MERIT RESEARCH JOURNALS
www.meritresearchjournals.org

Merit Research Journal of Education and Review Vol. 1(3) pp. 058-063, April, 2013
Available online http://www.meritresearchjournals.org/er/index.htm

Review

Uncertainties of digital information and preservation in today's and tomorrow's generation: an idea

Raghunadha T. Reddy[1], Dr. K. Kumar[1] and T. Sreenivasa Rao[2]

[1]Librarian, SVCET, Chittoor-517127
[2]Assistant Librarian and Head, ANGRAU Regional Library. S.V. Agricultural College, Tirupati – 517 502

Accepted April 05, 2013

The digital information generated by organizations from domains such as government, cultural heritage, science and education, audio, video, films, health, insurance, banking, etc are part of our digital universe or each one is a digital universe in its own right. There is good understanding in the Library of preservation issues and of the constraints of preservation strategies, which are taken into account in deciding what to take into the Library's collections and how it should be managed. So this paper is brief discussion of an idea: uncertainties of digital information and preservation in today's and tomorrow's generation.

Keywords: Digital Information, Preservation, Problems

INTRODUCTION

Preservation strategies in academic and research libraries are not new concepts. However, with an increasing amount of digital content, organizations have to cope with a new set of preservation issues. Digital preservation is in its infancy worldwide and presents some difficult technological issues. Since the creation of digital media, over 200 different storage mediums have been invented ranging from magnetic tape to CD-Rom. Each of these mediums presents a variety of their own preservation issues and also requires a diverse range of technology which in many cases is no longer manufactured. In addition to this, there are thousands of different formats in which data can be stored on each medium; and each type of storage format may also require a specific piece of software to interpret the data's meaning. National Study Report on Digital Preservation Requirements of India is conducted by Human Centered Design and Computing Group of C-DAC, Pune was entrusted with the responsibility to produce the national study report on digital preservation requirements of India by Department of Information Technology, Government of India during 2010-11.

The report has been prepared by involving the domain experts from 30 stakeholder organizations across India. It is duly submitted with Department of Information Technology for further processing and implementation.

As for the digital Universe Study Report by International Data Corporation, estimated size of Digital Universe in 2010 is 1.2 million Pentabytes or 1.2 Zettabytes. In 2020 may be around 35 Zettabytes (Includes the digital information from government, enterprise, social segments and all major forms of media –voice, TV, radio, print which would have completed the journey from analog to digital.)Estimated Size of Unprotected Data needing protection in 2020 is 18,000 Exabyte.

*Corresponding Author E-mail: raghuthima@gmail.com

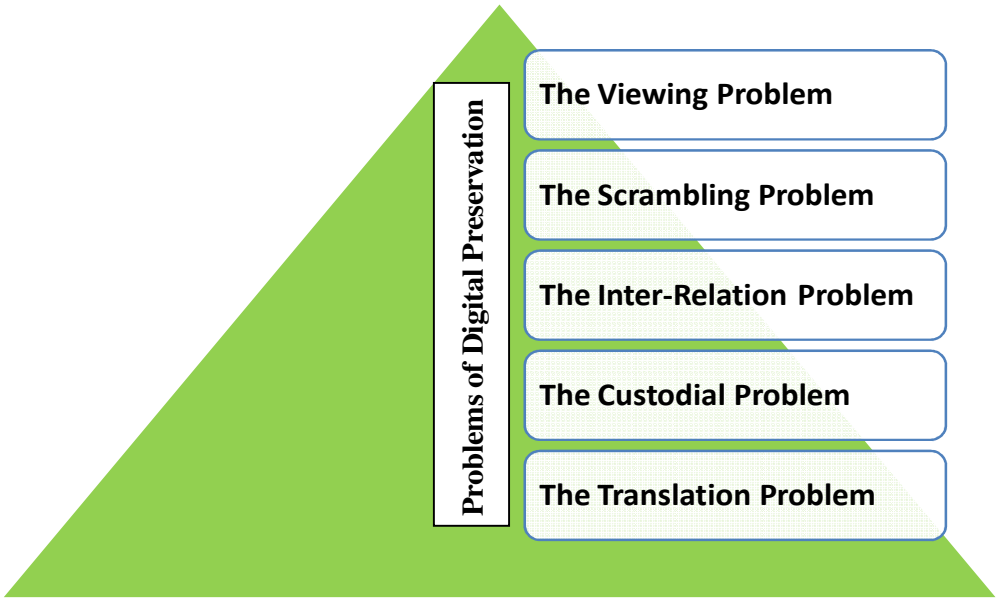

Figure 1. Problems in the area of Digital Preservation

Digital preservation objectives

Pursue the goal of long-term digital preservation of digital information (i.e. inclusive of born digital and reformatted digital content) across diverse domains; by creating the tools, technologies, best practices and sustainable infrastructure; to ensure usability of digital information through passage of time and obsolescence.

- To development in digital preservation to produce the required tools, technologies, guidelines and best practices.
- Spread awareness about the potential threats and risks due to digital obsolescence and the digital preservation best practices
- Develop the digital preservation repositories and provide help in nurturing the network of Trustworthy Digital Repositories as a long-term goal
- Define the digital preservation standards by involving the experts from organizations, consolidate and disseminate the digital preservation best practices generated through various projects.

Problems in the area of digital preservation

There are no clear standards in the area of digital preservation and with most institutions lacking resources already, how are they to tackle these issues? It has synthesized the issues surrounding digital longevity into five general areas. These are: (Figure 1)

The viewing problem

All digital formats require computer technology to view

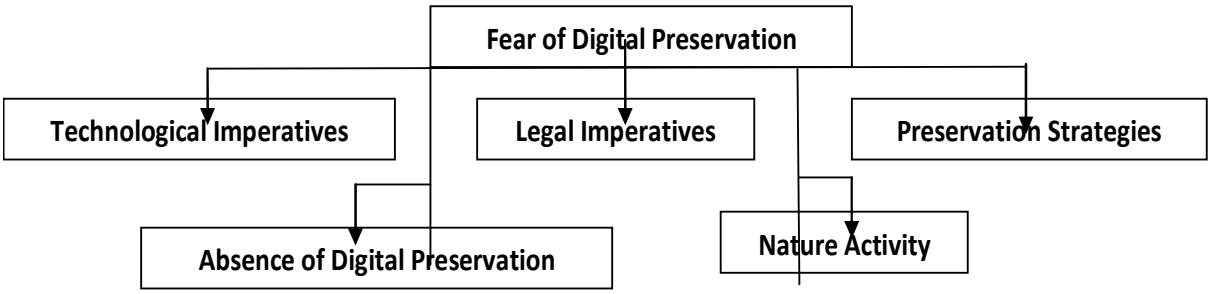

Figure 2. The fear of Digital Preservation in the future

them. By nature technology (software/ hardware/formats) move at such a rapid pace that, odds are, they won't be around when you want to view your data. This is unless you're viewing data right after you "preserve it" in which case, it's not really preserved now is it.

The scrambling problem

Data is often compressed or "scrambled" to assist in its storage and or protect it is intellectual content. These compression and encryption algorithms are often developed by private organizations that will one day cease to support them. If this happens you are stuck between a rock and a hard place. If you don't want to get into legal trouble you are no longer able to read your data; and if you go ahead and "do the unwrapping yourself" it's quite possible you are breaking copyright law.

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 are often overlooked.

The custodial problem

Who is the custodian of a digital document? Is it a librarian's job? What if someone changes the content without telling the custodian, after all digital content is dynamic and easily changed. So does the documents' custodian have to undertake version control? And then is it really preservation?

The translation problem

If we need software to interpret data 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 notices them, or does it really matter at all.

The fear of digital preservation in the future

Technological Imperatives

- Non-tangible bit stream form of digital data
- Obsolescence of file formats
- Obsolescence of computer hardware and software
- Obsolescence of storage devices and media
- Failure of devices
- Physical threats

Legal imperatives

- Information Technology Act
- Right To Information Act
- Public Records Act
- Indian Evidence Act
- Banking Regulation Act
- Banking Books Evidence Act
- Copyright Act / Digital Rights

Absence of digital preservation strategies

- Lack of awareness about digital obsolescence
- No digital preservation policy
- No standards, guidelines and best practices

Figure 3. Threats to digital preservation

- No tools and technologies
- No digital preservation infrastructure
- No digital preservation skills
- No quality control and audit procedures

Digital preservation strategies

- Awareness
- Digital Preservation Policy
- Digital Preservation Standards and Best Practices
- Digital Preservation Tools and Technologies
- Digital Preservation Course Curriculum
- Trustworthy Digital Repository
- Audit and Certification
- National Digital Preservation Infrastructure

Nature of activity

- Domain specific solutions
- Dynamic, cyclic, continuously evolving
- Long term commitment
- Dependable and trustworthy
- Interoperable
- Protective
- Accessible and Usable
- Sustainable

Fragile media

We can lose the digital content through poor selection of storage mediums. The progression of technology often motivates institutions to move from one storage medium to another through forced obsolescence. However it is interesting to note that new and "improved" mediums do not always deliver the same level of functionality. Ironically, many libraries who initially began migrating data to digital repositories under the banner of "economy" have found that it is too costly to keep up with technology and are using microfilm and acid-free paper to back up digital content. The "myth of the 100 year CD-Rom" is a good example of how we have been sold a potentially disastrous product for digital preservation on many levels whether it's personal photos, or important research data. When you compare the potential preservation value of a CD-Rom with an undetermined shelf life beginning at only two years, all of a sudden keeping paper records doesn't seem too bad.

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 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.

Bit rot

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 a period of 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. It 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. 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 migrated data to an acceptable format, for example all text files might be migrated to PDF thus only requiring copies of Adobe Acrobat to be preserved. 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. 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 Ikea 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.

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

Preservation strategies in academic and research libraries are not new concepts. However, with an increasing amount of digital content, organizations have to cope with a new set of preservation issues. Digital preservation is in its infancy worldwide and presents some difficult technological issues. Since the creation of digital media, over 200 different storage mediums have been invented ranging from magnetic tape to CD-Rom. Each of these mediums presents a variety of their own preservation issues and also requires a diverse range of technology which in many cases is no longer manufactured. In addition to this, there are thousands of different formats in which data can be stored on each medium; and each type of storage format may also require a specific piece of software to interpret the data's meaning.

REFERENCES

Aregu R (2001). Digitizing agricultural data for rapid agricultural modernization in Uganda: strengthening Ugandan NARS issues. In *Digital libraries: dynamic landscapes for knowledge creation, access and management*, Conference papers of the 4th International Conference of Asian Digital Libraries.

Breaks M (1999). "Management of electronic information". In: Proceedings of the future of libraries in human communication: *International Association of Technological University Libraries Conference*, Chania, Greece, May 17-21, 1999.

Fisser P (2001). "Using information and communication technology: a process of change in higher education". *Ph.D. thesis. Netherlands*: Twente University Press

Meyyappan N, Chowdhury CG, Foo S (2001). Design and development of a user – centered digital library system: some basic guidelines. In *Digital libraries: dynamic landscapes for knowledge creation, access and management*, Conference papers of the 4th International Conference of Asian Digital Libraries

Mutula S (2004). "IT diffusion in Sub-Sharan Africa: Implications for developing and managing digital libraries". *New Library World*, vol. 105 no.7/8, 281 – 289.