

A survey of preservation of digital resources in the national libraries with the membership of the International Internet Preservation Consortium (IIPC) and proposing a preservation plan for the National Library & Archive of Iran

Mitra Samiee ; Fellow member of the National Library and Archive of the Islamic Republic of Iran, Tehran, Tehran/Iran
Kuhyar Davallu; Fellow member of the Art University; Tehran, Tehran/Iran

Abstract:

Purpose: This research aiming at a survey of the existing status of digital preservation among the members of the IIPC was carried out by the identification of maintenance methods and strategies, digital repositories, standards, storage formats and tools and levels of access and security.

Method/ approach: The research is based on a descriptive and analytic survey. According to this method, the community in question has been surveyed within two groups of digital archiving and permanent access as regarded digital preservation using questionnaires. The questionnaires were sent electronically to the community in question including 11 members of the IIPC.

Finds: 90/91 of national libraries apply on-line magnetic media and tape library, 81/82 SAN technology, 90/91 METS standard and the OAIIS reference model, 90/91 PDF format, TIFF, MP3 and WAVA, 100 back-up supply, 45/45 access to the entire collection for free, access with restricted copyright and free access to the collection of digital records in part and 100 the mechanism of verification and access control management.

Conclusion: According to the results of this research, on-line magnetic media and tape library are appropriate for storage and long-termed preservation of back-ups and digital resources at national libraries due to their high quality and considerable life span and the SAN for better integration and allowing for the sharing of back-up facilities.

Introduction

From the very outset when the digital collections for libraries began to develop, the issue of their preservation and maintenance has been an inseparable part of the whole thing. As the bulk of digital resources increases, new information in a variety of scientific areas will be incorporated into fragile, unstable and viable media more and more. The loss of these library resources will have irreparable and detrimental effects on the posterity's intellectual perception. Therefore due to this issue's significance, IIPC (International Internet Preservation Consortium) which is composed of the world's considerable national libraries, formed in order to support the preservation of digital resources. The goals pursued by this international consortium, include: to perform group works within the legal framework of each country for identification, to develop and facilitate the application of some solutions for selection, collection building, preservation and

facilitation of access to internet information sources and global support of initiatives which would promote collection building and preservation and permanent access to digital resources. A proper understanding of the resources' characteristics and possible threats to the permanent access to them is the chief point which should be taken into account in order to achieve this purpose. In terms of fragility and loss of quality, digital storage media are vulnerable and there is no precise method of supporting and refreshing hardwares and softwares and new information carriers. So new procedures and strategies should be designed and implemented. [3] The other issue is the supply of the stored digital items in future. The question is how we can be ascertained that we will see and utilize the digital items in five, ten or fifteen years to come. Today in most of libraries and archives for the long-termed preservation of digital resources such strategies as below are adopted: bit-level preservation, information migration, technology emulation, technology preservation, information encapsulation, supply of back-up copies and etc. To this end, some tools are required to guarantee permanent access [10]. Despite striking technical achievements as regards the speed of processing and storage capacity, the preservation of digital material will go further than accumulation of information on a high capacity disc and again its retrieval. The problem is that we should be able to make the stored information accessible rather than making this digital information just presentable and further more, we should be able to keep them alive and accessible for the next generations and at the same time maintain the coherence and homogeneity of information as well as the intellectual property rights and copyrights of their owners. From 2000 onwards many of digital national libraries in the world including the USA congress have developed an initiative for the establishment of an appropriate infrastructure with a view to managing and preserving their valuable works in digital format. The library of American congress within the framework of NDIIPP (The National Information Infrastructure and Preservation Program) [12] the National Library of Australia within the framework of ICABS's third clause concerning the collection and preservation of digital resources and Netherland's National Library by means of DIAS (Digital Information Archival System) and other National libraries through other various programmes and projects have dealt with the preservation of their valuable digitized resources [7].

Similarly in Iran since 2001 some initiatives have been adopted by the NLAI to digitize the library resources and the development of a digital library has been officially proposed. To begin with a working group consisting of expert librarians and computer engineers embarked on a survey and review of research projects carried out throughout the world on digital library. This was done within a research project entitled " Digital National Memory "(HAMD)in research assistance , technology and programming . Later in 2008 the very project with a few changes under the title "digital national library" was put into operation. The project was intended to digitize all valuable and exclusive resources and to make them available to the public. Lots of old resources including manuscripts , publications of Ghajar period and old documents were scanned and digitized and were organized in a software named "Nama". Along with the increase of these digitized resources, some arrangements were to be made for the long-termed preservation and permanent access to them. This research seeks to put forward a practical and suitable paradigm for the preservation of the digital resources presently existing in the NLAI ,once it has surveyed the methods and strategies of digital preservation and the control access and information security in the national libraries which are members of IIPC.

History/background of the research

The preservation of written heritage, regardless of its format, is of paramount importance to all communities. Digital objects are prone to obsolescence, for they have been recorded with preservation strategies in a particular format on temporary and short-lived media and therefore require an encoding paradigm for use. So, it would be more reasonable to preserve all digital data permanently on standard and durable media by choosing an appropriate digital preservation strategy. Since the subject "Digital preservation " is new and unfamiliar, no research has been carried out on it in Iran, but many researches have been done on the very subject by libraries, archives and international organizations abroad of which I should here like to give a short account.

Titia Vander Werf-Darelaar (1999) in his article entitled" Long-term preservation of electronic publications "looks into NEDLIB research project and the technicalities of digital preservation strategies. The main purpose of this project is to give us insight into the advantages and disadvantages of different strategies adopted for long-termed preservation and to define authenticity in digital preservation. This project has taken the very first step to assess technicalities of preservation strategies by means of an experiment of technology emulation .Results of this experiment show assuming that appropriate emulators for obsolete platforms can host the future platforms, emulation can act as the foundation in this process. [9]

In a report of CEDARS research project (2002), the project along with the issues regarding preservation and methods of access to resources as well as digital preservation strategies have been analyzed .CEDARS is at the moment implementing pilot projects to assess and promote a selected strategy for digital preservation. The working group of CEDARS's data of preservation strategies attend to those preservation issues relating to migration, emulation and refreshment. [2]

Svein Arne Solbakk (2003)in his research article entitled " Critical technological and architectural choices for access and preservation in a digital library environment" describe some basic architectural choices for the access to and preservation of digital objects at the National Library of Norway .A digital repository is a core element for the handling of both access to and preservation of the digital objects .Strategies for giving access to the complete holdings include the use of a powerful search engine and the OAI protocol to harvest metadata from conventional catalogue systems to make textual or structured indexes. [8]

Reinhard Altenhoner (2006) in his article entitled "Data for the future, the German project, co-operative development of a long-term digital information archive (Kopal)" looks into Kopal research project and the long-term accessibility of digital documents, long-term preservation methods and strategies will be discussed in the paper .one of the unresolved problems of the global information society is ensuring the long-term accessibility of digital documents. The project Kopal tackles this problem. The Kopal system will be implemented in accordance with the Open Archival Information System (OAIS) framework. The project will present a stable and reusable platform for additional partners and users, especially for cultural heritage. [1]

Most of the researches in the area of digital preservation strategies in the libraries were carried out after digital libraries had been established and the challenges of digital resources preservation had arisen .Since 1990 when the concept of the digital library was for the first time brought up in the world and many of the National Libraries began to digitize their resources, the preservation strategies and maintenance of digital resources have been an inseparable part of the development of the digital collections for libraries .While making policies concerning long-termed digital preservation strategies ,many aspects such as cost-efficiency ,legal restrictions and users access requirements should necessarily be taken into consideration .With regard to a variety of digital preservation strategies in the survey of research projects, it was observed that in most of these projects, back-up supply, refreshment of media, migration to new media and technology emulation have been proposed as appropriate strategies for the preservation of digital resources.

Queries of research

- 1.How do the national digital libraries preserve their own resources?
- 2.which standards do these libraries apply for the long-termed storage of their resources?
- 3.Which strategies of digital preservation are applied at national libraries?
4. How is access control carried out at national libraries ?

The method of survey and information collection

In this research , a descriptive and analytic survey has been employed . according to this method the community in question has been surveyed in two groups digital archiving and permanent access regarding digital preservation and using questionnaires as a tool. The questionnaires were sent electronically to the community

in question including 11 members of the IIPC .The collected information were entered into the computer by means of E-mail and were processed and analyzed using the two softwares excel and SPSS .

Finds of the research

In this research the status of digital preservation of national libraries was analyzed covering many aspects such as digital archiving , storage tools , formats of digital resources preservation , way of storing resources at digital repositories , standards , and

way of users accessing the stored digital resources . Data analysis was done in response to the crucial queries of this research.

Query 1 . How do the national digital libraries preserve their own resources?

In this query , the portable storage and mass storage tools , were looked into .

Having a proper understanding and insight into different storage media is of great importance storage media have a wide range of proportion rate in terms of storage capacity for preservation and access .

Table1. All types of portable storage media

Type library	Tape library	Magnetic Tape	Online magnetic media	DVD	DLT
Italy	1	0	1	0	0
Australia	1	0	1	0	1
Iceland	1	0	1	0	0
The British Library	0	1	0	1	0
Denmark	1	0	1	1	0
Sweden	1	0	1	0	0
France	1	0	1	0	1
Canada	1	0	1	0	0
Norway	1	0	1	1	0
Finland	1	0	1	0	0
The Library of Congress	1	0	1	0	0
total	10	1	10	3	2
percent	90.91	9.09	90.91	27.27	18.18

In this respect, the finds show that %90.91 of national libraries apply online magnetic media and tape library,%27.27 DVD , %18.18 DLT and %9.09 magnetic cassette for storage and preservation of back-ups at libraries. Regarding mass storage technologies the finds show that %100 of the community in

question used SAN technology ,%45.45 DAS technology , %27.27 NAS for mass storage of their digital resources.

Query2. which standards do these libraries apply for the long-termed storage of their resources?

In this query , the storage formats and standards and standard of digital preservation of OAIS were surveyed.

Table2. required standards of digital preservation[5], [6]

Type library	METS	OAIS	PREMIS	DOI	VRA	EAC	MODS
Italy	1	1	1	0	1	1	0
Australia	1	1	1	1	1	1	1
Iceland	1	1	1	1	0	0	1
The British Library	1	1	1	1	0	0	1
Denmark	1	1	1	1	1	0	1
Sweden	1	1	0	0	1	0	1
France	1	1	1	0	1	1	1
Canada	1	1	1	1	1	1	0
Norway	1	1	1	1	1	1	1
Finland	1	1	1	1	1	0	1
The Library of Congress	1	1	1	1	1	1	1
total	11	11	10	7	9	6	9
percent	100	OAIS	90.91	72.73	81.82	54.55	81.82

The finds show that almost %100 of the community in question apply METS standard and the OAIS reference model ,%90.91, PREMIS, %81.82 MODS and VRA , %72.73,DOI, and %54.55 EAC . The wide usage of these standards and metadata by the national libraries in question stands for the importance of these standards in digital preservation of resources.

About the formats of storing the information of text , image and audio-visual files in the national libraries[4], The finds show that almost PDF,TIFF,mp3,WAVE formats in%90.91 ,PDF/A format

in %81.82,MPEG format in %72.73 ,DOC format in %63.64,HTML ,XML and JPEG2000 formats in % 54.55,JEPG format in %45.45 and ASCII format in %36.36, national libraries have been applied .

Query 3. Which strategies of digital preservation are applied at national libraries?

In this query , digital preservation methods and strategies in national libraries were surveyed.

Table3.Types of digital preservation strategies

Type library	Technology preservation	Technology Emulation	Information Migration	Encapsulation	Digital Archaeology	Data backup	Persistent Identifiers
Italy	0	1	1	1	0	1	1
Australia	1	0	1	1	1	1	1
Iceland	0	0	0	0	0	1	0
The British Library	0	1	1	1	1	1	1
Denmark	0	0	1	1	0	0	1
Sweden	0	0	1	1	0	1	1
France	1	0	1	1	1	1	1
Canada	1	0	1	1	1	1	1
Norway	0	0	1	1	0	1	1
Finland	0	0	1	1	0	1	1
The Library of Congress	1	0	1	1	0	1	1
total	4	2	10	9	4	11	10
percent	36.36	18.18	90.91	81.82	36.36	100	90.91

%100 of libraries apply the strategy supplying back-up ,%90/91 of the Libraries use information migration strategies and permanent identifiers to locate digital resources and for the long-term preservation of their digital resources.% 81/82 employ information encapsulation strategy in OAIS reference model.%36/36 have attempted to use the strategy of technology preservation and digital archaeology and % 18/18 apply technology emulation strategy.

Query4. How is access control carried out at national libraries ?

In this query , The range of users, methods of accessing digital resources, Methods of protecting digital objects, types of mechanisms of security management and access control in national libraries were surveyed.

Table 4.methods of accessing digital resources

Type library	free access to the entire collection	access by restricting copyright	access to the library digital resources chargeable to users	free access to a part of the collection of digital resources	delivering them in the form of CD_ROM and DVD	visit the digital archive of library and authorize them to access the resources.
Italy	0	1	1	1	0	0
Australia	1	1	0	0	0	0
Iceland	1	0	0	0	0	0
The British Library	0	1	1	1	1	0

Denmark	0	1	0	0	0	0
Sweden	0	1	0	1	1	0
France	0	1	0	1	1	1
Canada	0	1	0	0	1	0
Norway	1	0	0	0	0	0
Finland	0	1	1	1	0	0
The Library of Congress	0	1	0	0	1	1
total	3	9	3	5	5	2
percent	27.27	81.82	27.27	45.45	45.45	18.18

%81.82 provide the users with the method of access by restricting copyright. 45.45 % of the libraries allow for free access to a part of the collection of digital resources and also delivering them in the form of CD-ROM and DVD .% 27.27 of the libraries allow for free access to the entire collection and access to the library digital resources chargeable to users .and % 18.18 too enable the users to visit the digital archive of library and authorize them to access the resources.

About Methods of protecting digital objects against the unauthorized entrance of user ,the finds show that %100 of the community

being studied apply the method of fire walls and virus protection and also they use updated antiviruses against the penetration of viruses into their digital library system. %72.73 utilize the method of firm authorization of the users to benefit from the digital resources of their library. %54.55 draw on security alarms to guard digital objects against the invasion of viruses and the users intrusion.Fire walls and virus protection and the use of updated anti-viruses rank as the most frequent methods of protecting digital objects against unauthorized entrance of users) and penetration of viruses into libraries.

Table 5. All types of mechanisms of security management

Type library	key management	encoding	digital signature	access control	verification mechanism	control and monitoring routes
Italy	1	1	1	1	1	1
Australia	1	1	1	1	1	1
Iceland	0	0	0	1	1	0
The British Library	1	1	1	1	1	1
Denmark	0	1	0	1	1	0
Sweden	1	1	0	1	1	1
France	1	1	1	1	1	1
Canada	1	1	1	1	1	0
Norway	1	1	0	1	1	0
Finland	1	0	1	1	1	1
The Library of Congress	1	1	1	1	1	1
total	9	9	7	11	11	7
percent	81.82	81.82	63.64	100	100	63.64

%100 apply verification mechanism and management of access control for the security of the resources information at their digital library system .%81.82 use key management and encoding. % 63.64 make use of control mechanism and monitoring routes and digital signature to guarantee the security of resources information of digital and national libraries system .None of the national libraries being studied has drawn on the mechanism of the management of data integrity. Thus, it was excluded from the table.

Discussion and conclusion:

Studies carried out in the area of digital preservation and the survey of the status of digital preservation strategies in the National Libraries with the membership of IIPC have led the researcher to some finds which could pave the way for the implementation of appropriate strategies of digital preservation in the NLAI .

- According to the results of this research online magnetic media and tape library are appropriate tools for storage

and long-term preservation of back-ups in national libraries due to their high quality and life – span. Also SAN technology is the most suitable choice for the mass storage of digital resources .

- The SANs are regarded as one of the most common technologies in mass storage in most of national libraries because of their better integration and allowing the sharing of back-up facilities and keeping the crowd and accumulation between storage tools (for example for the back-up) far from internet networks.
- the design and implementation of the OAIS established a paradigm or a general framework for the construction of a trusted information repository for long-term preservation and access to digital resources of national libraries.
- The mets standard and the OAIS reference model are used to describe the requisite information for implementation , documentation and evaluation of digital preservation processes in national libraries because the establish a basic and comprehensive framework and afterwards EAD metadata and PREMIS and VRA should be implemented for national libraries.
- There is a wide variety of formats for various types of information resources(including textual, audio , video and image, webpages, databases). The largest amount of stored information at the national libraries in question are in textual and then image files , the textual files in PDF and the image files have been store in TIFF and JPEG .The finds show that WAV and MPEG are more frequent formats in audio-video files .
- in most of the National Libraries in question, a blend of the strategies "back-up supply" ,"information migration" and "permanent identifiers" has been adopted to locate and preserve the digital resources of libraries in the long term. Cost-effectiveness and user access requirements to digital resources are considered the most important and frequent factors in the selection and use of an appropriate preservation strategy at libraries. The strategy of migration aims at the maintenance of intellectual content of digital objects and accessibility for the users to retrieve information.
- Defensive procedures such as fire-walls, virus protection and the use of updated antiviruses against virus invasions have been applied in the system of national digital libraries.
- The mechanism of access control management, verification, key management and encoding as mechanisms of security management have been employed to implement the security of information within the system of digital national libraries in question.
- According to the result, %81/82 of libraries make use of the method of access with restriction of copyright in order to access the digital resources of their libraries. Therefore, it is proposed to the NLAI to apply access with restricted copyright for high-quality copies and to apply free access at the site for low-quality access.

Suggestions

- writing a suitable software for the digital library of existing resources in NLAI
- implementation METS standard and the OAIS in software system of national library in NLAI
- implementation of the proper security system in software system of national library in NLAI
- application of preservation strategy and proper storage tool for digital resources and its implementation at NLAI
- application of proper descriptive and preservation metadata (PREMIS , MODS) in software system of national library in NLAI
- interoperability of this metadata and security and preservation standard in software system of national library in NLAI
- Making a written and comprehensive policy for digital preservation at NLAI.

Reference:

- [1] Altenhoner, R. (2006). Data for the future , the German project (co-operative development of a long-term digital information archive, Kopal). *Library Hi Tech*, 24(4), 574;
- [2] Calanag, M. L., Tabata , k., & Sugimoto, S. (2002). *Linking collection management policy to metadata for preservation – a Guidance model to define metadata description levels in digital archives*. Paper presented at the proc.Int.Conf on Dublin Core and Metadata for e-Communities2002. [On-line] available: <http://www.bncf.net/dc2002/program/ft/paper4.pdf>. (accessed 6 April 2008);
- [3] Gatenby, P. (2005). " Getting started: what needs to be in place to maintain access to digital collections". from [http://archive.ifla.org/IV/ifla71/papers/032e-Gatenby.pdf\(6Aug.2011\)](http://archive.ifla.org/IV/ifla71/papers/032e-Gatenby.pdf(6Aug.2011));
- [4] Library of congress (2007). " Formats, evaluation factors, and relationships". 10 Feb2008, from http://www.digitalpreservation.gov/formats/intro/format_eval_rel.html (6Aug.2011);
- [5] Idem. (2011). "Standards at the library of congress". from <http://www.loc.gov/standards/> (6Aug.2011);
- [6] "Mets :Metadata Encoding & Transmission Standard". (2010). from [http://www.loc.gov/standards/mets/\(6Aug.2011\)](http://www.loc.gov/standards/mets/(6Aug.2011));
- [7] Nordland, Lori. Podolsky. (2007). " The long and short of IT : the International Development Research Centre as a case study for a long-term digital preservation strategy". A thesis Master of Arts of Archival Studies. Winnipeg, University of Winnipeg and University of Manitoba;
- [8] Solbakk, S. A. (2003). Critical technological and architectural choices for access and preservation in a digital library environment. *Library Review*, 52(6), 251-256;
- [9] Werf-Davelaar, T. v. (1999). Long-term preservation of electronic publications: The NEDLIB project. *D-Lib Magazine*, [on-line] available: <http://www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html>;
- [10] Wijnagaarden, H. v. (2004). An exploration of strategies and recent developments in digital preservation: outline of an international inquiry by the Koninklijke Bibliotheek. Retrieved from <http://archive.ifla.org/IV/ifla70/papers/025e-Wijnagaarden.pdf>.

Author Biography

Mitra Samiee obtained a BA degree in Library and Information Science in 1998, then received an MA in Library and Information Science in 2002., Then received PhD in Library and Information Science in 2010. She is Fellow member of staff NLAI.. She work in the doyenne of researches department, university lecturer and in charge of HARAM (i.e. National Digital Memory) project's standards and digital content in the NLAI .She is Manager website of The department of Iranian & Islamic study researches, the NLAI. She teaches as a visiting lecture in the At the management faculty of Tehran university, National library's department of administration, the organization of management and country planning, Azad university of Iran, the broadcasting Company of the Islamic republic of Iran .She is also a member of the Iranian Library and Information

Science Association. She is written 27 papers and 2 books which are published in Iran. Her research interests include digital libraries, digital preservation, PREMIS and its usage in digital libraries, and data security and so on. Mitra Samiee is the corresponding author and can be contacted at: samiei.mitra66@gmail.com

Kuhyar Davallu obtained a BA in English translation 1996 and MA in English literature from Tehran University in 2000. He has been teaching ESP courses to art MA students for the past 12 years (at state universities). Furthermore He has been contributing to library and information science and IT especially by writing a few articles on the field and also by collaborating with a few projects at the NLAI such as the development of Islamic and Iranian studies website .kuhyar Davalu is the corresponding author and can be contacted at: kuhyardavallu@yahoo.com