Digital Preservation

Digital preservation is the active management of digital information over time to ensure its accessibility. Preservation of digital information is widely considered to require more constant and ongoing attention than preservation of other media. This constant input of effort, time, and money to handle rapid technological and organizational advance is considered a major stumbling block for preserving digital information. Indeed, while we are still able to read our written heritage from several thousand years ago, the digital information created merely a decade ago is in serious danger of being lost, creating a digital Dark Age.

Digital preservation is the set of processes and activities that ensure continued access to information and all kinds of records, scientific and cultural heritage existing in digital formats. This includes the preservation of materials resulting from digital reformatting, but particularly information that is born-digital and has no analog counterpart. In the language of digital imaging and electronic resources, preservation is no longer just the product of a program but an ongoing process. In this regard the way digital information is stored is important in ensuring its longevity. The long-term storage of digital information is assisted by the inclusion of preservation metadata.

Digital preservation is defined as: long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required for. Long-term is defined as "long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely"."Retrieval" means obtaining needed digital files from the long-term, error-free digital storage, without possibility of corrupting the continued error-free storage of the digital files. "Interpretation" means that the retrieved digital files, files that, for example, are of texts, charts, images or sounds, are decoded and transformed into usable representations. This is often interpreted as "rendering", i.e. making it available for a human to access. However, in many cases it will mean able to be processed by computational means.

Preservation strategies in academic and research libraries are not new concepts. However, with an increasing amount of digital content, organizations have to cope with a new set of preservation issues.

Digital preservation is in its infancy worldwide and presents some difficult technological issues. Since the creation of digital media, over 200 different storage mediums have been invented ranging from magnetic tape to CD-Rom. Each of these mediums present a variety of their own preservation issues and also require a diverse range of technology which in many cases is no longer manufactured. In addition to this, there are thousands of different formats in which data can be stored on each medium; and each type of storage format may also require a specific piece of software to interpret the data's meaning.

So what is a library to do in order to protect digital content? There are no clear standards in the area of digital preservation and with most institutions lacking resources already, how are they to tackle these issues?

The issues surrounding digital longevity into five general areas:

The Viewing Problem - All digital formats require computer technology to view them. By nature technology (software/hardware/formats) move at such a rapid pace that, odds are, they won't be around when you want to view your data. This is of course unless you're viewing data right after you "preserve it" in which case, it's not really preserved now is it.

The Scrambling Problem - Data is often compressed or "scrambled" to assist in its storage and or protect its intellectual content. These compression and encryption algorithms are often developed by private organizations that will one day cease to support them. If this happens you're stuck between a rock and a hard place. If you don't want to get into legal trouble you are no longer able to read your data; and if you go ahead and "do the unwrapping yourself" it's quite possible you're breaking copyright law.

The Inter-Relation Problem - Digital information is often linked to other items. This is much more evident in the digital world than the physical. If these links aren't maintained the information is either incomplete, incorrect, or just plain doesn't make any sense. Unfortunately, due to the diversity of digital linkages and the relatively recent identification of these issues, they're often overlooked. A simple example of this is links on web pages which have died, never to be resolved again.

The Custodial Problem - Who is the custodian of a digital document? Is it a librarian's job? What if someone changes the content without telling the custodian, after all digital content is dynamic and easily changed. So does the document's custodian have to undertake version control? And then is it really preservation?

The Translation Problem - If we need software to interpret data (due to formats etc), and software changes version to version will it be translated differently in subsequent versions? Even if the software claims it will sometimes it might lose formatting, a font? This is particularly dangerous where the changes are subtle or so small that no one notices them, or does it really matter at all?

Digital Preservation Management

'Digital Preservation' is critical to the proper management of digital objects. Active management and intervention ensures digital objects will be accessible and understandable. The biggest risks to the accessibility and use of digital objects are advances in hardware and software leading to the loss of their meaning. Why Does Digital Preservation Matter to YOU and YOUR business? Legal frameworks, National legal frameworks often require organizations to provide adequate records of business processes, communications and many other types of data for many years after their creation, accountability & protection from litigation. Recent legal cases have shown the importance of being able to search and recover archived emails quickly and in a legally admissible manner. Protecting the long term view/Access to digital data is critical to ensure business continuity and to support decision making with a long term view. For research in particular preserving data may be crucial for identifying long-term trends. The valuable intellectual assets of organizations are increasingly in digital form. This data represent both intellectual property and a considerable investment of time, effort and money. It would there-fore be foolish not to protect and preserve these assets adequately. Reuse Repositories of digital information and the tools to mine, analyze and repurpose them represent a society's intellectual capital. Effective and affordable digital preservation solutions are essential to transfer digital data into valuable assets for business.

Digital preservation is a set of activities required to make sure digital objects can be located, rendered, used and understood in the future. This can include managing the object names and locations, updating the storage media, documenting the content and tracking hardware and software changes to make sure objects can still be opened and understood.

1. "Digital preservation combines policies, strategies and actions to ensure access to reformatted and born 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."
2. "The act of maintaining information, in a correct and Independently Understandable form, over the Long Term."
3. "All activities concerning the maintenance and care for/curation of digital or electronic objects, in relation to both storage and access."

What does 'long-term' mean in the context of Digital Preservation?

1. "five years or more"
2. "a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future."

3. "Data should normally be preserved and accessible for not less than 10 years for any projects, and for projects of clinical or major social, environmental or heritage importance, the data should be retained for up to 20 years, and preferably permanently within a national collection, or as required by the funder's data policy."

What do we need to preserve?

Various aspects of the digital objects may be needed to be preserved.

- The lowest level of preservation requirements includes preservation of the bit stream, this does not however ensure understandability, readability or usefulness of the digital object. The biggest risk in terms of understandability is that the meaning (and even the names) associated with values in a dataset, although known to the data producers, is not available to the users; without this the data is essentially useless.
  Another aspect is that, even for users within the same sub-discipline, terminology drifts and meaning is lost; users in different (sub) disciplines will require even more help with the semantics of the data.
- A more complex approach may strive to preserve not only the 1s and 0s but also the meaning so that it remains readable and understandable. Such an approach requires the preservation of additional information (representation information, technical metadata etc.)
- Even more ambitious preservation approaches try to preserve understandable content in such a way that the provenance and source of the digital object also remains clear. Thus the users can have trust that the object is authentic, accurate,

Why should we care about Digital Preservation?

- **Storage media/data carrier problem**
  Digital objects are much more 'fragile' than traditional analogue documents such as books or other hard copy mediums. Digital objects are fragile because they require various layers of technological mediation before they can be heard, seen or understood by people. Digital objects are also much more venerable to physical damage. One scratch on CD-ROM containing 100 e-books can make the content inaccessible, whereas to damage 100 hard copy books by one scratching move is - fortunately - impossible. A flash memory stick can drop into glass of water or get magnetized, portable hard drive or laptop can slip from your hands and get irreparably damaged in a second.
  Digital objects require pro-active intervention to remain accessible. While you can put a book on a shelf and return to it in upwards of 100 years and still open it and see the content as it was intended by the author/publisher, the same approach of benign neglect to a digital object is almost a guarantee that it will be inaccessible in the future.
- **Hardware obsolescence**
  Even if you returned to the digital object in five years to find the disk is in perfect

condition and you have software that can open the file, but if that file is on a disc your computer doesn't have a drive for you will not be able to access it.

- **Software and format obsolescence problem**
  Alternatively the software or file format can become obsolete for a number of reasons. For example software upgrades may not support legacy files; the format take up is low and the industry does not produce compatible software; software which supports the format may be bought by a competitor and withdrawn form the market place. Without the intervention of digital preservation techniques the information contained will no longer be accessible.

Some of the benefits of Digital Preservation

- **Legal**
  National legal frameworks often require organisations to provide adequate records of business processes, communications and many other types of data for many years after their creation.

- **Accountability & protection from litigation**
  Recent legal cases have shown the importance of being able to search and recover archived emails quickly and in a legally admissible manner.

- **Protecting the long term view**
  Access to digital data is critical to ensure business continuity and to support decision making with a long term view. For research in particular preserving data may be crucial for identifying long-term trends.

- **Protecting investment**
  The valuable intellectual assets of organizations are increasingly in digital form.

  This data represents both intellectual property and a considerable investment of time, effort and money. It would therefore be foolish not to protect and preserve these assets adequately.

- **Reuse**
  Repositories of digital information and the tools to mine, analyze and re-purpose them represent a society's intellectual capital.
  Effective and affordable digital preservation solutions are essential to transfer digital data into valuable assets for business.