Digital Preservation: Implementation and Challenges

Madhura Pimputkar¹, Prof.Suhasini Vijaykumar²

¹²Bharti Vidyapeeth Institute of Management and Information Technology, Navi Mumbai

Abstract—Digital preservation is the active management of digital information over time to ensure ongoing access to the data. It is the management and maintenance of digital objects (the files, or groups of files, that contain information in digital form) so they can be accessed and used in future. It is error-free storage of digital information, with means for retrieval and interpretation, for the complete time span the information is required for. Digital preservation systems aim to ensure that long-lived digital data will be usable in the distant future.[8]

This paper presents the concepts and observations on the contents of digital preservation. It discusses about the overview of technological approaches and strategies to digital preservation and the challenges in implementation of these strategies

I. INTRODUCTION

Digital Preservation has emerged as an important challenge for information systems in almost every domain. The field is increasingly recognized and has progressed in the last decade. Digital preservation is dealing with the problem of maintaining digitally encoded information so that it remains accessible and understandable over very long periods of time. Applying a suitable preservation strategy for individual (groups of) digital items can be a technically challenging task. The required effort typically depends on the individual objects and characteristics like hardware and software dependencies, data formats, or the integrity of the object, to name a few. However, in recent years, preservation strategies and corresponding technologies have been developed that enable us to preserve a large range of digital materials. Key aspects target the prevention of data loss as well as format/hardware obsolescence.[10]

Another increasingly difficult problem digital preservation is facing, is the management of the continuously growing volumes of data. It is a challenge for research data archives to establish and maintain robust, long term trusted and reliable digital repositories while keeping track of new insights from digital preservation research. [9]

II. ISSUES IN PRESERVATION

A. Keeping the data

Every digital file consists of a series of zeros and ones, or bits (binary digits). These streams of bits need to be preserved and retained over long period of time, without loss or damage, to ensure the survival of digital materials. There are many threats to any attempt at preserving these bits of data.

Storage media can deteriorate over time, leading to corrupted files. Storage media can become obsolete and unsupported by present computers and the software that understands, supports and provides access to them. The bits may be ignored, abandoned, accidentally deleted, get corrupted or maliciously destroyed. Removable media may be left unattended and forgotten, files stored on a shared network drive might be left carelessly, or a third party cloud storage may fail due to some reason. Maintaining a systematic and reliable process for bit preservation remains a basic and prime requirement to ensure long term digital preservation. Storage media must be checked, monitored and refreshed. Redundancy can be implemented by replicating or backing up files, introducing diversity in dependent technologies and avoiding catastrophic disaster at a single location. Check sums must
be generated and frequently recalculated to identify any loss in data and ensure that the integrity of the bits can be verified in an efficient and automated manner. The locations in which digital materials are stored should be carefully recorded and maintained, and responsibility for their preservation should be allocated.[3] [7]

B. **Keeping the meaning of the data**

Restoring the information that is encoded within a stream of a bits requires computer software that is designed to render, manage and handle, analyze or otherwise interact with the particular encoding or format of the data. Over a period of time, the encodings may get changed or file formats may change, and the software applications that interact with them may become outdated. Understanding the technology on which particular digital materials are dependent enables suitable action to be taken to ensure their preservation. A proper and suitable preservation planning process might result in digital files getting migrated from one format to other format, the emulation of software that has chances of being obsolete, or the employment of alternative software applications to render the data. Each of the options presents its own advantages and limitations and these options need to be evaluated carefully to find a solution, possibly on a case by case basis. [3]

C. **Maintaining trust in the data**

Digital materials have the ability to remain fluid over time, being modified or updated with ease, being damaged by media failure, or decoded into human readable and understandable information in an unreliable inconsistent or inaccurate manner by rendering software. For an end user to believe in the result of digital preservation work it requires careful planning and consideration of the entire life cycle of the digital data and who or what has interacted with them over time. Information management systems need to be able to access essential contextual information regarding the business procedures of the creating agency. Credibility, authenticity and integrity of digital resources can be equally important in other sectors. The application of data integrity techniques and the maintenance of audit trails can provide confidence that a digital object has remained unchanged since deposit in an archive. [3]

D. **Acting in a timely manner**

Prioritizing and planning digital preservation activities and applying them in a timely manner can be critical not just in avoiding and reducing loss but in ensuring the optimal use of limited resources. Where the opportunity exists to get involve early in the life cycle, digital materials can be shaped to sustain better into the future. The choice of file format, the capture of crucial documentation or the description of key relationships in the metadata may require a small investment up front, but could deliver considerable savings further down the line. [11] Early involvement to head off obsolescence of the technology may provide greater confidence and assurance of long term sustainability but with the risk that involvement may not ultimately be required and resources were wasted. Timely action may reduce unnecessary activity, but it may increase the effort required to research obsolete technology in a particular case requiring expertise that is no longer present. Appropriate action should be taken on a case by case basis. [3]

E. **Coping with the data deluge**

The rate of data creation is expanding is far more than the rate of developments in data storage. While this places demanding pressures on selection processes, policies and other organisational decision making it also poses technological questions. Simple and easy preservation processes that function effectively at one level may not necessarily scale easily to handle very large amount of data or perhaps very large files. The technology to work at scale is progressing rapidly, with growing expertise for handling large audio visual collections, research data and web based archives.
But some repositories still face lot of challenges in developing and maintaining scalable architectures and procedures and processes to handle increasing quantities of data. The technical and managerial challenges in accessing, managing and providing access to digital materials on this scale should not be ignored. It is important to remember that selection, appraisal and disposal are important components in any digital management activity.[3]

III. STRATEGIES

Many digital preservation strategies have been proposed, but no one strategy is perfect or suitable for all data types, situations, or institutions.

A. Migration

Migration is the process of transferring digital materials from one hardware or software type to another or from one generation of computer to another. Migration can also involve transfer of data to non digital media such as paper or microform, or the transfer to a more appropriate medium for example floppy disc to CD-ROM. Migration ensures the integrity of the object, preserving the essential and significant characteristics of the data and retaining the ability to retrieve or view despite changes to technology.[5]

With migration it is not always possible to replicate the same as the original whilst ensuring compatibility with the new technology. It is important that if organizations decide to use migration they need to determine what aspects to be migrated and need to be preserved and determine if migration is appropriate. Migration store data in latest format before they become obsolete. Another approach to migration is migrating to standard formats. As standard formats are seen to be less volatile that other non standard formats it helps to ensure the integrity of data. The process of migration can also include refreshing. Refreshing is the process of copying or moving digital information from one long term storage media to another. This is done without making any changes to the object or the bit stream. Refreshing guarantees that the information is stored on a newer media before the old media gets out date.[4]

**Challenges:**

1) Very expensive as it requires assessment of formats and observation of the development of formats.
2) It is a continuous task which needs to be repeated on a regular basis.
3) As a stand-alone procedure, it does not protect against obsolescence of devices file formats and data carriers.
4) Information may be lost during migration.

B. Emulation

Emulation is the process of replicating the functions of one system using a different system so that the second system behaves and appears like the first system and the original digital material is thought to still be available in its original form. Emulation is a way to tackle technological obsolescence as it provides a method of preserving the functionality of accessing the digital information which may be lost with the software or hardware when it becomes outdated. This strategy does not encompass preserving aging hardware and original operating system software. The goal of emulation is to retain the look and feel of the digital object as well as its functionality. The principle of this strategy is to copy the technical context of the resource allowing the original objector a refreshed copy of the original object to be used in the future. [4] [5]

The advantages of emulation are that the original data is not changed or modified in anyway which helps to maintain the integrity of data and records and the functionality. It is also an efficient
process in that once the data is emulated no further action is needed, until it needs to be emulated again as a result of technology advances.[4]

The process of emulation includes the creation of emulators, which are programs that translate code and instructions from one computing environment to be rightly executed in the other. This strategy of emulation can be a costly and time consuming process.

While emulation is a used technique of preservation its practical benefits are again currently not widely exhibited, but what testing has been done has been positive. More study in this area will be required and is currently being undertaken.

**Challenges:**

1) The process is very costly and expensive.
2) A large number of different emulators are required to implement this strategy. (depending on formats and required software).
3) It can’t be assured that emulators will run on future operating systems and file systems.

**C. Encapsulation**

Encapsulation is the wrapping up of digital objects and metadata needed to provide access to the object. It reduces the chances that components required to decode or interpret the object will be lost. It is seen as a viable solution to technological obsolescence for file formats because all the information to interpret the bits is available.

It can be attained by using physical or logical wrappers that provide a relationship between the object and the supporting information. The types of supporting information that should be included in the encapsulation include the representation of information used to interpret the bits correctly for access; the record to describe the source of the object; the context to describe how the object is related to other information outside the container; reference to provide identifiers to objects to uniquely identify them; and facility to provide evidence that the object has not been altered or modified. Encapsulation focuses on overcoming the problems of the technological obsolescence of file formats by making the details of how to interpret the digital object part of the encapsulated information. This technique involves building the original application that was used to develop or access the digital object on future computer platforms. Part of the encapsulation process may be migrating the entire record to a more easily documented format.[4]

Encapsulation is an essential part or abstract of emulation and a number of methods are being developed using its theory, notably the universal preservation format.

**Challenges:**

1) The preservation system may have to automatically generate archive package for encapsulation. There may be no application that can perform this function.
2) Increases the storage overheads while encapsulating the format specification with digital preservation.[6]

**D. Universal Virtual Computer (UVC)**

Universal Virtual computer is a computer that builds a layer between the hidden computer platform and the underlying software. It is a form of emulation and all files are backed up in a format that they can be accessed and interpreted by the universal computer. To read and access the data in the future would only need a single layer between the UVC and the computer at the time.[4]
Challenges:
1) The approach is expensive.
2) It is not very easy to implement.

E. Technology preservation

This involves preserving working models of essential computer hardware with the programs that run it and on it. This is not a feasible preservation option in the long run due to the costs associated with maintaining the aging computers and the staff and training required to retain and run the technology. Technology preservation is more of a disaster recovery technique for use on digital objects that have not been subject to a proper digital preservation technique. It offers the potential of coping with media obsolescence, assuming the media hasn’t deteriorated beyond readability. It can expand the window of access for obsolete media and file formats, but is eventually a dead end, since no obsolete technology can be kept functional indefinitely time period. This is not a approach that an individual institution can implement. [4]

Challenges:
1) This is not a strategy that an individual institution can implement.
2) Maintaining obsolete technology in usable form requires a substantial investment in equipment and personnel.
3) Difficulty in maintenance of old machines due to lack of replacement parts.
4) Knowledge of operating hardware and software components is required.
5) Relatively costly and expensive.

F. Bit Stream Preservation

Bit stream preservation is applied as a basis for other preservation strategies, it is no stand alone. Bit Stream preservation is the process storing the binary code of the digital object. The object will not be viewable without the original creation software and hardware, thus it needs to be used jointly along with other preservation techniques to ensure accessibility.[1] [4]

Challenges:
1) It is the most basic approach and cant be used alone independently.
2) This approach cant be used as long term maintenance as it only deals with the question of data loss due to hardware failure. [1] [2]

G. Durable/Persistent media

This approach focuses on the use of media that is designed to be superior than others. Use of such high quality media provides longevity. This technique lessens the need for regular refreshing and minimizes loss resulting from media deterioration.

This technique does not tackle media obsolescence or physical loss, it prominently means that the short term preservation or storage methods will last longer. Many organizations and archives have to complete migration or emulation by using formats that have a longer shelf life.[4]

Challenges:
1) This approach does not solve the problem of media obsolescence and physical loss.
2) It does not address the issue of obsolescence of encoding and format.
IV. CONCLUSION

There are different theories on the best methods to preserve digital material, and various techniques have been developed. They vary from preserving the original technology on which the archival digital objects ran, to preserving only the crucial properties of an object, which are defined independently of any specific hardware or software platform. Each strategy has its own advantages and limitations. There is no one single and perfect solution to the issues and challenges of digital preservation. Digital preservation continues to present a complex challenge which is why there are various strategies and a number of various applications currently being undertaken by various organizations around the world to tackle the digital preservation challenge. The choice of which technique to choose or whether to develop a new strategy is the best option is depend on the nature of the material to be preserved, the objective of retention and the organization.

REFERENCES