Journal of Information and Knowledge Management (JIKM) Vol 1 Special Issue (2022)

The Reliability of Data Integriti Preservation in Archival Institutions

Normashidayu Abu Bakar[1], and Nordiana Mohd Nordin[1]

[1]School of Information Science, College of Computing, Informatics and Media, UiTM Selangor Branch, Puncak Perdana Campus, 40150 Shah Alam, Selangor, Malaysia

Email: 2021155511@student.uitm.edu.my

Received Date: 30 August 2022
Accepted Date: 21 September 2022
Published Date: 1 November 2022

Abstract. Digital preservation is a broad term that includes everything that needs to be done to keep digital materials accessible even if the media fail or technology changes. The advent of digital storage has profoundly altered our surroundings. Data is growing in size, thus individuals are migrating their data to the digital. This research will identify the factors that influence digital preservation, identify the benefits of digital preservation in archival institutions, measure the extent of digital preservation among members of archival institutions, and identify the data integrity of digital preservation in National Archives of Malaysia.
Keywords: Digital archives, Data integrity, Archival institutions, Cloud storage, Records management and Information management.

1 Introduction

Digital preservation is defined as “the series of managed activities necessary to ensure continued access to digital materials for as long as necessary and refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological and organisational change (Lynam, 2021). Data Integrity is described as "the extent to which data are comprehensive, consistent, accurate, trustworthy, and reliable and that these attributes are maintained throughout the data's life cycle." (PIC/S, 2021). Today's archivists and records administrators face challenges in ensuring that digital records are legitimate, dependable, and usable for the long haul. The life expectancy of the medium the data is stored on is determined by the technology used to produce them. Digital materials, however, are constantly at risk of being corrupted or tampered with because of all the activities taken to use, store, manage, and even preserve them. The profession has yet to reach a consensus on how to best protect

digital data's lifespan. In today's digital age, organizations are increasingly vulnerable to various security risks, including those posed by information systems such as viruses and hacking tools (Adu, 2014). Two main variables contribute to digitalization in the archive institution: application and human factor characteristics. For the sake of archival preservation, the data must be kept in immaculate condition so that it may be accessed by the public when needed. To guarantee that digital preservation is accessible, trustworthy, and used for future generations, archival institutions must maintain conservation integrity. According to this study, it concerns data loss, manipulation, and data corruption. A record's authenticity is enhanced as it moves from the site of origin to the preservation location.

According to Duranti (2007), records that have been separated from their initial place can be preserved in archives because they meet three fundamental criteria: transparency in preserving records, confidentiality, and consistency. Records must be made transparent. Preservation means that someone can be trusted to keep the records safely. Possibly more than ever before, the digital world has necessitated that archivists develop new solutions briefly, adapt theories and methods from other fields, and interact with disciplines that have not been archivists' traditional standards. It resulted in the existential conversation about what archives do and the development of new methodologies, funding options, methods of organizational cooperation, and ways of thinking. Currently, archivists recognize that to "sell" their needs and objectives for records operations, they must establish persuasive theoretical solutions and learn how to implement them in practice. To keep their validity, trustworthiness, integrity, and usability, "Electronic records must be actively managed" (International Council on Archive, 2008). This management includes digital preservation operations to ensure these features are maintained over time.

2 Literature Review

The vulnerabilities that digital information faces and possible solutions face are well-documented in the academic and professional literature (Zsuzsanna, 2014). Challenges can be broken down into two categories: Preserving their potential to identify, process, render, interpret, and make use of the bit streams by minimizing the damage to the data element, which includes the storage devices and the bit streams recorded there (passive preservation) (active preservation). The first point deals with the numerous dangers to the storage media affecting the stored coding system. Although the medium will degrade with time even if preserved in the best possible circumstances, media collapse and operational faults can negatively impact its usefulness. Digital content must be regularly migrated to newer formats (Zsuzsanna, 2014). As a result, a tight maintenance schedule for data storage and systems is needed to safeguard data from viruses, power loss, natural disasters, unauthorized access, and intentional destruction.

2.1 Authenticity

ISO 15489 defines authenticity as a record that is what it claims to be and is made or sent by the person at the time it is claimed (InterPARES, 2002). Because data must

be safeguarded against illegal addition, deletion, modification, and use and concealment to be authentic, the term not only pertains to their identification but also involves their integrity, which is defined as being entire and unaltered. There are no alterations or corruptions in authentic records. An item's identity and whether or not it has been altered also implies that there was an original from which to compare it. As soon as it is saved and closed in a digital realm, the original records, even as the creator made it on the screen, disappear (InterPARES, 2002).

2.2 Data Integrity

An original state is what the IT Governance Institute (ITGI 2004:22) defines as "integrity." According to this definition, information integrity refers to how accurate a representation is to the condition or subject matter it portrays. Bovee et al. (2003) explain the four parts of integrity, which is a part of how information is made:

2.2.1 Accuracy.

This information fits with the natural world or ideas that the user is interested. Most people think it doesn't have any mistakes. Completeness means having all the parts we need or having enough information to make a choice.

2.2.2 Consistency.

Values for any of the attributes must be recorded the same way more than once in different places and times. So that everything is the same, these values must always be the same.

2.2.3 Existence.

This is a crucial piece of information used in auditing.

Bovee et al. (2003) state that if we need to validate information, it should pass any testing to ensure there are no false or redundant entities, fields, or values.

2.3 Reliability of Records

A record's reliability and usability must be evaluated in addition to its legitimacy and integrity. These characteristics are critical when archivists decide how to preserve and appraise a collection. Usability refers to the extent to which future end users can view and interact with saved material in terms of accessing, displaying, and accurately interpreting the data in the long-term preservation of digital records (Mason, 2007). While the reliability of a record as a factual statement is reflected by its trustworthiness, the opposite is true. In determining a record's validity, it is necessary to examine its form and the degree of control exercised over its development (Roeder et al., 2008). Thus, the digital record must be accurate, factual, and reliable in any administrative or corporate context. (ISO 15489-1, 2001). For a record to be considered usable, it must be easy to locate, retrieve, and use. Preserving an electronic record necessitates knowing its

specific qualities, which can be found by identifying, authenticating, and extracting its essential metadata, according to the IRMT (2009). As they put it, "in what format was a digital product created and stored?" would be answered by the identification procedure. Is this a digital picture? Documents created with Microsoft Word 2000 must be checked to see if a copy exists in an MS Word 2007 or an MS Word 2000 document, for example.

2.4 Skills of individual

Individual users often have limited knowledge about appropriate archival tools or necessary techniques for management and preservation (Debra A. Bowen, 2018). The staff needs to know a lot about how to use ICT tools. According to Jain and Mnjama 2016, most archivists and records managers do not know enough about technology to deal with the challenges of ensuring digital records are kept for a long time. These (theoretical) recommendations by archives are implemented in two ways: on the level of policies and strategies and on the level of practical solutions where time for planning is limited. Insufficient financing is frequently stated as the primary reason for the lack of coordination between the two, as evidenced by the following report. Electronic documents in archive institutions have not received the attention they need because of a lack of funding. Policy and strategy are excellent, but unless they are implemented, they have little value (O'Shea, 1997). It necessitates substantial resources, compliant organizations, committed management, and suitably trained people. The implementation is likely the most challenging aspect of digital preservation to complete. Sometimes, archivists feel they are expected to provide answers and solutions to situations beyond their ability. When it comes to serving the public, archive institutions must comprehend and manage changes in their environments to adapt service delivery in the future while still meeting the mission of their organizations (O'Shea, 1997).

2.5 Cloud Storage

The advent of cloud computing has profoundly altered our surroundings. For this reason, many people are shifting their data to the cloud (Aldossary & Allen, 2016). As a result, cloud storage has become the new norm. Data saved in the cloud is vulnerable to several challenges. These concerns range from using virtual machines to share resources in the cloud to problems with the actual cloud storage service itself. For example, cloud-based data must be secure while maintaining its integrity and readily accessible. Furthermore, because the cloud service provider is untrustworthy in managing authentication and authorization, distributing cloud data among multiple users remains a problem. For a wide range of services, cloud computing offers a cost-effective and flexible solution via the World wide web (Nepal et al., 2011). An IT paradigm shift and a new computing model across pooled resources such as bandwidth, storage server, processing power, services, and applications are referred to as "Cloud Computing" or "Cloud Computing." This new approach has become quite popular and garnered much attention from academic and industrial researchers.

2.6 Application Factors

Obsolete computer hardware and software threaten the integrity of digital records unless careful measures are taken to ensure their usage over time. For the dependability and safety of digital records, a clear and consistent mechanism must be utilized to monitor the integrity of every digital item's content, context, and structure. As a result of hardware and software upgrades, digital preservation typically necessitates the transfer of data from one format or configuration to the next. Because of this, the cost of transferring data (refreshing) or building and maintaining data (emulation) to accommodate outmoded data can be prohibitive for some organizations (ICST, 2002; Lavoie & Dempsey, 2004; Navarrette, 2009). On the other hand, research on long-term digital preservation costs is littered with studies that fail to provide reliable and comprehensive data. Researchers imply that digital records are vulnerable to loss and destruction due to the fragility of the magnetic and optical media on which they are stored and the unexpected failure of the reading and writing equipment on which they are used (Sambo, Urhefe, and Ejitagha, 2017). The" 2011 Data Breach Investigations Report" reported that hacking and malware are the common causes of data breaches, with 50% hacking and 49% malware (Sultan Aldossary, 2016).

2.7 E- Records

Appraising records from electronic environments is not much different from appraising traditional records, but there are still some slight differences. So that electronic records can be kept for a long time, the archive has to look at their technical condition, context, authenticity, and value. The specialized circumstance of an electronic record and whether or not it has been changed or tampered with since it was made require archivists to learn new skills and find new ways to evaluate the record. According to the International Organization for Standardization (ISO) standard 15489: 2016, this section defines the fundamental concepts and principles governing the generation, capture, and administration of records. It serves as the foundation for a number of International Standards and Technical Reports that offer more guidance and education on the concepts, strategies, and practices for producing, capturing, and managing records.

2.8 Data Migration

It is possible to migrate data every two to three years, but it will demand a large financial commitment, continual human attention, and employee training (Siew Lin et al., 2003). The ability to analyze and recommend the best new formats, the time to design and evaluate migrating pilot projects, and the ability to form and refine migration processes are all necessary for a successful digital migration project. In addition, the file is vulnerable to corruption as it is being re-converted. Over time, formatting can change, and data can be lost. A weird depiction of a document with no way to recover the actual data might be caused by a machine, software, or human mistake. Ensuring the long-term preservation of electronic records requires the development of best practices and methods.

3 Methodology

The study has targeted a focus population and unit of analysis to complete this research. A target population is a group of individuals that are from the same group of the sample. The target population for this study is mainly focused on the individuals involved directly with the archival institution and the digitization of the archival materials. The only group chosen is because the individuals have the knowledge and experience relating to the archival materials that have been digitized and stored in the cloud. So they will most like to give the most accurate response and feedback regarding the research objectives.

Interviews took place in the National Archives of Malaysia. Purposive sampling is the right size if it is big enough to answer the research questions and do what the study set out to do. Saturation is reached when collecting more data wouldn't lead to discovering a new theoretical category that would help understand and explain the studied event. But over time, the word "saturation" has come to mean more and more. For example, Weller et al. (2019) suggest using saturation as salience because they found a direct link between salience and the frequency of an item, theme, or behavior in the studied population. In the "saturation" tool, the sample size can be 10 units if the research goal is to find out about the most popular ideas or a larger size if the goal is to find out about a wider range of ideas. The interview is conducted using phenomenological analysis. Interpretative phenomenological analysis (IPA) explores how participants make detailed sense of their personal and social worlds. The main currency for an IPA study is the meanings particular experiences, events, and states hold for participants (Smith & Osborn, 2009). According to Smith et al. (2009), it is advised to choose between 3 and 10 for studies based on interpretative phenomenological analysis, but indicate that the appropriate sample size depends on several factors specific to the study concerned, including the level of study for student work. There were 3 staff from 10 Archives Management Division in the Digital Archives section has been chosen.

4 The Conceptual Framework of Proposed Methodology

A record's integrity in digital preservation refers to its completeness, consistency, accuracy, and correctness without any alterations (Kiltz, Lang & Dittman, 2007). If the document is considered "full and uncorrupted " in all its essential qualities throughout its history, it might be regarded as an authentic record.". When procedures like prevention, tracking, and validating modifications to preserved objects are pursued, the integrity of a digital asset is established. If users use existing programs, programmers test their programs on a non-productive machine, control processes are audited, and the system managers and auditors have access to the system. Such a method will eventually ensure storage media's preservation, maintenance, and preservation (Bishop, 2004). Authenticating data on untrusted servers has arisen as a significant concern. When accessing data, these technologies ensure that the storage server isn't altering or misrepresenting the contents. On the other hand, archival storage necessitates assurances regarding the authenticity of data stored on storage servers (Ateniese et al., 2007). In 2007, the NSFC provided additional funding for the work on preserving the integrity

and authenticity of digital information, particularly born-digital material (Fund No. 70773088).

Figure 1: The Framework of the Data Integrity and Digital Preservation

The framework structure shows that factors affecting data integrity reliability are the application and individual factors. The application factor consists of the software related to the computer used in the archival institution and other places. Many kinds of applications available can be used to communicate and go through the operations of the archival institutions in preserving the digital archives that have a practical value to the community. To ensure that digital items' authenticity and useful access are maintained throughout time, digital preservation is a combination of regulations and workflows that need active maintenance of those objects (Baucom, 2019). Archivists constantly combat issues like obsolescence. Hardware is routinely replaced, software is abandoned, and file formats are upgraded frequently. Migration of older digital items into new formats is a typical tactic used to combat obsolescence. Long-term digital preservation is crucial in order to keep these digital assets accessible to future users. Accessing data tampered with or deleted is not enough to determine that data has been lost or altered—many servers in an archive store vast amounts of data that are rarely accessed (Baucom, 2019).

5 Findings

5.1 RO1: To Identify the Factors That Influence Digital Preservation

All respondents agreed that the skills of archivists affect digital preservation. They have a variety of backgrounds, and not necessarily those with IT knowledge will only handle the system due to IT background. The skills of archivists play the biggest role in this digital preservation as it involves policies or regulations. Indeed, as archivists, the input is from the archivists themselves. Using the Archives Management System (AMS), they will follow all the archivist's requested processes. Databases for IT are more about the use of systems. The role of archivists in digital preservation is a huge involvement in skills, input or policy requirements, and indeed need to be provided by archivists.

5.2 RO2: To Measure the Extent of Archival Knowledge or Archives in Digital Preservation

Whenever the archivists handle any project, they will have the training to transfer technology to technical staff and archivists. That is one of the ways they intend to get knowledge for the use of the system. Respondent 2 was sent to Australia to see how the problems or the implementation to the National Archives as references and to add some knowledge about the digital records. However, they are not used to developing new databases since they used the readymade database from the government. They are also aware of cloud storage usage even though they use it for personal purposes (e.g.: email and Google Drive).

5.3 RO3: To Identify the Data Integrity in Archival Institutions

Archivists got the authenticity from a legitimate source. It is from the department to the agency, and there must be a transfer list to determine the materials. Even in archives, they have to recheck the materials. For integrity, the National Archives cannot identify the document's authenticity when it is already preserved. AMS can identify the records or content of materials transferred from agencies that public offices are not composed or not corrupted. Therefore, only materials that are created in public office are accepted. However, testing the document's authenticity is not their task since it is already included in digital forensics bodies.

6 Conclusion

This paper was aimed at establishing factors that influence data integrity by digital preservation sustainability in the National Archives of Malaysia in, to propose a model of digital preservation. To conclude, archives or archivists do not have ultimate control over the authenticity of a document. An item-by-item inspection is impossible given the volume of documents produced each year and the widespread occurrence of inadequate archives. Authenticity, on the other hand, is remained unverifiable.

References

A.Prakash, et al. l (2015). Improving Cloud Security Using Multi-Level Encryption and

Adu, K. K. (2014). Framework for digital preservation of Electronic Government in Ghana (thesis).

Aldossary, S., & Allen, W. (2016). Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions. International Journal of Advanced Computer Science and Applications, 7(4). doi:10.14569/ijacsa.2016.070464

Aldossary, S., & Allen, W. (2016). Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions. International Journal of Advanced Computer Science and Applications, 7(4). doi:10.14569/ijacsa.2016.070464

Allen, William, et al. l (2016). Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions. Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions. 7(2016) 4, 2016 www.ijacsa.thesai.org

Ateniese G, Burns R, Curtmola R, Herring J, Kissner L, Peterson Z, et al. 2007, Provable data possession at untrusted stores, in "Proceedings of the 14th ACM Conference on Computer and Communications Security", ACM, New York, NY, USA, pp. 598– 609.

Ateniese G, Di Pietro R, Mancini LV, Tsudik G. 2008, Scalable and efficient provable data possession, in "Proceedings of the 4th International Conference on Security and Privacy in Communication Networks", ACM, New York, NY, USA, pp. 9:1– 9:10.

Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., & Song, D. (2007). Provable data possession at untrusted stores. Proceedings of the 14th ACM Conference on Computer and Communications Security - CCS '07, 598. https://doi.org/10.1145/1315245.1315318

Authentication. International Journal of Innovative Research in Information Security 2 (2015), 1-8, doi www.ijiris.com

Baucom, E. (2019). A Brief History of Digital Preservation. 18. 'GXP' Data Integrity Guidance and Definitions, MHRA, March 2018.

Bishop, M. 2004. Introduction to Computer Security, Addison: Wesley.

Bovee, M.W. 2004. Information quality: a conceptual framework and empirical validation. [Online]. Available WWW:http://www.bsad.uvm.edu/Research/FacPubs/details?author=265 (Accessed 9 December 2005).

Copeland, A.J. (2011), "Analysis of public library users' digital preservation practices", Journal of American Society for Information Science and Technology, Vol. 62 No. 7, pp. 1288-1300.

CSA, "The notorious nine cloud computing top threats in 2013," The Notorious Nine Cloud Computing Top Threats in2013.pdf.

Farrell, M & Morris, C. 2006. The Status of e-government in South Africa. ST Africa Conference, Pretoria, South Africa. http://researchspa.csir.co.za/DSPACE/bitstream(Accessed on 22 November, 2013).

Guo, W, Fang, Y, Pan, W, & Li, D (2016) "Archives as a trusted third party in maintaining and preserving digital records in the cloud environment", Records

Guo, W., Fang, Y., Pan, W., & Li, D. (2016). Archives as a trusted third party in maintaining and preserving digital records in the cloud environment. Records Management Journal, 26(2), 170-184. doi:10.1108/rmj-07-2015-0028

ICSTI/CODATA/ICSU. 2002. Seminar on preserving the record of science, 14-15 February 2002, UNESCO, Paris, France.

International Records Management Trust (IRMT). 2003. Electronic government and electronic records: E-records readiness and capacity building: An electronic discussion paper, 19. London: IRMT.

International Records Management Trust. 1999. Electronic records. London: IRMT.

International Records Management Trust. 2003. Summary of actions and strategies. An electronic government and electronic records: e-records readiness and capacity building. Available: http://irmt.org/evidence/ediscussion/ObjectivesandStrategies.pdf. (Accessed 04 April 2014).

International Records Management Trust. 2009. E-Records readiness assessment tool. http://irmt.org/documentsbuilding_integrityIRMT_project_proposal.pdf.

IRMT. 2011. Managing Records as Reliable Evidence for ICT/e-Government and Freedom of Information. White Paper for Senior Management. London: IRMT

Jain, P & Mnjama, N. 2016. Managing knowledge resources and records in modern organizations. Hershey: IGI Global.

Jayapandian N A , Md Zubair Rahman A M J. (2018). Secure Deduplication for Cloud

K. Hashizume, D. G. Rosado, E. Fernaʹndez-Medina, and E. B. Fernandez, "An analysis of security issues for cloud computing," Journal of Internet Services and Applications, vol. 4, no. 1, pp. 1–13, 2013.

Kamala, G. 2010. E-government and e-records: challenges and prospects for African records managers and archivists. ESARBICA Journal, 25: 146-163.

Kiltz, S., Lang, A & Dittman, J. 2007. Taxonomy for computer security incidents, in Cyber warfare and cyber terrorism, edited by LJ. Janczewski and AM. Clark. Information Science Reference (IGI Global).

Li, yiben, et all (2016). Intelligence cryptography approach for secure distributed big data storage in cloud computing. Journal homepage: Information Science. (2016), 13, doi:www,elsevier.com/locate/ins

Liu, J., & Du, P. (2009). Long-term preservation of digital information in China: Some problems and solutions. Program, 43(2), 175–186. https://doi.org/10.1108/00330330910954389

Management Journal, 26 (2),170-184, Retrieved from https://doi.org/10.1108/RMJ-07-2015 0028

Mason, S. 2007. Authentic digital records: laying the foundation for evidence. Information Management Journal, 5: 32-40. modern organizations. Hershey: IGI Global.

O'Shea, G. (1997.). Research Issues in Australian Approaches to Policy Development. 7.

Sinn, D, Kim,S, Syn, & S,Y. (2017) "Personal digital archiving: influencing factors and challenges to practices", Library Hi Tech, 35(2),222-239, Retrieved from https://doi.org/10.1108/LHT-09-2016-0103

Sneha Tripathi, (2018) "Digital preservation: some underlying issues for long-term preservation", Library Hi Tech News, 35(2),8-12, https://doi.org/10.1108/LHTN09-2017-0067

Storage Using Interactive Message-Locked Encryption with Convergent Encryption, To Reduce Storage Space. Brazilian archives of biology and technology An International Journal, 61(2018), 1-13, doi http://dx.doi.org/10.1590/1678-4324- 2018160609

W. A. Jansen, "Cloud hooks: Security and privacy issues in cloud computing," in System Sciences (HICSS), 2011 44th Hawaii International Conference on. IEEE, 2011, pp. 1–10.

Yu, Y., Au, M. Ho., Ateniese, G., Huang, X., Susilo, W., Dai, Y. & Min, G. (2017). Identity-based remote data integrity checking with perfect data privacy-preserving for cloud storage. IEEE Transactions on Information Forensics and Security, 12 (4), 767-778.