Current Issues in Digital Asset Management and Preservation

Adi Alter, Digital Preservation Enthusiast and Rosetta Product Manager, Ex Libris

Ex Libris
a ProQuest Company

The New York Times, NASA, the Egyptian Scientific Institute, the Glasgow School of Art, and Eyebeam Studio - what do all of these institutions have in common? Each has experienced a loss of assets due to natural disaster, vandalism, or just poor preservation standards. In some cases, when assets were damaged it was possible to salvage them through restoration techniques. In other cases, the material, including irreplaceable historical documents, simply ceased to exist. In all of these instances, the use of digital asset management and preservation strategies would have made the recovery process much easier.

Digital asset management and preservation is a market with significant growth. Some projections state that market revenue will increase to over four billion dollars in 2019, at an annual growth rate of almost 29%, with the library market segment showing a similar trend.

Institutions that once saw these tools as nice-to-have now see them as a must, due to the exponential growth of digital data. It is estimated that the global volume of digital content - video, pictures, analog and digital research documents, websites, and financial records - will exceed 40 zettabytes (40 trillion gigabytes) by 2020 and 180 zettabytes in 2025, compared to 8 zettabytes in 2015.

Consequently, discussions about digital preservation that were once theoretical now have an immediate impact. Some of the major topics which are currently discussed include:

Data Sustainability

How to ensure long-term access to content in a world of constantly changing formats

Storage

How to wisely select and manage ever-increasing storage

Organization & Standards

Which systems, policies, and standards should be implemented within the organization, and how to ensure cooperation from the various stakeholders involved

Data Sustainability

Data Risks

There was a time when people thought that storing data on digital media such as a tape, flash drive, or even a CD was a permanent solution. But as technology has advanced, recognition has grown that storing content is not enough. Organizations must ensure that data will always be accessible.

Data formats have a finite lifespan: as new formats are introduced, old formats become obsolete and new applications can no longer render them. For example, content created by historical word processors cannot be opened by current word processors. As new technology is introduced into the market at a faster and faster pace, obsolete content on a significant scale is now a pressing problem.

In addition, managing content on a large scale requires that rich metadata be associated with that content, otherwise it is impossible to keep track of all the content managed by the digital repository. Missing metadata may ultimately make items inaccessible to most users.

The consequences of corrupted media, missing rendering applications, or missing metadata can be disastrous.

Complex Data

Commonly, data storage aims to preserve content such as documents, images, and audiovisual content. In such cases, one file usually represents a single data set. However, modern data can be much more complex. For example, storing an online newspaper requires preserving more than a single page – every page must integrate embedded images, hyperlinks, and audiovisual files as well as maintaining the appropriate association between these files.

Complex data is a major challenge for numerous entities. Various institutions such as cultural heritage organizations and research centers harvest snapshots of websites on a regular basis, every week or even every day. There are initiatives to preserve social data, such as tweets related to major elections, for future research. Other initiatives include preserving digital games together with their record-breaking game sessions. For all of these cases, the preservation of data is complex and challenging.

Current Solutions

In order to address the risk of formats' finite lifespan, digital asset management and preservation solutions manage and support digital content for the long term using two main strategies:

The **first strategy** is content migration, in which content is converted from an out-of-date format to a current format. For example, to ensure future accessibility to a Word 1.0 document, a migration process would convert it to PDF/A. Until recently this was a theoretical procedure, but now organizations are starting to implement content migration due to the need to manage massive amounts of legacy data that were generated using software that is now outdated.

The **second strategy** is emulation, which involves building a software emulator that enables running old content on new platforms. Emulation, although complex to develop and implement, has proven to be a highly effective measure and we expect to see increased use of emulation in the future.

Storage

As the digital universe grows and more digital data is generated, there is an increasing need for advanced storage solutions. Nowadays, it is not unusual for an institution to have millions of files and terabytes or even petabytes of data which need to be stored for the long term. In deciding upon a storage solution, institutions must consider many factors, such as:

- **Storage type** – which type of storage should be used – tape, flash, cloud, etc.
- **Location** – which content should be stored on the premises, which should be stored remotely or in the cloud, and how many replications of the different content types should be available
- **Scalability** – the complexity and costs of adding more storage capacity
- **Speed** – the required retrieval speed that is needed for content; for example, content that will be available online requires relatively fast retrieval
- **Reliability and data refresh** – the failure rate and lifespan of a storage solution and the difficulty of migrating content to a new storage solution

There are different parameters that influence storage decisions, such as content type, what the content will be used for, and who its users will be.

All of these factors inform the tradeoff between cost and efficiency. Since the selected storage solution will serve the organization for many years to come, and since storage decisions are expected to have high costs associated with them, organizations need to carefully examine the different parameters and choose a solution that handles this tradeoff in the best way. Typically, this will require using a hybrid storage solution – that is, using different storage types in different cases.

Organization and Standards

Many people are involved in the digital asset management and preservation process, including librarians, archivists, researchers, and the IT personnel who are in charge of enabling access to millions of documents. All of these stakeholders are prone to uncertainty about which preservation policies to use and how to standardize the process to ensure that these policies will actually be followed within the organization.

Policies

People across the institution must follow the same preservation policies to handle digital data consistently and correctly. For example, institutions need to create policies regarding what will happen to the hard disks of scholars who have retired, how donated collections should be handled, and which types of users can access what kind of data.

Similarly, there is a need to decide what to preserve and sometimes, even more importantly, what not to preserve. A balance must be struck between the resources required for preservation and the ramifications of potentially mishandling data with future value.

Finally, once policies are set, correct implementation must be checked against the initial policies.

Organizational Requirements

For legal reasons, national and state institutions such as libraries and museums need to preserve cultural heritage content which is stored on a variety of media. Universities and researchers are obligated to maintain and share research data, as research must be preserved for several years due to regulations and the funders' requirements. In addition, research must be reproducible, and so research data must be preserved in order to enable future replication. Finally, institutions such as libraries have become heavily involved in digitizing printed media such as old books, manuscripts, and historical maps, enabling access to their valuable content from virtually everywhere.

Standardization

There are many standards for managing elements of the digital preservation process, including ISO 16363, PREMIS, METS, and the Nestor Seal of Approval. One of the fundamental standards is the Open Archival Information System (OAIS), now known as ISO 14721, which describes characteristics of a digital preservation system, and which is currently in the process of review and renewal.

Organizations and vendors are struggling to understand, implement, and maintain the various standards. Each standard requires both human and technical resources and is very detailed, so the proper selection and use of standards can be a significant barrier.

The Future

Experts believe that the digital asset management and preservation industry will mature in the next five to ten years as the need for these services becomes more urgent. We are currently seeing the beginning of some interesting trends that promise to become industry practices.

Consolidation

Various entities within the organization, such as the IT, library, and administration departments, maintain their own digital asset management and preservation activities, and each tends to have different types of data, policies, and systems. But in the near future, organizations will have a central entity to establish a "digital repository backbone". This consolidated system will manage the entire digital content lifecycle and offer a unified solution for preserving and managing data. A centralized system will also be able to support different workflows, usage types, discovery needs, and access rights.

Standardization

In terms of organizational standards, there is no need to reinvent the wheel - existing standards such as OAIS are comprehensive and well-maintained. The challenge remains the application of these standards, and the coming years will see more efforts in the implementation of standards which will determine how institutions manage the full digital data lifecycle, from deposit to dissemination.

Commoditization

Currently, solutions are specialized, but in the future, digital preservation will be more of a commodity. While large institutions will remain the main users of preservation services, individuals will also want these services for their own personal data. It is also possible that digital preservation solutions will become "Preservation as a Service," even for personal use.

Conclusion and Recommendations

Digital asset management and preservation is a complicated field that has become increasingly essential for a variety of institutions. The number of policy choices that must be made regarding accessibility, storage, organization, and standards can be overwhelming. Similarly, technical considerations such as the many tools and data types involved in digital preservation are currently the domain of specialists.

Although each of the above-mentioned challenges can be addressed with specific products and services, the optimal solution is a single, universal platform. This ensures a uniform treatment of all essential variables based on industry standards and practical experience. As the digital asset management and preservation market grows, so too will the need for and sophistication of universal digital asset management and preservation platforms.

To choose the right platform for the organization, stakeholders should start the process through internal consultations with each other. It is likely that various parts of the organization are encountering similar challenges, and different departments should join forces to find the best, universally beneficial solution. This will also ensure better utilization of the selected product and the ability to share costs between various stakeholders. In addition, organizations should pick a solution that is easy to use, because easy-to-use technologies are the ones most likely to see consistent, widespread, and long-term application. Implementation of the system should occur regardless of the development level of institutional policies, which can be formulated more practically once the technology is in place. Waiting until all cross-institutional policies are in place may dramatically postpone the entire implementation and place the implementation itself at risk. It is therefore recommended to start with early adopters and a handful of workflows. Once the system is active and successful, the majority will eventually join the ride.

And finally, as various industry-leading organizations have painfully learned, it can never be too soon to implement a digital asset management and preservation solution. Choosing the right solution can save your digital assets from the next data disaster.

About Ex Libris

Ex Libris, a ProQuest company, is a leading global provider of cloud-based solutions for higher education. Offering solutions for the management and discovery of the full spectrum of library and scholarly materials, as well as mobile campus solutions driving student engagement and success, Ex Libris serves thousands of customers in 90 countries.
Visit http://www.exlibrisgroup.com

About Ex Libris Rosetta

Rosetta is a complete digital asset management and preservation solution that addresses the ever-growing need to collect, archive and preserve the digitally-born and digitized materials stored at academic institutions, cultural heritage organizations, and archives, ensuring data integrity and access over time. With an emphasis on workflow optimization and open architecture, and including an easy-to-use web-based user interface, Rosetta has already been adopted by dozens of institutions worldwide to manage their valuable digital assets.
http://www.exlibrisgroup.com/category/RosettaOverview

Ex Libris
a ProQuest Company