

DIGITAL PRESERVATION AND CONTENT MANAGEMENT IN THE CLOUD AGE: ISSUES AND CHALLENGES

Sreekala P.K., Jayasree V & Dr. M D Baby

Sreekala P.K. is working as Professional Assistant at Cochin University of Science and Technology.

Jayasree V is working a Professional Assistant at Cochin University of Science and Technology

ABSTRACT: Digital preservation is a formal venture to ensure that digital information of continuing value remains accessible and usable. It involves planning, resource allocation, and application of preservation methods and technologies, and it combines policies, strategies and actions to ensure access to digital content, regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time. Digital information is the lifeblood of institutions and organizations and the amount of digital content is increasing at a rapid rate. Managing all these data is becoming difficult and time consuming without an efficient and robust content management system. Content Management System and Digital Preservation do different but overlapping jobs. This paper highlights the need and importance of digital preservation and cloud computing based content management system for an effective way of enhancing scholarly communication. The paper also discusses about various issues and challenges of digital archiving in the cloud age.

Keywords: Digital Preservation, Content Management, Cloud Computing

1. Introduction

Now a days paper documents are being replaced by digital counter parts. Growing use of digital documents brings new users into the digital world and adds to the demand for simple means to store and share files. The challenges can be handled efficiently if we use a robust content management system. A content management system allows institutions and other organizations to create, edit, manage and publish any type of digital information such as text, images, video, sound, documents etc. A cloud computing based content management system is rapidly evolving the field and there is an increasing demand for storage of both born-digital archives and digitised material, and an expectation that public access to this content will continue to expand.

This paper intends to give an understanding of cloud storage and its potential contribution to their digital preservation and content management system. Planning and decision making to preserve and distribute valuable collections of any archives need to consider many factors. Cloud storage services proves to be a better solution for an integrated digital preservation environment consisting of systems, storage, policies, and people providing secure back up for preservation copies . Storage and band width needs depend upon the anticipated usage, security concerns and performance requirements.

2. Digital Preservation

Digital preservation is the active management of digital content over time to ensure ongoing access. It is the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value. This encompasses not just technical activities, but also all of the strategic and organizational considerations that relate to the survival and management of digital material. The goal of digital preservation is to maintain the ability to display, retrieve and use digital collections in the face of rapidly changing technological and organizational infrastructures and elements.

3. Need for digital preservation

The rapid growth of information has led to technological changes and the digital archiving has become a necessity. Digital preservation ensures continued access to and usability of digital information records over long period of time. Digital archive and preservation services are becoming more prevalent and a basic requirement beyond traditional libraries and content repositories.

Disaster recovery strategies and backup systems are not sufficient to ensure survival and access to authentic digital resources over time. Digitized preservation includes digitized analogue content and born digital content like text, videos, audio, e-mail, websites, research data, database, software etc. Rapid growth of digital objects that require archiving and data heterogeneity is another factor. Data continues to grow in to terabytes, petabytes and exabytes and data has a tendency to be lost.

The key features of Digital Preservation System are

- Digital preservation system is an effective system for securely moving content in to the archive making sure that they are unchanged and safe.
- Digital Preservation is a data management system which controls the digital objects and disposes them at the right moment.
- The administration role of digital preservation is the controlling of digital objects.
- The techniques of preservation always make sure that digital objects are usable and accessible whenever they are required.
- It also allows the objects to be found and downloaded to appropriate users.

4. Issues in digital preservation

Information held digitally is highly fragile. There is a chance for loss and destruction since they are stored in fragile magnetic and optical media that deteriorate rapidly. Digital media are subject to destruction. Even after we have the bits held somewhere safe and protected the problems still exists. Information technology is continually updating and the file formats are becoming unreadable by the software and IT systems of tomorrow surprisingly.

An “Active Preservation” system will be a best tool to recycle files in to formats that can be read today and to do so in a way that is automated and validated. This will preserve the full behaviour of the digital objects and not just their appearance.

5. Content Management system

Content Management Systems and Digital Preservation do different but overlapping jobs. Digital information is the lifeblood of all institutions and organizations in the cloud era. The process of managing and disseminating this vast content of information is a challenge and their effective control is a crucial part of organisations. The need for storage and retrieval of high volume of content under strict rules controlling the sharing of information among users and organizations can be attained through a proper content management system. The Content Management System can be defined as “a computer application that allows publishing, editing and modifying content, organizing, deleting as well as maintenance from a central interface”. The function and use of content management systems is to store and organize files, and provide version-controlled access to their data. The issues and challenges regarding the management of digital contents efficiently can be handled by a content management system namely Cloud based Content Management System.

6. What is Cloud?

The term *Cloud* means any computing provision provided by an external supplier over a network – this could be simple external hosting of an application, could be use of external services for a particular task, or it could be something rather more complicated which allows an organisation to grow and contract the resources it uses according to its needs. Cloud computing is the delivery of scalable IT resources over the Internet, as opposed to hosting and operating those resources locally, such as on a college or university network. Those resources can include applications and services, as well as the infrastructure on which they operate.

7. Cloud Digital Preservation Service

A Cloud based Content Management System is capable of handling a heterogeneous set of user devices in a distributed environment. Cloud computing is one of the most rapid growing area of Information Technology. A cloud service which provides digital preservation of information and data includes a comprehensive management and curation function that controls infrastructure, information, data and storage services. Cloud based digital archives and preservation services can offer significant advantages regarding ease of use, power, data centres, security and high availability. Rather than depending solely on internal infrastructure cloud storage helps to achieve cost saving also.

8. Types of Cloud Services

There are three major types of cloud services available

1. Software as a service (SAAS)

Applications or software is delivered as a service to the customer who can access the program from any online device. Some of these Web-based applications are free such as Hotmail, Google Apps, Skype, and many 2.0 applications, while most business-oriented SAAS, such as SalesForce, is leased on a subscription basis. There is usually little customization or control available with these applications. However, subscribers benefit from low initial costs, have access to (usually 24/7) support services, and needn't worry about hosting, installing, upgrading, or maintaining the software.

2. Platform as a service (PAAS)

With PAAS, a computing platform is provided which supplies tools and a development environment to help the institution build, test, and deploy Web-based applications. Customers don't need to invest in the infrastructure required for building Web and mobile applications but can rent the use of platforms such as Windows Azure, Google AppEngine, and Force.com. Applications which are built using these provider's services, however, are usually locked into that one platform.

3. Infrastructure as a Service (IAAS):

This type of cloud computing is also sometimes referred to as IAAS or Hardware as a Service and it involves both storage services and computing power. Amazon's Web Services, one of the major players in this area, offers two main products including the Elastic Compute Cloud (EC2), which provides computing resources, and Simple Storage Service (S3) for data storage.

9. Types of Cloud

According to the subscription there are different types of cloud

1. **Public Cloud:** Public cloud services enable the cloud infrastructure to be used by general public and they offer users the benefits of rapid scalability and low initial set up costs.
2. **Private Cloud:** A private cloud is owned and operated by an organisation, or a third party on behalf of an organisation.
3. **Community Cloud:** A community cloud is shared among two or more organizations that have similar cloud requirements.
4. **Hybrid Cloud:** Hybrid clouds are a composition of two or more distinct cloud infrastructures (private, community, or public) that although remaining distinct are interconnected to allow data transfer.

10. Issues and challenges

Digital preservation strategy utilizing cloud computing brings a range of legal questions like legal requirements in terms of management, preservation and access placed upon archives and their parent organizations, by their donors and funders via contracts and agreements or via legislation by Government (eg. Accessibility, availability, information security, retention, audit and compliance, Public Records Act etc); legal obligations relating to third party rights in or over the data to be stored (eg. Copyright, data protection) and legal elements of the relationship between an archive and a cloud service provider or providers.

Adoption of cloud based content management services presents a number of challenges. The staff will need to acquire skills to manage contracts and in particular to monitor services to ensure service level agreements. It is important to note that any failure of the service resulting from a problem with in house or other services is likely to be outside of the terms of the service level agreement. No service, whether cloud based or not can guarantee 100% availability. We should make assurances from our cloud partners about the

arrangements they have for the security of our data, both in terms of the physical and technical security and the level of network connectivity our cloud service provider has.

11. Conclusion

Digital preservation and content management in the cloud age requires a combination of technological and human factors, automatic processes managed and implemented by manual input and policy. The successful integration of right cloud service in our preservation environment will equip the library professionals with the tools to support and manage the preservation of digital collection. Cloud computing will play an important role in the future Internet services enabling on-demand provisioning of applications, platforms and computing infrastructures.

References:

1. Anderson S, “ Feet on the ground: a practical approach to the cloud nine things to consider when assessing cloud storage” A V Preserve 2014,<
<http://www.avpreserve.com/wp-content/uploads/2014/02/Assessingcloudstorage.pdf>>
2. Diptoneel Kayal, “ CcMS: A Cloud computing based content management system” International Journal of Information and Computation Technology 3,no. 4 (2013), 329-334.
3. David S. H. Rosenthal and Daniel L. Vargas, “Distributed Digital Preservation in the Cloud,<http://www.lockss.org/locksswp/wp-content/uploads/2013/01/IDCC2013.pdf>
4. <http://preservica.com/resource/content-management-and-digital-preservation-white-paper/>