Research Dimensions

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017

IMPACT FACTOR : 5.11UIF

Problems Faced in Digital Preservation

Mrs.Tank Manisha Kishor

Librarian

L.B.P.M.Mahavidyalaya, Solapur

Mobile No. 09420088246

E-mail id- manishatank11@gmail.com

Abstract: Today, information technologies that are increasingly powerful and easy to use, especially those that support the World Wide Web, have unleashed the production and distribution of digital information. If we are effectively to preserve for future generations the portion of this rapidly expanding corpus of
information in digital form that represents our cultural record, we need to understand the costs of doing so and we need to commit ourselves technically, legally, economically, and organizationally to the full dimensions of the task. Failure to look for trusted means and methods of digital preservation will
certainly exact a stiff, long-term cultural penalty.

Keywords: Digital preservation, metadata, detoriation, Digital obsolescence

Introduction: We all know that in this era of knowledge, the important commodity is information or data. This information is to be preserved for its long use. Centuries before this information is stored using many different materials, including stone, vellum, bamboo, silk, and paper. To keep this information safe and long lasting the information scientists always in need of media in which this information is stored for long duration. With the development in technology this process of preservation becomes easier and long lasting. Now this information storage and preservation is done with newly developed digital media i.e. digital preservation process is used to preserve the valuable information. Digital preservation is nothing but, the set of processes, activities and management of digital information over time to ensure its long time accessibility. The goal of digital preservation is to preserve material

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017
IMPACT FACTOR : 5.11UIF

resulting from digital reformatting and also the born digital information without analog or print format. Digital preservation is an ongoing process. So, the way the digital information is stored is important in ensuring its longevity. The long term storage of digital information is accompanied by preservation of metadata.

Thus digital preservation is defined as," long-term, error free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required for.

Digital preservation standards: To standardize digital preservation practice and provide a set of recommendations for preservation program implementation, the Reference Model for an Open Archival Information System "OAIS" was developed. The reference model includes the following responsibilities that an OAIS archive must follow,

- Negotiate for and accept appropriate information from information Producers.
- Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.
- Determine, either by itself or in conjunction with other parties, which communities should become the designated Community and, therefore, should be able to understand the information provided.
- Ensure that the information to be preserved is Independently Understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.
- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.
- Make the preserved information available to the Designated Community.
- Develop core competencies in audio preservation engineering. Participants noted with concern that the number of experts qualified to transfer older

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017

IMPACT FACTOR : 5.11UIF

recordings is shrinking and emphasized the need to find a way to ensure that the technical knowledge of these experts can be passed on.

- Develop arrangements among smaller institutions that allow for cooperative buying of esoteric materials and supplies.
- Pursue a research agenda for magnetic-tape problems that focuses on a less destructive solution for hydrolysis than baking, relubrication of acetate tapes, and curing of cupping.
- Develop guidelines for the use of automated transfer of analog audio to digital preservation copies.
- Develop a web-based clearinghouse for sharing information on how archives can develop digital preservation transfer programs.

Problems Faced in Digital Preservation: The unique characteristic of digital forms makes it easy to create content and keep it up-to-date, but at the same time brings many difficulties in the preservation of this content. Margaret Hedstrom points out that "...digital preservation raises challenges of a fundamentally different nature which are added to the problems of preserving traditional format materials." Normally in the process of digital preservation some problems are observed.

- **The Viewing Problem** - All digital formats require computer technology to view them. By nature technology (software/hardware/formats) move at such a rapid pace that, after some days the technology may not be available to view the data. This is of course unless you're viewing data right after you "preserve it" in which case, it's not really preserved now as it.

- **The Scrambling Problem** - Data is often compressed to assist in its storage and or protect it's intellectual content. These compression and encryption algorithms are often developed by private organizations who due to some reason will one day cease to support them. If this happens we will stuck between a rock and a hard place. In this crucial condition if we go ahead and uncompressed by yourself, it's quite possible that it may be the violation of copyright act.

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017
IMPACT FACTOR : 5.11UIF

- **The Inter-Relation Problem** - Digital information is often linked to other items. This is much more evident in the digital world than the physical. If these links aren't maintained the information is either incomplete, incorrect, or just plain doesn't make any sense. Unfortunately, due to the diversity of digital linkages and the relatively recent identification of these issues, they're often overlooked. A simple example of this is links on web pages which have died, never to be resolved again. The so the important information linked with this is lost or we are not able to use it.

- **The Custodial Problem -** Who is the custodian of a digital document? Is it a librarian's job? What if by some mistake someone changes the content without telling the librarian what happen to the originality of information. After all, digital content is dynamic and easily be changed. And anybody can't able to notice it. Then what about the preservation.

- **The Translation Problem -** If we need software to interpret data (due to formats etc), and software changes version to version, will it be translated differently in subsequent versions? Even if the software claims it will sometimes it might lose formatting, a font? This is particularly dangerous where the changes are subtle or so small that no one can notices them,

- **Physical deterioration-**

  The media on which digital contents are stored are more vulnerable to deterioration and catastrophic loss than some analog media such as paper. While acid paper is prone to deterioration, becoming brittle and yellowing with age, the deterioration may not become apparent for some decades and progresses slowly. It remains possible to retrieve information without loss once deterioration is noticed. Digital data recording media may deteriorate more rapidly and once the deterioration starts, in most cases there may

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017

IMPACT FACTOR : 5.11UIF

already be data loss. This characteristic of digital forms leaves a very short time frame for preservation decisions and actions.

Digital obsolescence -Another challenge is the issue of long-term access to data. Digital technology is developing quickly and retrieval and playback technologies can become obsolete in a matter of years. When faster, more capable and less expensive storage and processing devices are developed, older versions may be quickly replaced. When a software or decoding technology is abandoned, or a hardware device is no longer in production, records created with such technologies are at great risk of loss, simply because they are no longer accessible. This process is known as Digital obsolesce.

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017
IMPACT FACTOR : 5.11UIF

- **Migration-**Migration is the transferring of data to newer system environments . This may include conversion of resources from one file format to another (e.g. conversion of MS word document to PDF), from one operating system to another (e.g. windows to LINUX), from one programming language to another (e.g. from C language to JAVA), so the resource remains fully accessible and functional. Resources that are migrated run the risk of losing some type of functionality since newer formats may be incapable of capturing all the functionality of the original format, or the converter itself may be unable to interpret all the nuances of the original format. The latter is often a concern with proprietary data format.

- **Replication-** Creating duplicate copies of data on one or more systems is called replication. Data that exists as a single copy in only one location is highly vulnerable to software or hardware failure, intentional or accidental alteration, and environmental catastrophes like fire, flooding, etc. Digital data is more likely to survive if it is replicated in several locations. Replicated data may introduce difficulties in refreshing, migration, versioning, and access control since the data is located in multiple places.

- **Emulation-** Emulation is the replicating of functionality of an obsolete system. Examples include emulating WordPerfect 1.0 on a Macintosh. Emulators may be built for applications, operating systems, or hardware platforms. Emulation has been a popular strategy for retaining the functionality of old video game systems. The feasibility of emulation as a catch-all solution has been debated in the academic community.

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017

IMPACT FACTOR : 5.11UIF

- **Metadata attachment-** Metadata is data on a digital file that includes information on creation, access rights, restrictions, preservation history, and rights management. Metadata attached to digital files may be affected by file format obsolescence. ASCII is considered to be the most durable format for metadata because it is widespread, backwards compatible when used with Unicode and utilizes human-readable characters, not numeric codes. It retains information, but not the structure information it is presented in. For higher functionality, SGML or XML should be used. Both markup languages are stored in ASCII format, but contain tags that denote structure and format.
- **Digital sustainability–**Digital sustainability encompasses a range of issues and concerns that contribute to the longevity of digital information. Unlike traditional, temporary strategies and more permanent solutions, digital sustainability implies a more active and continuous process. Digital sustainability concentrates less on the solution and technology and more on building an infrastructure and approach that is flexible with an emphasis on interoperability, continued maintenance and continuous development. Digital sustainability incorporates activities in the present that will facilitate access and availability in the future.

Threats to digital preservation:

- **Massive storage failures**

  Basically no matter how much money you spend on the system housing your data there are still many ways in which it can fall over and create opportunities for data to be lost. This may be from hardware/software failure or an act of war. The longer you try to store data the more likely this will occur.

- **Mistaken erasure**

  Sometimes people accidentally delete things and if it's the only copy, then it's gone. On the other hand sometimes people think that they no longer need a

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017

IMPACT FACTOR : 5.11UIF

piece of data and delete it on purpose only to find that it was in fact useful. The longer you try to store data the more likely this will occur.

- **Bitrot**

  No affordable digital storage is completely reliable over a long period of time. For example some CD's have recently been shown to have a life span of only 2 years which could cause significant problems for anyone relying on them. Other media such as magnetic tape also suffers various types of bit rot. The worse thing about this threat is that is often undetected until it's too late to recover the material. You would very nearly have to employ someone to check all your media all the time to minimize data losses which would make most of these mediums too expensive to seriously consider in a preservation project. Bit rot is inevitable with any storage medium over a period of time.

- **Outdated media**

  Over time all kinds of digital media become outdated. Technology is driven by innovation which unfortunately leads to very short periods of relevancy before redundancy. Data stored on redundant media becomes effectively useless if the appropriate hardware is not available to read it. This is a particularly difficult issue to manage where data is stored over long periods of time. Ideally, long term data storage should be technology independent, however this is not practical. A Cornell University website (mentioned above in another post) has actually documented the lifespan of various storage media with floppy disks lasting a whopping five years.

- **Outdated formats, applications and systems**

  As hardware becomes redundant, so do file formats and the software which interprets them. A good example of this is *Word Perfect*; try to find a computer today which can read a *Word Perfect* document properly. Fortunately, system and format redundancy does not usually happen at quite as rapid a pace as hardware.

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017
IMPACT FACTOR : 5.11UIF

This is a difficult problem for long term storage and there are two common, but awkward, solutions. The first is to preserve a copy of the appropriate software and make it available wherever that data is stored. This becomes increasingly unmanageable as the types of systems required increases. The second is to migrate data to an acceptable format, for example all text files might be migrated to pdf thus only requiring copies of Adobe Acrobat to be preseved. However, during the migration process it is possible to lose data. It is also a costly process in terms of work hours and expertise.

- **Loss of context**

  Some data can be related, and this relationship can be vital to data interpretation. A good example of this might be the Rosetta Stone, discovered in Rashid, Egypt. The stone is engraved with hieroglyphics in three different languages and without the "key" of what these symbols meant none was able to read the inscription. It took a French scholar Jean François Champollion fourteen years to decipher the inscription. Can you imagine if you had to take that amount of time to decipher each document on your PC because someone had forgotten to preserve the relationship between that document and its key? It would be like trying to assemble like furniture without instructions, a complete waste of time. Unfortunately, if this relationship is not identified and preserved when information is first stored it is unlikely to ever be recovered. The longer the data is kept without this relationship, the less likely it is to ever be resolved.
- **Intentional attacks:**

  Unfortunately in the world we live in there are some people who intentionally destroy or damage digital assets for a variety of reasons. As much of the information is currently located in open access repositories accessible via the internet it is also vulnerable to attack. This is a threat to both long and short term storage.

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017
IMPACT FACTOR : 5.11UIF

- **Lack of resources**

  Many institutions simply do not have the resources, usually financial, to consider digital preservation. These strategies are often overlooked as low priority and are likely to remain so until a major data loss scares people into action.

- **Organizational failure**

  This is a massive threat to long term digital storage of any kind. Technology is so dynamic not only in innovations but also movement with vendors and competition killing off what seemed to be at one point very strong tech players. For this reason it would be a folly to rely too heavily on any one vendor/system/sponsoring organization because they change and often change quickly. Digital assets which need to be preserved long term must be protected from the failure of any one organization. Unfortunately this is easily said but hard to plan for in such a dynamic environment.

Conclusion: Digital preservation is thus very important aspect to preserve important information. But there are some points to keep in mind so that the preservation actually beneficial. Keeping in mind these problems and the threats observed the procedure of preservation to be followed.

References:

1. M. Altman and G. King. “A Proposed Standard for the Scholarly Citation of Quantitative Data”,D-Lib, 13, 3/4 (March/April). 2007.
2. M. Altman, et. al, “Data Preservation Alliance for the Social Sciences: A Model for Collaboration” Proceedings of DigCcurr07, Chapel Hill. April 2007.
3. G. King, “An Introduction to the Dataverse Network as an Infrastructure for Data Sharing”,Sociological Methods and Research, 32, 2 (November, 2007): 173–199.
4. M. Altman , "A Fingerprint Method for Verification of Scientific Data" in, Advances in Systems,Computing Sciences and Software Engineering,

ISSN: 2321-5488 Volume – 5| Issue – 1| January, 2017

IMPACT FACTOR : 5.11UIF

(Proceedings of the International Conference on Systems, Computing Sciences and Software Engineering 2007) , Springer Verlag. 2008.

5. Altman, M., Adams, M., Crabtree, J., Donakowski, D., Maynard, M., Pienta, A., & Young, C..“Digital preservation through archival collaboration: The Data Preservation Alliance for the Social Sciences.” The American Archivist. Forthcoming 2009.

6. Ferrara L., Young, C. (in press). “From Preserving the Past to Preserving the Future: The Data- PASS Project and the challenges of preserving digital social science data”. Library Trends.Forthcoming 2009.

7. Altman, M., “Transformative effects of NDIIPP, the case of the Henry A. Murray Archive”.Library Trends. Forthcoming 2009.

8. Micah Altman ; Bryan Beecher,; Jonathan Crabtree; with Leonid Andreev, Ed Bachman, Adam Buchbinder , Steve Burling, Patrick King, Marc Maynard, “A Prototype Platform for Policy-Based Archival Replication“ Against The Grain, (Forthcoming) Winter 2009.