools, like PLATTER, DRAMBORA, DSA and ISO 16363. Hence, the OAIS model should not be considered as a guideline as such, but better as a mature platform model to be adopted.
However OAIS is a conceptual architecture model that has to be defined and achieved by a specialized software, not a blueprint for system design; as a matter of fact it is assumed that it will be used as a reference model and a guide while developing a specific implementation, to provide identified services and content.
It is now quite clear that the most suitable software structure for a digital archive, with ample possibilities for research, is a relational database that meets the ISO standards, e.g. such as SIARD, which has already achieved a good maturity of use. It is then very probable that both the OAIS model and a Relational Database like the European SIARD, or something similar like the American Preservica's RDB, will make up the basis of future long term digital archives; it will be only the experience that will be accumulated with the multiplication of implementations to clarify definitively if this solution is really the best way to achieve digital preservation on a large scale, for a variety of domains and a variety of objects of a different nature.
It is also necessary to respect the standards already established, not only in the creation of digital archives, but also and especially by the "producers" of digital documents, which must respect them at least in the choice of file formats, thus avoiding unnecessary subsequent conversions with possible loss of information. The best solution would actually be that the "producers" of documents

created them from the beginning as "standard digital objects", that could already perfectly be inserted, complete with metadata, in future preservation structures. If a true "shared solution" for archiving will be established, software packages will certainly also be put on the market, offering the "producers" a pre-packaged environment for the editing of standard digital documents.

Careful attention must also be paid to the recording of metadata, descriptive, technical, structural and navigational ones, which also must also comply with shared standards to allow not only inter-archives research capabilities, but also the possibility of easily transferring contained data from an archive to another.
Other issues, strictly pertaining to the preservation objective rather than simple document storage, should not be even forgotten, like the large field of security management procedures and technologies, such as:

- perimeter security, including for example security of the data farm access, password policy,user authentication and authorization function, firewall, anti-intrusion control;
- data integrity, including for example all the checksum and protection technologies, antivirus,fire protection, backup and remote copying, recovery procedures;
- cybersecurity, that is defence from hackings;
- software evolution: one of the biggest challenges in the field of digital librarianship is simply trying to evolve as fast as technology, because it is needed to also keep up with emerging file formats and software systems to read those formats;
- copyright issues: copyright can be a huge barrier for making any digitized materials accessible, with some institutions focusing solely on public domain material, other owning some of their copyrights such as a university-run press, and more lacking the resources to track down copyright holders.
- software license problems: in some cases the manufacturer does not sell the ownership of the program or digital document to the end user, but only issues a license for use with expiration. After this term, any use of the product could cause legal problems. Recently in the US, Adobe has warned its customers of continue the use of old versions of Photoshop, which they had purchased at the time, as their licenses had expired; the same firm on 12 January 2021 has suspended the possibility to execute the Flash Player, that has been largely used by WEB and SW producers as component of their production.
- issue of "user authentication" of the original and "user permission" about accessibility restrictions to the content or the right to use it. This necessitates controls as to who can access data and for what purpose: Archives cannot accept material without permission to archive from the intellectual property right holder.

Organizations and Companies implementation

Each organization or Company can have its own long-term digital archive, above all if particular confidentiality requirements exist, like for instance in the case of CIA or Avionics companies.
This type of solution however is very costly, both relating to HW/SW implementation/maintenance and skilled personnel requirements, and can be better supported by State and Public Institutions or specialized commercial archives than single private companies. That's why private "on subscription storage sites" are currently under development to accommodate digital documents in the long term,

constituting a sort of natural evolution of current on-cloud storage, like for instance Preservica in U.S. This type of solution could be developed especially in favour of companies, which not wanting to overload their internal systems but intending to keep technical, legal or administrative documents, could save them on external archives.
It is not excluded that such long-term archives may tomorrow also host documents that private citizens could save, even in order to preserve personal memories, without the need to worry about their possible obsolescence: the external archive would carry out any necessary migration, by making also the necessary hardware available to users.

The situation is rapidly evolving, and the same trend of substituting paper with digit will put pressure on the preserving instruments and conservation's providers development. Think for instance that from 2023 the U.S.National Archives do not any more accept documents and records on paper but only digitised copies, however obviously keeping their fundamental obligation to preserve them indefinitely.

The American Situation

In the United States people is realising the risk of not preserving digital records, and private industry, beside public institutions, is offering preservation services. Fortunately they are reacting to that danger. Some last events in the on-line US world, regarding some cloud services which wiped out people records or severely hardened storage rules, contributed to make the public perceive the reality of the preservation problem: cloud services are not conservation services. But now the situation is changing, if not for private citizens at least for companies and institutions.
In U.S. there are several initiatives undergoing, both private and public, for instance:

Chronopolis
is a digital preservation program for the preservation of long-lived digital data collections. It accomplishes this through the development and implementation of a preservation data grid and its supporting human, policy, and technological infrastructure. Chronopolis is a geographically distributed preservation network. All data in the network are replicated among three geographically dispersed partner sites. This geographic distribution ensures that no single catastrophic event will affect the content.
Originally funded by the Library of Congress, the Chronopolis digital preservation network has the capacity of preserving hundreds of terabytes of digital data.
Format obsolescence is not an immediate concern of the Chronopolis system. Instead, this and consequent migration task is regarded as the responsibility of the data providers. The single, overriding commitment of the Chronopolis system is to preserve objects in such a way that they can be transmitted back to the original data providers in the exact form in which they were submitted. Due to this shortcoming the service provided can not be considered a true and complete document preservation process, in that Chronopolis does not safeguard the archived documents from software obsolescence, resulting from file format abandonment or the future unavailability of adequate interpretation software.

Cyark
Curation would not be complete without the possibility of effective preservation of new kind of digital documents, like for instance 3D and virtual reality types, whose use is continuously rising in the engineering, archaeology and arts domains. Luckily the issue has been faced in the US with the CyArk, which is a non profit organization founded in 2003 to digitally record, archive and share the world's cultural heritage: after 16 years in operation they have recorded over 200 monuments on all 7 continents, including Mt. Rushmore, Pompeii and the ancient Mayan city of Tikal.

It creates an accurate surface model of the monument or artwork using a combination of digital recording technologies. The complete data set is then backed up to tape and this gold copy is sent to Iron Mountain's secure underground facility in Boyers, PA. , where it is implemented as a comprehensive data protection, management and archiving solution.

NDSA & DLF
The National Digital Stewardship Alliance (NDSA) is a consortium of organizations committed to the long-term preservation of digital information. The mission of the NDSA is to establish, maintain, and advance the capacity to preserve U.S. digital resources for the benefit of present and future generations. The Digital Library Federation (DLF) is a community of practitioners who advance research, learning, social justice, and the public good through the creative design and wise application of digital library technologies.

Preservica
It provides an OAIS modelled archive and software platform that future-proofs all types of digital content against technology obsolescence, ensuring it remains accessible and trustworthy over decades to meet legal, compliance, governance and brand value needs. Available as a cloud-hosted (SaaS) or on-premise solution, the software is already used by a growing global client base – from major corporations and government bodies to iconic cultural institutions. These include EU Commission, HSBC, the World Bank, Associated Press, BT, Amnesty International, Yale University, Texas State Library, MoMA, the UK National Archives, London's public transport network, Library and Archives Canada, 21 US state archives, several Universities and Museums, 16 national and pan-national archives. Recently it has introduced Preservica Starter, that is a brand new set of FREE and low-cost digital preservation solutions, making it easy and affordable for institutions of all sizes to preserve, curate and share digital content online in minutes.

The Library of Congress
The 548 digital collections at the Library of Congress comprise both material that has been digitized and born digital collections, including more than 13,000 electronic serial titles, half a million electronic books; in 2022, the Library managed 21 petabytes of digital collection content, comprising 914 million unique files. It leads and participates in communities developing open formats and standards, and use wherever possible open source software. Digital preservation efforts are distributed throughout many units at the Library of Congress and include programs related to digital content packaging and ingest, monitoring and reporting of digital storage, sustainable digital file formats, metadata and more. It holds hundreds of digital collections about 132 different parts of the human knowledge, each one comprising hundreds or even thousands of different documents.

Main subjects are:

•Newspaper	3,101,487
•Photo, Print, Drawing	1,199,676
•Manuscript/Mixed Material	487,258
•Book/Printed Material	466,943
•Periodical	403,209
•Legislation	363,218
•Notated Music	125,897
•Web Page	96,409
•Personal Narrative	93,100
•Film, Video	73,583
•Audio Recording	63,637
•Map	57,861
•Web Archive	32,944
•Software, E-Resource	7,632
•Event	4,960
•3D Object	2,716
•Exhibition	13
•Research-Center	1
•Service	1

WEB document inclusion in a collection to be retained is rising, both of public or private kind. Websites are selected for archiving by Library Recommending officers. Sites in the web archive are generally representative samples of web content that document an event or cover a particular theme or subject area for Library's thematic and event collections.

U.S. National Archives and Records Administration
The National Archives and Records Administration (NARA) is the nation's record keeper.
NARA manages the Federal government's archives and Presidential Libraries, operates museums, conducts education and public programs, provides oversight of government-wide records management activities, and provides temporary storage of other agencies' records on their behalf.
NARA holds over 12.5 billion pages of permanently valuable archival Federal and Presidential records in traditional (analogue) formats, and 900 terabytes of electronic archival records, included the rising number of digital borne Government documents.

USC Digital Repository
The USC (University of Southern California) Digital Repository is a long-term home for digital collections of research, cultural and business value. Its cloud archive is supported by USC's Center for High-Performance Computing, the sixth-fastest academic supercomputer in the United States.
It provides digitization, restoration, preservation, storage and digital asset management.
It has also opened its services to commercial and educational users, providing the expertise and resources to allow them to manage complex collections of not just books, but film, TV, video tape and digital assets, with a current 380 terabyte total.

WEB Archiving
Preserving the web, or 'web archiving', refers to the practice of taking a copy of a website or of particular content published on the web to act as a record. A web record might consist of an entire website or only the text from a few pages. Web records require urgent attention because the web by nature is ephemeral, with very large changes from an year to another. For this very reason it is especially important to capture web content when it is the only version of a record.
Currently some organizations are active in this task, as:

- Library of Congress
- International Internet Preservation Consortium
- University of California Digital Library's Web at Risk project.
- University of Illinois at Urbana-Champaign's ECHO DEPository project
- Internet Archive
- WASAPI, a collaboration with LoCKSS, Stanford University DLSS, University of North Texas Libraries, and Rutgers University
- End of Term Web Archive
- Stanford Libraries
- Federal Depository Library Program (FDLP) Web Archive with selected U.S.Government Web sites
- MirrorWeb, 880 millions page, more than 45000 Websites, more than 400 TB stored
- WebArchiving.net

but not all are long-term preservation archives.

The European situation

All the preceding issues and characteristics of the general digital preservation topic are valid both in U.S. and Europe, but the latter has two more huge problems to face, that is multiplicity of language and legislation among nations.
The European variety of languages poses not only an obvious problem of preserved documents understandability, but also a technical problem about metadata: in which language have they to be expressed ?
As for the different legislation among States the most severe outcome regards copyright rights and authenticity guarantee of the stored documents, as well as mandatory legal preservation terms.
During recent years the EU authorities have promoted and funded several research projects about the general issue of digital preservation, aimed to identify best methods, procedures and tools for a common digital preservation framework, to be realized within European States by national authorities and private firms. A list of them follows:

- APARSEN
- ARCOMEM
- BLOGFoREVER
- BRITISH ATMOSPHERIC DATA CENTRE
- CASPAR
- CESSDA
- CESSDA ERIC
- CESSDA SaW
- DASISH
- DELOS
- DPE
- DPS
- DwB
- EARK
- eArchiving
- ENSURE
- ERPANET
- KEEP
- LiWA
- PARSE.Insight
- PLANETS
- PrestoPRIME
- PROTAGE
- SCAPE
- SEEDS
- SERSCIDA
- SHAMAN
- SODA
- SPRUCE
- TIMBUSUK DATA ARCHIVE AND THE NATIoNAL ARCHIVES
- WF4EVER

A Web Archiving & Preservation Task Force (WAPTF) was first formed in 2010 in an effort to coordinate national web archiving programmes. In recent years, however, new developments in web-archiving have emerged and many more organisations have turned their attention to their own institutional requirements for managing and archiving their web records and collections.

In all EU States are currently underway several national preservation initiatives, tailored to national needs and characteristics, possibly using the results come from European projects, but apart from Europeana platform without a binding common architecture within the Union. Although the emission of a Digital Single Market Strategy, that aims to open up digital opportunities for people and business and enhance Europe's position as a world leader in the digital economy, the overall picture seems to be that of an EU commission continuing to propose projects and sound advices that others will have to materialize, while member States, during this crisis years, traipse realizing national specific projects without a unique European design, if not the adoption of OAIS as architecture. It is anyway worthwhile to add that UK situation is better than that of EU, because they followed a national policy, founding in 2001 the Digital Preservation Coalition, and developed their solutions in collaboration with the United States.

The future tasks

As the old saying has it, prediction is difficult, especially about the future. Trying to predict the future certainly generates mistakes, for the simple reason that the future is imagined as an extension or an extrapolation of the present, while technological advances and unexpected events disrupt reality by providing previously unthinkable tools and deep changes in the world.
In this way it is possible to think of great centres of digital preservation with immense banks of flash memories, while perhaps instead the problem will be trivially solved using unified software all over the world and synthetic crystal memories, with capacities of thousands of terabytes, that will be possible to buy in every store for a few bucks.
The digital revolution has already begun to change the world, and will even more affect the way of life in the future:

- Internet of Things, by connecting all the machines and most of objects at our disposal, both at home and in the office, to our network and under remote control, will profoundly change our lives.

- Avatars governed by Artificial Intelligence, will become our normal means of communication both with computers and with the institutions and companies that will rely on them. Talking to a human being will become an exceptional privilege.
  Also files edited by A.I. will become a problem just because of the gigantic rate of their production.

- Robots will be used more and more widely not only in offices and companies, causing employment problems, but also in hospitals, forcing us to trust a surgeon-robot.

- Automated guided vehicles, trains and planes will have to use millions of billions of data every day to get us to our destination safely, but those data will also have to be stored for a reasonable time, both for the carrier's administrative activities and as documentation available to the authorities in case of unfortunate accidents.

- Genomic medicine will need, to be applicable and effective, all information on the individual phenoma, progressively collected during human existence to be used in case of illness or injury.

- Social and economic infrastructure will be entirely based on digital applications and Internet, becoming more effective but also more fragile, because it will need the perfect working of every parts for properly functioning.

It is therefore necessary to consider that the gigantic adaptation effort that awaits us will bring to the foreground many problems, resulting from the "digitalisation", far more relevant in the immediate respect than the need for long-term preservation of digital documents. It could therefore been overlooked. But two allies exist:

Storage of data, performed for normal activity needs and for short term re-use, will maintain at least for some time the registered digital documents, that could possibly be later recovered for long-term preservation.

Internet surely holds multiple copies of the most part of every document produced in the world, so theoretically it should be ever possible to recover it, provided that we know where to look for it. Search engines could help, but private citizen non personal or firms not confidential relevant documents are not registered by them: a solution could be the automatic registration of such documents, when they are produced or locally stored, on a specialised global search engine to be used by researchers in the future.
Who knows ? It is therefore better, for all this reasons, to be cautious and not to venture forecasts too far in time, but limit them to a few of the next years. The ongoing ever rising acceleration of hardware and software upgrades within the lifecycle of an IT device, plus the normal trend of evolution and substitution of these products, will strain the process of protecting digital documents from obsolescence, creating an increasing burden on digital archives personnel and resources, above all if external provider will be not capable to provide timely solution to companies.
This could end in a sudden crisis at every moment, and this could be very dangerous for some type of industries, like the following example clearly shows.

The Boeing case

When comparing this rapid pace of software and hardware changes (CAD S/W versions change every 6 to 12 months, CAD generations change every 10 years) with the lifecycle of certain highly complex industrial products, it appears that digital preservation is not always a process started for a mere intellectual satisfaction, but a compelling necessity.
Let's take for instance the planes produced by Boeing: 707 was designed using paper drawings, then 747 with 2D CAD, 777 by 3D/2D CAD and finally 787 by modelling 3D.
Data volume reached more than 4 Terabytes only for the last of these projects, but every plane flying in the world needs a complete running copy of it, with all the machine production and maintenance history, for a total of several petabytes of data for the entire Boeing flying fleet.

On the other hand airplanes have lifecycles of several decades, as the 707 program that began in 1952 and it is still in use today: as a matter of fact the last commercial passenger flight was carried out by Saha Airlines in 2013, but as of February 2021, of the 1,010 units produced, 524 of them in military versions are still operational, mainly in the US Air Force. During the operative life of each of those aircrafts all their relating data have to be maintained and possibly updated.
Let's remark that the same is valid for ships, trains, modern buildings, as well as to a lesser extent for trucks, cars, motorcycles, earth moving machinery and heavy equipment. Not to talk of spacecrafts, satellites and heavy weapons.
So it exists, just right now, a challenging condition for a large part of industries, but only a fraction of them are reacting in whatever way. Boeing and others have however done something to face the problem now for avoiding future serious troubles.
To face this issue in 2010 the Long Term Archiving and Retrieval (LOTAR) project has been started, promoted by leading American and European aircraft producers, like Boeing, Airbus, Dassault, General Dynamics, Eurocopter, Israel Aerospace Industry, BAE Systems, Lockeed Martin and others. Given that traditional legacy retention and retrieval methods do not support complex

digital product definition data, it has been established as the objective of LOTAR International to develop, test, publish and maintain standards for long-term archiving (LTA) of digital data, such as 3D CAD and PDM data. These standards will define auditable archiving and retrieval processes. Use of the standard series by other branches of industry such as the automotive or shipbuilding industry is possible. The results are based on the ISO 14721, open Archival Information System (OAIS) Reference Model. The spadework in long term archiving of the LOTAR project will be a benefit also for all other digital preservation projects relating to huge size and highly complex data.

Conclusions

Normally the conclusions of a topic are the ideal place to beat the bass drum and develop the full orchestral focus on the work done, simultaneously indicating the great prospects that open up in the future. Given the subject matter and the reality of the facts, this does not seem to be the most appropriate tone in this case, but rather a whispered one, and a much more modest reflection is perhaps required.
As it has been seen leading archival institutions and technical environments are fully aware of the issue, much less within political institutions, that especially now are besieged by much more pressing problems; the public opinion is just slowly realising that a problem could exist.
Heritage and research institutions are already concretely facing the issue or are currently organizing themselves to do it.
Industry and Services, mainly in the U.S., have begun to employ solutions on the market for the long term conservation of digital documents, both in house and in outsourcing. Other high-tech industrial companies, that have a particular problem because of the complexity and long lifecycle of their products, are funding projects to attain the same result.
We should also think to develop other kind of personnel expertise for Preservation Archives, whose skills have to be some in between an archivist and a IT engineer, because they are a decisive element of every future solutions. Digital content producers even have to converge toward common standard formats, possibly joining their metadata to the newly produced documents.
Strategy makers, policy makers, professionals, users and managers are called worldwide to understand the huge problem they are facing, and confront it for not leaving to their future successors an inextricable problem to solve.
And we are only at the beginning of a long, long way ahead.

Bibliography

5D Data Storage by Ultrafast Laser Nanostructuring in Glass
A DNA-Based Archival Storage System
A Guide to Web Preservation
About - Digital Preservation (Library of Congress)
About - E-Ark Project
About the Digital Library Federation - DLF
Approaches to Managing and Collecting Born-Digital Literary Materials
Arch Mission Foundation Announces Our Payload On SpaceX Falcon Heavy
Archives Portal Europe
Archiving the Phenome- Clinical Records Deserve Long-term Preservation
Archiving_of_Radio_and_TV_Programmes
Ars Electronica
Assessing Digital Preservation Needs in the UK
Caring for archives – The National Archives
Collecting Digital Content at the Library of Congress
Consultative Committee for Space Data Systems CCSDS
Contact with lost NASA satellite IMAGE confirmed
Contents - Digital Preservation Handbook
Creating metadata - Stanford Libraries
Critically Endangered - Digital Preservation Coalition
Data Base Preservation: the international Challenge and the Swiss Solution
Data storage lifespans_ How long will media really last
Digital Collections - Library of Congress
Digital Data Preservation Practices to Ensure Long-Term Information Accessibility
Digital Library Federation - DLF
Digital Preservation and Permanent Access
Digital Preservation at the Library of Congress
Digital Preservation Coalition Digitisation & Digital Archiving
Digital Preservation formats
Digital Preservation System (DPS) • European University Institute -HAEU
Digital Transitions and the Impact of New Technology on the Arts
Digitally Endangered Species - Digital Preservation Coalition
DPE Digital Preservation Europe
eArchiving Standards & Specifications
E-ARK, European Archival Records and Knowledge Preservation
Electrical Switching and Bistability in Organic Polymeric Thin Films and Memory Devices
European Commission Report on Bringing Europe's Cultural Heritage Online Digital Single Market
Europeana
File Formats - Digital Preservation Coalition
File formats and standards - Digital Preservation Handbook
Future Digital Libraries Research and Responsibilities
How DNA data storage works - ExtremeTech
Interagency Science Working Group, Open Archival Information System (OAIS) Standard Reference Model
International Internet Preservation Consortium - IIPC
Internet Archive Digital Library of Free Books, Movies, Music & Wayback Machine
Introduction to Digital Formats for Library of Congress Collections
ISO 147212012 - Space data and information transfer systems -- Open archival information system (OAIS)
-- Reference model
ISO Publicly Available Standards
ISO_IT_Strategy_2017-2020
Jisc UK
Knee-Deep_in_the_Data_Practical_Problems in Applying the OAIS ReferenceModel to the Preservation of

Computer Games
Library-of-Congress-Digital-Strategy-2019-2023_v1.1.2
LNE Durability of syylex-glass-dvd-accelerated-aging-report
Long Term Need for Data Exchange Standards
Long term preservation formats
Longevity of Electronic Art
Long-Term File Formats – National Archives of Australia
Long-term management and storage of digital motion picture materials
Long-term preservation of biomedical research data
Long-Term Preservation of Digital Documents - SpringerLink
Long-term-preservation-of-scientific-data
Mass storage and long-term document preservation
M-Disc optical media reviewed_ Your data, good for a thousand years
Media Durability
Microsoft Has a Plan to Add DNA Data Storage to Its Cloud - MIT Technology Review
Nano-bio-computing lipid nanotablet - Science Advances
NASA IMAGE- Long-Lost Satellite Tech Is So Old NASA Can't Read It
New materials for next-generation data storage, skyrmions, chirality
OAIS implementation Case studies - wiki.dpconline.org
OAIS Model application in digital preservation projects
OAIS Reference Model (ISO 14721) The fundamental standard for digital preservation
Open Source Software For Digital Preservation Repositories
Our 5D storage crystal joins Tesla Roadster on incredible space journey - Optoelectronics Research Centre -
University of Southampton
Overview of emerging nonvolatile memory technologies
Overview of EN 9300 LOTAR standards
Personal Archiving - Digital Preservation (Library of Congress)
Preserving Computer-Aided Design (CAD)
Preserving Interactive Multimedia Art
Preserving Long-Term Access To United States Government Documents In Legacy Digital Formats
Preserving the Web
Princeton University Library Digital Preservation Action Plan
Recommended Formats Statement
Research on another permanent data Storage Solution
Research on Digital Preservation within projects co-funded by EU
Saving the World Wide Web - Digital Preservation (Library of Congress)
SIARD format descriptioning
SpaceX Hid a Second, Secret Payload Aboard Falcon Heavy
Systematic approach towards WEB Preservation
Technical Guidelines for Digitizing Archival Materials for Electronic Access
The Five Organizational Stages of Digital Preservation
The Long-term Preservation of Databases
The Open Archival Information System (OAIS) Reference Model Introductory Guide
The Preservation of Web resources Handbook
The quest to save today’s gaming history from being lost forever - Ars Technica
UK Digital Preservation Handbook
Venice Time Machine Project
Web Archiving (Library of Congress)
Web Archiving in the United States_A 2017 Survey
Web preservation demands access - Digital Preservation Coalition
World without history ?