Digital Preservation **Handbook**

Digital Preservation Briefing

Who is it for?

Senior administrators (DigCurV Executive Lens), operational managers (DigCurV Manager Lens) and staff (DigCurV Practitioner Lens) within repositories, funding agencies, creators and publishers, anyone requiring an introduction to the subject.

Assumed level of knowledge

Novice.

Purpose

- To provide a strategic overview and senior management briefing, outlining the broad issues and the rationale for funding to be allocated to the tasks involved in preserving digital resources.
- To provide a synthesis of current thinking on digital preservation issues.
- To distinguish between the major categories of issues.
- To help clarify how various issues will impact on decisions at various stages of the life-cycle of digital materials.
- To provide a focus for further debate and discussion within organisations and with external audiences.

Gold sponsor

Silver sponsors

Bronze sponsors

Reusing this information

You may re-use this material in English (not including logos) with required acknowledgements free of charge in any format or medium. See How to use the Handbook for full details of licences and acknowledgements for re-use.

For permission for translation into other languages email: handbook@dpconline.org

Please use this form of citation for the Handbook: Digital Preservation Handbook, 2nd Edition, http://handbook.dpconline.org/, Digital Preservation Coalition © 2015.

Contents

Why Digital Preservation Matters

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

This section together with [Preservation issues](#) is designed as a briefing for those new to digital preservation. It is structured into four inter-linked sub-sections. In addition it has a close relationship to the [Getting started](#) section, which is also particularly designed with those new to digital preservation in mind.

Digital preservation: the challenge of a generation

Any digital object can be considered in scope for digital preservation: born digital or digitised, corporate or personal, innovative or routine. Digital preservation can encompass texts and images, databases and spreadsheets, vectors or rasters, programs and applications, desktop files and enterprise systems, email and social media, games, movies, music and sound, entire web domains and individual tweets. Digital collections can derive from laptops or desktops or smart phones; from tablets, souped-up servers or hulking great mainframes. They can be snapped at the end of a selfie stick or beamed from sensors deep in space; they can be generated by tills and cash machines, by satellites and scanners, by tiny sensitive chips and massive arrays. They can be stored in repositories or data centres or USB sticks. There is no digital object or system that is not provisionally within scope for digital preservation.

Pervasive, changing and ubiquitous, digital technologies are a defining feature of our age. Digital materials are a core commodity for industry, commerce and government. They are fundamental for research, the law and medicine. The creative industries, cultural heritage and the media depend on reliable access to digital materials while families and friends extend and sustain their relationships through digital interactions.

But digital materials - and the opportunities they create - are fragile even if they also have the capacity to be durable through replication. Digital platforms change and the long chains of interdependence on which they depend are complicated and fluid. Their longevity and utility is threatened where contents or contexts are lost: engagement and exploitation are enabled when digital materials endure. The greater the importance of digital materials, the greater the need for their preservation: digital

preservation protects investment, captures potential and transmits opportunities to future generations and our own.

Already we have made great strides in averting a "digital dark age". There are a growing number of repositories all over the world that can claim a long track record of keeping digital materials well over many decades (for example the UK Data Archive founded in 1967). This gives us a broad foundation of experience and collaborative professional networks to draw on.

It is a shared, generational challenge.

The always emerging digital preservation challenge

The unifying characteristic of digital materials is their machine-dependency. Information can only be accessed and functions can only be executed through a computer. As technology becomes more sophisticated this dependence becomes an ever more elaborate chain of inter-dependencies that are hard to track and tricky to maintain.

So long as the IT sector remains innovative in its provision of new tools and technologies, digital preservation managers will respond by devising effective strategies for ensuring the durability and usability of new digital materials, so digital preservation will remain an always-emerging challenge.

To ensure the value of digital materials in the long run we need to ensure access, which in turn means we need to understand and mitigate rapid changes in technology and organisations (see Preservation issues).

Digital material can often only be archived well in digital form: there is no non-digital equivalent such as paper that retains all the essential information and functionality it provides. Too often it has been necessary to print out digital material for archiving and then even re-digitizing the printed copy later because there has been no local capacity for managing born digital material.

Today we have a growing and effective body of approaches, experience, and collaboration to address the challenges. Digital preservation is an important, necessary and doable endeavour with simple first steps all can undertake (see Getting started).

What is in scope?

Simply because everything could be in scope for a digital preservation strategy does not mean that everything should be preserved.

The question is less what can be preserved so much as what should not be lost. Selection, appraisal and disposal are significant components in any digital management activity. In the context of an expanding digital universe, a determined effort to identify, process and retain digital material of enduring value means on one hand that the right material is available to the right people at the right time in the right format; and on the other hand material is identified that can be actively removed or benignly neglected.

Digital material provides profound new opportunities for access and use of repositories. If digital collections exist in a fast changing environment, then we should expect that our users do too. Users of digital materials are likely to be using technology that is not yet fully developed in ways that we cannot fully anticipate, in places we may never visit and for purposes that we may struggle to predict. So any meaningful answer to the question of 'how can we preserve digital materials' will rapidly resolve to 'what can we do to ensure that these digital materials can be used'? Preservation planning will only succeed when user needs are fulfilled.

All of this indicates a requirement that wherever possible the long term viability of digital materials should be defined early not late. Preservation action is needed at the start of the life of a digital object, not always at its end. Creation, management and archiving of digital materials are no longer at opposite ends of a process but are integrated all the way through. By extension, preservation is no longer simply a concern for memory institutions in the long term but for everyone interested in using and accessing digital materials.

Who needs to be involved?

The ability to preserve digital materials depends upon a wide range of stakeholders. Principal among these are the creators of digital content, whose involvement in their preservation might involve, for example, consideration of standards in terms of format and media, and ensuring enough contextual information is available to enable their management by others. Creators may often be unaware of their pivotal role. This could be for all kinds of reasons, but a vital part of any digital preservation effort is the effective dialogue with creators of digital materials to inform and advocate the value of their engagement (to them and others).

If the creators of digital materials have a responsibility to enable long term access, then this responsibility is borne even more fully by those who provide the infrastructure and environments in which they are created. In some cases this may be a corporate function, with the provision of corporate tools and services which are preservation ready. In other cases responsibility will be borne by external service providers who host digital infrastructure for clients.

The nature of digital technology dictates that it is not feasible simply to hand over stewardship of the resource at some point in the future, without having managed it sufficiently to facilitate sustainability.

In some cases, institutions will manage their own digital legacy: large institutions that create digital materials may most sensibly be the ones to manage them in the long term, thus maximising return on their initial investment. But in other contexts co-operative models for long-term preservation have emerged involving a number of organisations. Both subject specialist and expert centres have emerged offering specific preservation solutions for specific types of digital material.

For some organisations, it may prove more cost-effective to contract all or part of their digital preservation activities to a third party. Whilst it may be advantageous to outsource, it is important to remember responsibility remains with the organisation. Staff will need to be sufficiently aware of digital preservation issues, particularly as they relate to legal, organisational and contractual problems, to manage these third party contracts effectively.

Any institution which places value on digital resources in general needs to ensure the long-term preservation of digital materials. A significant number of institutions have not only taken that role on for themselves but have offered wider leadership in addressing the practical implications of digital preservation.

Ultimately however, digital preservation cannot be perceived as solely a concern for archives, libraries, museums and other memory institutions: it is a challenge for all who have an interest in creating, using, acquiring and making accessible, digital materials.

Resources

Why Digital Preservation is Important for Everyone

https://www.youtube.com/watch?v=qEmmeFFafUs&index=43&list=PLEA69BE43AA9F7E68

Short Library of Congress video produced in 2010 for the non-specialist audience explaining how traditional information sources such as books, photos and sculptures can easily survive for years, decades or even centuries but digital items are fragile and require special care to keep them useable. Rapid technological changes also affect digital preservation. As new technologies appear, older ones become obsolete, making it difficult to access older content. (2 mins 51 secs)

Preservation Issues

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Introduction

This section together with Why digital preservation matters, is designed as a briefing for those new to digital preservation. It is structured into three inter-linked sub-sections covering Threats to digital materials, Organisational issues, and Resourcing issues. It links to more detailed treatment in other sections of the Handbook as appropriate, but has a particularly close relationship to the Getting started section, which is also particularly designed with those new to digital preservation in mind.

Digital preservation can often seem daunting at first. It is important to realise than those with existing skills in either information management or information technology within organisations are well placed to build on and apply these skills to digital preservation activities. However, it may require initially learning some new unfamiliar terminology (see Glossary), extending skill sets, and sometimes working in new ways.

Threats to Digital Materials
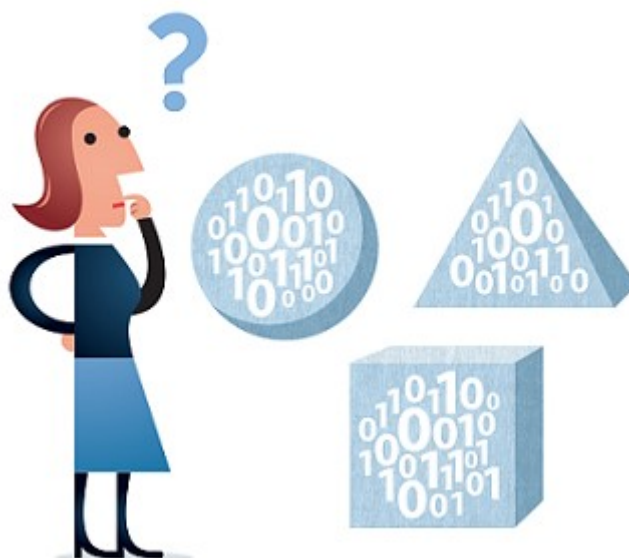
Keeping the data

Every digital file is formed from a series of zeros and ones, or bits (binary digits). These streams of bits need to be captured and retained over time, without loss or damage, to ensure the survival of digital materials. There are an array of threats to any attempt at preserving these bits. Storage media can decay over time, leading to corrupted files. Storage media may become obsolete and unsupported by contemporary computers and the software that understands and provides access to them. The bits may be ignored, abandoned, accidentally deleted or maliciously destroyed. Removable media could be left on a shelf and forgotten, files stored on a shared network drive might be left without an owner, or a third party cloud storage provider could go out of business.

Maintaining a systematic process for bit preservation remains a fundamental requirement in ensuring long term digital preservation. Storage media must be monitored and refreshed (See Legacy media). Redundancy must be introduced by replicating or backing up files, introducing diversity in dependent technologies and avoiding catastrophic disaster at a single geographical location (see Storage). Checksums must be generated and frequently recalculated to identify any loss and ensure that the integrity of the bits can be verified in an efficient and automated manner (see Fixity and checksums). The locations in which digital materials are stored should be carefully recorded, and responsibility for their preservation allocated.

Keeping the meaning of the data

Reconstructing the information that is encoded within a stream of a bits typically requires computer software that is designed to render, manipulate, analyse or otherwise interact with the particular encoding or format of the data. Over time, the encodings (or file formats) may change, and the software applications that interact with them may go in and out of favour. Although unusual for well known file formats, less well used file formats may become obsolete over time, as the software that renders them is no longer supported (see File formats and standards).

Understanding the technology on which particular digital materials are dependent enables appropriate action to be taken to ensure their preservation. A considered preservation planning process might result in the migration of digital files from format to format, the emulation of obsolete software, or the employment of alternative software applications to render the data (see Preservation

action). Each of the options presents its own advantages and disadvantages and these need to be evaluated carefully, possibly on a case by case basis (see Preservation planning).

While file format obsolescence has not emerged as the overwhelming danger that was previously perceived, challenging subtleties remain. It may be possible to find a method for rendering an old file format (perhaps by emulating some obsolete software), but how accurate is the rendering, is it legal to run the software, and how much will this complex effort cost the preserver and the user?

Maintaining trust in the data

Digital materials have the potential to remain fluid over time, being edited or altered with ease, being damaged by media failure, or decoded into human readable information in an unreliable or inaccurate manner by rendering software. For an end user to have trust in the result of digital preservation work it requires careful consideration of the entire lifecycle of the digital materials and who or what has interacted with them over time. Information management systems need to be able to link to essential contextual information regarding the business procedures of the creating agency. Authenticity and integrity of digital resources can be equally important in other sectors. For example, scholars will need to feel confident that references they cite will stay the same over time, courts of law will need to be assured that material can withstand legal evidential requirements, government departments may well have legally enforceable requirements regarding authenticity, and so on. This issue overlaps with both legal and organisational issues and it may be one which is best resolved within individual sectors rather than through generic procedures.

The application of data integrity techniques and the maintenance of audit trails can provide confidence that a digital object has remained unchanged (except by necessary preservation action) since deposit in an archive (see Fixity and checksums, and Information security). Ultimately its authenticity to a user may depend much more on the broader trustworthiness of the preserving organisation as a whole. Maintaining high quality preservation processes based on current best practice and validated by appropriate audit and certification will be crucial (see Audit and certification).

Keeping the context of the data and its dependencies

The meaning of digital information can be dependent on additional information that may have been implicit within the context it was originally created or used in, but less clear when revisited at a later date. Identifying, understanding and capturing relevant contextual information can be vital to a successful preservation effort. This might be as simple as capturing the units of measurement used within a spreadsheet, the scale of a map, or the point of origin within a CAD drawing. As digital information continues to be created in a more complex and interconnected manner, it may be necessary to retain the place of particular digital materials within a wider context of associated information resources. What may be seemingly simple and stand alone documents may actually depend on related files, referenced fonts and may have pointers to related information on the web. What might be viewed as a simple web page may have been generated on the fly from live data sourced from different locations on the Internet.

Understanding the data, how it will be used, its dependencies and its context will enable it to be captured for preservation in an appropriate manner and documented in a sufficiently explicit manner to enable the intellectual content to be retained and understood on into the future (see Metadata and documentation).

Acting in a timely manner

Prioritising digital preservation activities and applying them in a timely manner can be crucial not just in avoiding loss but in ensuring the best use of limited resources. Where the opportunity exists to intervene early in the lifecycle, digital materials can be shaped to survive better into the future. The choice of file format, the capture of critical documentation or the description of key relationships in the metadata may require a small investment up front, but could deliver considerable savings further down the line (see Creating digital materials). Where this is not possible, and risks to the data have been identified, the best timing for preservation action can be unclear. Early intervention to head off technological obsolescence may provide greater confidence of long term sustainability but with the risk that intervention may not ultimately be necessary and resources were wasted. Just in time action may minimise unnecessary activity, but increase the effort needed to research obsolete technology in a particular case requiring specialist knowledge that is no longer current. Appropriate action should be taken on a case by case basis.

Coping with the data deluge

Research reported by David Rosenthal noted that the rate of data creation is expanding by about 60% per annum; that developments in data storage allow are expanding at about 25% per annum; and that data centre budgets are expanding at about 2% per annum (Rosenthal, 2014). While this places challenging pressures on selection policies and other organisational decision making it also poses technological questions. Simple preservation processes that function effectively at one level will not necessarily scale easily to work with very large volumes of data or perhaps very large individual files. The technology and understanding to work at scale is moving forward rapidly, with growing expertise for handling large audio visual collections, research data and web based archives (see Content-specific preservation). But some repositories still face significant challenges in developing and maintaining scalable architectures and procedures to handle growing quantities of data. The technical and managerial challenges in accessioning, managing and providing access to digital materials on this scale should not be underestimated. It can be important to remember that selection, appraisal and disposal are significant components in any digital management activity.

Organisational Issues

While technological issues can be challenging, there are also numerous challenges which relate to organisational issues. These include how digital preservation is organised and delivered, or how those responsibilities change over both time and the lifecycle of digital materials. There are common digital preservation challenges faced across organisations, yet every organisational context will be different. It is vital to ascertain organisational drivers and tailor practical solutions to meet these needs. There is no one size fits all approach for digital preservation.

The creation, preservation and access for digital materials are widely distributed. As a result, there is an increasing need to go beyond the confines of individual organisations, or even countries, to maximise the benefits of the technology, address common issues, and to overcome the challenges cost-effectively.

In-house or outsource?

The decision whether to do all or part of digital preservation via a third-party or in-house, or perhaps a combination of the two, is often a complex one. Digital preservation may be undertaken in-house if there is sufficient staffing and infrastructure but outsourcing some activities or support can be cost-effective, and can leverage internal capabilities and capacity.

Outsourcing specific tasks or services from a repository is by no means a new phenomenon. Repositories have contracted out some of their operations for decades. Of critical importance is having and retaining sufficient knowledge to be able to prepare effective specifications and monitor performance. Outsourced work must be easily verified and quality checked, and this is best enabled via careful design of the specification, and the reporting providing by the 3rd party. Cost will clearly be a key consideration when deciding whether or not to contract out digital preservation but there are also other factors to consider such as legal issues. For example, legal provisions due to privacy or confidentiality may influence whether outsourcing is appropriate or not. The advantages and disadvantages of each option will need to be balanced in light of the individual organisation's mission and responsibilities (see Procurement and third party services and Cloud services).

Collaboration

There is a significant overlap in the digital preservation issues being faced by all organisations and across all sectors so it makes sense to pool expertise and experience. There are compelling reasons and, in some cases, political pressure, to engage in greater collaboration within and between organisations in order effectively to confront and overcome the challenges of digital preservation.

Most organisations readily acknowledge the benefits of increased collaboration but also indicate the potential difficulties that can arise in the form of differing agendas, timescales, or funding mechanisms. None the less, it is often possible to collaborate in specific areas or with different levels of intensity that moderate these potential difficulties. Some of the most high-profile and successful initiatives in digital preservation of recent times have been collaborative in nature (see Collaboration).

Organisational change

The modern digital world is a place of both rapid technological and organisational changes. Organisations re-organise internally, merge, or cease to operate with increasing frequency. Digital preservation is a long-term activity and the likelihood of it being affected by organisational change increases over time. This may affect a repository not only through changes to its parent organisation, but through changes to its major depositors and users, suppliers, or collaborators. Organisational change is therefore a major risk to be managed (see Risk and change management).

Organisational structures

The nature of the technology and dependencies in the preservation of digital materials are such that there are implications for organisational structures. Many of the activities converge, for example decisions about acquisition and preservation should sensibly be made at the same time. Organisational structures will need to cross boundaries in order to draw on the full range of skills and expertise required for digital materials. Assigning responsibility for preservation of digital materials acquired and/or created by an organisation will inevitably require involvement with personnel from different parts of the organisation working together. This can potentially present difficulties unless underpinned by a strong corporate vision which can be communicated to staff (see Collaboration, Advocacy, and Staff training and development).

Roles and responsibilities

There are some existing repositories which undertake responsibility for specific subject areas or specific formats. In the UK, for example, the UK Data Service undertakes responsibility for selected social science research data, while the British Library's National Sound Archive assumes responsibility for its collection of sound recordings. Each repository will need to consider its own collection policy and the broader landscape of collecting institutions and remits within which it sits.

The digital environment demands engagement with a large group of stakeholders. The lifecycle approach to digital preservation advocated in the Handbook has significant implications for the way organisations responsible for long-term preservation need to interact and collaborate with creators, publishers and other intermediaries, and each other.

Creators of digital materials need to be able to understand the implications of their actions in terms of the medium to long-term viability of the digital material they create. Whether it be a record created during the day-to-day business of the department, a digital copy of analogue collection material, or a "born digital" resource, guidance and support as well as an appropriate technical and organisational infrastructure will assist in facilitating greatly improved prospects for efficient management and preservation (see Creating digital materials).

Selection

The enormous quantity of information being produced digitally, its variable quality, and the resource constraints on those taking responsibility to preserve long-term access, makes selectivity inevitable if the objective is to preserve ongoing access.

In the digital environment non-selection for preservation may almost certainly mean loss of the item, even if it is subsequently considered to be worthwhile.

In cases where there may be multiple versions, decisions must be made in selecting which version is the best one for preservation, or whether more than one should be selected. Sampling dynamic resources as opposed to attempting to save each change, may be the only practical option but may have severe repercussions if the sampling is not undertaken within a well-defined framework and with due regard to the anticipated contemporary and future needs of the users.

Some consideration also needs to be given in the selection to the level of redundancy needed to ensure digital preservation. There needs to be a clear understanding of who will undertake that responsibility and for what period of time. Otherwise, even if several copies are stored in various repositories, all of those repositories might, for a variety of reasons, cease maintenance of the digital object at some point (see also Acquisition and appraisal).

Balancing security and access

There has always been a strong link between preservation and access. Repositories need to ensure that their digital materials are safe and secure, but most also provide access to a variety of users. Access by real users can provide a valuable steer to the design of preservation facilities, helping to avoid unnecessary actions but also validating and introducing a feedback cycle.

Many types of digital material selected for long-term preservation may contain confidential and sensitive information that must be protected to ensure they are not accessed by non-authorised users. In other cases there may be legal or regulatory obligations on the repository affecting access. There can be tensions between these two roles and a need to strike a balance between security and ease of access (see Access, and Information security).

Legal compliance

Legal issues are not simple in digital preservation. Multiple copies and derivative versions often exist of digital materials, and there may be associated software and metadata with them from different sources. Digital content is generated by a wider group of creators and incorporates more diverse formats and intellectual property rights (IPR) than applies in the analogue world. The law also often lags behind technological change and digital preservation needs. Some of the key legal issues that affect repositories in collecting, preserving, and providing access to digital materials are:

- Any legal requirements in terms of management, preservation, and access placed upon the repository and its parent organisation, by donors and funders via contracts and agreements or via legislation by Government (e.g. accessibility, availability, information security, retention, audit and compliance, Public Records, Legal Deposit, etc.);

- Those legal obligations relating to third party rights in, or over, the digital materials held by the repository (e.g. copyright, data protection); and

- The legal elements of any relationship between a repository and any third-party provider or providers (e.g. terms of service contracts and service level agreements).

For further guidance and resources to help address these issues and manage associated risks, see Legal compliance and Procurement and third party services sections respectively of the Handbook.

Resourcing Issues

Budgets and costs

The cost of digital preservation cannot be easily isolated from other organisational expenses, nor should it be. Digital preservation is essentially about preserving access over time and therefore the costs for all parts of the digital life cycle are relevant. In that context even the costs of creating digital materials are integral in so far as they may need to include cost elements which will ultimately facilitate their long-term preservation (see Creating digital materials).

The ability to employ and develop staff with appropriate skills is made more difficult by the speed of technological change and the range of skills needed. It is also limited by resource constraints on organisations which may well need to manage growing traditional collections and digital collections without additional resources.

Nonetheless the exercise of calculating costs, however complex, is a valuable and necessary task to establish cost-effective practises and a reliable business model. The cost of the labour required for digital preservation will be the most significant by far and includes not only dedicated experts but varying proportions of effort from many staff such as administration, management, IT support, legal advisers etc.

Other major issues to impact costs include organisational mission and goals, including the type and size of collections, the level of preservation committed to, the quantity and level of access required, and time frame proposed for action. These are discussed in detail in the section on Business cases, benefits, costs, and impact.

The relationship of costs and institutional strategies and activities such as Collaboration, Procurement and third party services, Legal compliance, Staff training and development, or Standards and best practice are also discussed in the relevant sections of the Handbook.

Staffing and skills

Digital preservation involves a range of skills and organisational roles. Typically digital preservation draws on a range of skills which are not normally found in combination. That means larger organisations will likely need to assemble multi-disciplinary teams while in smaller organisations it will be necessary to rely on a distributed team or sources of support.

There are three main issues to consider with respect to staffing and skills:

- Firstly, although there have been considerable improvements in recent years, digital preservation teaching often lags behind current best practice or is wholly theoretical within relevant information management programmes for new entrants into the profession. So individuals with practical skills and experience are in high demand and staff can be hard to recruit.

- Secondly, job descriptions can be hard to script, especially when agencies are effectively starting from scratch with a new role. To this end a number of research projects have attempted to describe generic skills needed for digital preservation, using as a basis the assumption that different skills are required at different levels of an organisation. Tools like the DigCurv Skills framework allied to the Digital Preservation Coalition's Vacancies section can be very useful when describing new roles. Larger organisations with multi-disciplinary teams may be able to recruit to roles that are 'digital' variants of existing professional categories such as archivist, librarian or records manager, but for most organisations new types of roles must be created.

- Finally, staff working in digital preservation frequently report the need to engage in active career development. Given the expectation that technology and the needs of users develop through time, so the staff involved in meeting these changing requirements will need to find ways to have their skills constantly refreshed, such as through specialist briefings and professional networking (see Staff training and development).

Facilities

Effective digital preservation requires some basic facilities or infrastructure, typically technological in nature, on which operational workflows and the processing of digital material can be based. While these may be rudimentary or at least small scale in nature when an organisation takes its first steps in digital preservation, ramping up operations to address large quantities of data will require considerable investment in the facilities required to support it.

Storage

With the typical requirement of replicating preserved data to avoid loss, storage hardware remains amongst the most important digital preservation facilities. Storage technology has changed rapidly over recent decades. Archives widely used media such as CDs or DVDs for long term storage, but the rapid developments in magnetic media have brought fast and reliable storage that has made handheld media redundant. Enterprise storage systems now provide large storage volumes at reasonable cost. While they have finite lifespans, typically of around 4-8 years, they are easy to monitor and then replace when they reach end of life (see Storage).

Organisations may also wish to consider cloud services to "rent" preservation infrastructure. The flexibility of the cloud allows relatively rapid and low-cost testing and piloting. Cloud services can provide easy, automated replication to multiple locations and access to professionally managed digital storage and integrity checking. Repositories can add access to dedicated tools, procedures, workflow and service agreements, providing a digital repository system tailored for digital preservation requirements via specialist vendors (see Cloud services).

Digital repository systems

Many of the core requirements for preserving digital materials are provided in an automated fashion by dedicated digital preservation systems, or trusted digital repositories. A repository application will uniquely identify each digital object placed within it. It will manage the storage of that object, identify its characteristics and help a repository manager to plan its preservation. It will also facilitate access to the object. While basic preservation can be provided on an ad hoc basis at a small scale, a dedicated repository application is essential to managing digital materials effectively over time. The OAIS model provides a high level model for the functions required by a repository (see Audit and certification for more information on certification of trusted digital repositories and Tools for repository systems and components).

High performance computing

Increasing volumes of data require not only more storage but also greater computational power. Characterising and assessing the technical characteristics of data, indexing data to enable search and access, integrity checking and a host of other tasks require considerable computational performance. Those dealing with these big data, be it research data or web archives have typically looked to high performance computing, and technologies such as Apache Hadoop running on clusters of commodity hardware to meet this need.

Digital preservation laboratory

A number of larger organisations have developed lab environments within which an array of old and new technology can be applied for the stabilisation or ripping of data from obsolete media, and has been championed by organisations working with personal digital collections. Specialist drives for reading magnetic media, robots for processing large numbers of optical disks and write blockers for allowing access to hard drives without changing the bits in the process, are just some of the equipment that could be useful here. Media recovery companies offer an alternative approach that may be preferable in high volume cases, albeit with less control of the process and the need to move media offsite (see Digital forensics).

Resources

How Toy Story 2 Almost Got Deleted: Stories From Pixar Animation: ENTV

https://www.youtube.com/watch?v=8dhp_20j0Ys

Entertaining and informative story of how 'Toy Story 2' was almost deleted from Pixar Animation's computers during the making of the film and how the film was saved by one mom's home computer (2 mins 26 secs)

References

Rosenthal, D., 2014. *Talk "Costs: Why Do We Care?", DSHR's Blog*, Tuesday November 18 2014. Available: http://blog.dshr.org/2014/11/talk-costs-why-do-we-care.html