Digital Preservation, IAML Berkeley, August 2002

Difficulty in finding a speaker – thanks to Deborah Woodyard for assistance ⏩

[Summary of contents]

The aim of this presentation is that of a wake up call in relation to a problem that is perhaps not yet widely appreciated. ⏩

Just in case anybody is expecting it, I shall not be talking about techniques for preservation of the content of analogue media through digitization ⏩

A reminder of preservation issues in existing (analogue) media ⏩

A brief survey of carrier formats where change has been continual ⏩

A look at what digital actually means, and why context is so important ⏩

Obsolescence problems ⏩

Digital media degradation ⏩

Digital preservation and digital stores

⏩

Digital is forever? ⏩

I hope to demonstrate that this is rather a dangerous illusion: there is no simple and one time solution to the preservation of digital objects.

About the oldest information carriers remaining from earlier civilizations other than cave paintings are at most a few thousand years old (and it has to be said, lasting rather better than many more modern carriers!) Most of our modern media have no hope of remotely approaching this degree of longevity. Permanent retention of cultural information will always need active intervention whether in the analogue or the digital domain.

Recently I caught the end of 'Star Trek – the movie' on the TV. In it, a captured 300 year old Voyager probe was transmitting a signal back to earth, that nobody could either read or understand. This seemed a rather appropriate metaphor for this presentation.

B[lack] and stay black until [⏩] appears at Analogue audio

[A brief survey of preservation issues in existing media]

Before plunging into the digital arena, I'm going to spend a little time looking at existing analogue audio media.

We probably all know something about the fragility of acidic paper, and the risks involved with any remaining nitrate film stock. Analogue sound also has its fragile carriers which include so-called acetate discs, novelty discs on cellophane-laminated card and wax cylinders.

[Format variety]

Let's consider the variety of carrier formats used during the relatively brief history of audio recording (around 120 years).

⏩

Analogue audio ⏩

- cylinders (mostly wax) *fragile* ⏩
- coarsegroove discs (mostly shellac) ⏩
- metal or glass platters coated with a thin layer of typically cellulose nitrate – used for private recordings, archive broadcasts, etc. from the time before tape usage was common *very fragile* … [leads into next bullet without change]
- cellophane-laminated card or other material … even chocolate ⏩
- microgroove discs (vinyl) ⏩

- steel wire (and steel tape??) ⏩

- flexible plastic in strips, coated with magnetizable oxides of iron, chromium, etc. and latterly micro-particles of ferromagnetic metals. Tapes have been produced in various widths from 4mm up to 50mm (2 inches) in various track configurations and playback speeds. They have been stored on open spools, and in different types of cassette, as well as formats such as the 8 track endless cartridge.

The problems with shellac or vinyl discs, or magnetic tape are well-understood and we've learned quite a lot about their ageing characteristics and optimum storage requirements.

⏩

Digital audio

Digital media are of course much more recent and already include a further refined form of magnetic tape as ⏩

Digital Audio Tape (R-DAT) ⏩

and there is the ubiquitous compact disc or CD. ⏩

More recently extensions to that format such as the "enhanced CD" have appeared, containing audio material together with data and images related in some way to it, and intended to be played using a computer. The additional content is usually software-specific and might therefore become unusable in quite a short time. I find it rather ironic that some record companies are at one and the same time producing enhanced CDs needing a computer to access their full contents, while also producing copy-protected audio CDs with embedded encoding to prevent them from being played in a computer! ⏩

Betamax and VHS tape cassettes have sufficient capacity to be used for storage of digital audio, encoded into the video bandwidth by use of a PCM encoder. My own institution decided in the early 1980s, as PCM encoders became easily available, to use the Sony betamax video cassette for digital archiving on the basis that it was acknowledged to be technically superior to VHS. Unfortunately the VHS format won the battle for mass acceptance, and became the de facto standard.

We now have a relatively urgent conservation programme for transferring the (digital) contents of around 10,000 betamax cassettes to CD-R before the continued availability of equipment to play them becomes too great a technical risk. Archiving was transferred to VHS from the early 1990s until quite recently when CD-R became widely available. It seems to me quite likely though that the contents of over 10,000 VHS tapes will similarly need to transferred to another medium in probably not much more than ten or fifteen years time. ⏩

The minidisc was born, nearly died and then was born again once portable minidicsc recorders became widely available. It seems likely to survive for the immediate future because of its easy portability, though it could quickly give way to MP3. ⏩

Does anybody remember Philips' competitor to Sony's minidisc when that first appeared? The Digital Compact Cassette or DCC came, and went, in the twinkling of an eye. I think the NSA has a few examples of the format, and possibly one piece of equipment on which to play them ... but I'd hazard a guess that by now they might have degraded to the extent of not being fully usable anyway. ⏩

The digital carrier of the moment is DVD in audio and video flavours. I wonder though how long it will last. ⏩ A recent news item headed: '100Gb in your pocket' described an optical disc the same size as a CD or DVD that is capable of storing up to 100Gb of data, far outstripping the 0.65Gb capacity of a CD or the 9.4Gb of DVD. The extra data is crammed in simply by reducing the size of the mark that the laser makes on the disk to record each bit. A commercial product is expected by 2005 or 6.

It uses the same red laser technology as CDs and DVDs so there is backward compatibility for playback equipment. Before long I don't doubt we'll see a move towards use of violet or even ultra-violet lasers together with the above technique and multiple layer construction to squash in unbelievably huge amounts more data.

Whatever next? In John Boorman's film Zardoz, the entire memory of the community was stored in a single large crystal … maybe that would have the kind of physical longevity and capacity we need but there might be an issue of continued readability if encoding standards and/or mechanisms changed!

⏩ ⏩

As archivists we should always take note that certain formats use so-called lossy methods of signal compression in order to maximize the amount of data that can be stored. These include the minidisc, DCC, DVD and MP3. In bald terms, this means that part of the original data is deliberately discarded prior to writing the file to tape or disc. The processes, like the creation of JPEG images, use some very clever algorithms developed out of the psychology of perception, and the end result can often be difficult to distinguish from the original. Make no mistake though … it is <u>not the same</u> as the original and the datastream contains no means for reconstructing a perfect copy of the original. ⏩

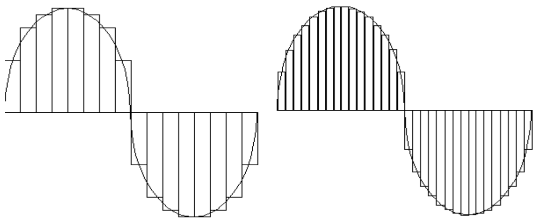
Only lossless compression – or no compression at all – should ever be used in the archiving of digital originals.

⏩

So what do we actually mean by digital anyway?

My own impromptu definition is that it is the representation of complex (analogue) data (whatever form that may take) by encoding in any two-state or binary system.

⏩



Analogue audio signals are digitized by sampling the level of the waveform at regular intervals and recording the result as a digital value. You'll notice in this diagram of a simple waveform that the blocks show regular samples taken of the level of the wave. On the first wave the outline would be very rough indeed when reproduced. For the second wave by simply doubling the sampling rate we already have a considerable improvement in smoothness. If we use a sufficiently high sampling rate and resolution (specified by the number of bits needed to register it) it is theoretically possible to reconstruct a more or less perfect copy of any analogue original, though there is an inevitable trade-off between improved accuracy and larger file sizes and transfer rates.

⏩

Taking the well-known 'red book' CD as an example, ⏩ it uses 16 bit sampling to register 65,536 different amplitude levels of the audio waveform at a rate of 44.1 kHz (44,100 samples taken every second).

This generates data at a rate of 705,600 bits every second for each channel, so that's a little under 606 Mb an hour for stereo. ⏩

The DVD Audio format has a much higher specification and registers samples at 24 bit resolution and a sampling frequency of 192kHz – that's 4,608,000 bits per second per channel and a little under 4Gb an hour for stereo. In each case we're looking at the raw data rate as generated whereas what needs to be recorded is rather more, as we shall see shortly.

⏩

Data packets, redundancy and error checking ⏩

A continuous stream of bits (binary digits) is conceptually grouped into 'words' of 8, 16, 24, 32, 64 or even 128 bits. Note that I say conceptually – there are no physical markers in a data stream so all digital devices count their way through. If something cause just a one bit loss, the entire data set effectively shifts by one bit and become instantly meaningless. Look at these examples of a bitstream … ⏩

00010100100101000001010011010100110011101

… I've marked in orange the start and end of the part of the bitstream that reads … ⏩ IAML in standard ASCII characters.

Assume that something has caused a counting problem and shifted the whole stream leftward by just one bit. ⏩ The bitstream is still the same except that my orange markers have moved leftward by one place …

00010100100101000001010011010100110011101

… those four bytes do still produce readable characters in this case but they now read as … ⏩

§†¶¶ ⏩

To circumvent such problems, error checking and data redundancy are built into all digital data streams, to enable the detection of errors and the reconstruction of data in the event of loss or corruption of one or more bits. A CD devotes around half of its capacity to error checking and data redundancy. Critical data streams needing very high levels of robustness have higher levels of redundancy and more complex error checking, and consequently have even higher storage requirements. (When the compact disc was first introduced it was said that its built-in error checking could cope with a small hole drilled through the playing area and reconstruct the missing data. While that was a case of overselling, it is true that radial scratches are capable of being ignored, though tangential scratches can make a disc partially unusable – or even totally unusable if the table of contents sector has been damaged.)

⏩

[The issue of context]

Here are four samples extracted from bitstreams. Can anyone spot the difference between them? ⏩

- 11010011001010100001101010111110001 ⏩
- 11010011001010100001101010111110001 ⏩
- 11010011001010100001101010111110001 ⏩
- 11010011001010100001101010111110001 ⏩
- – text (ASCII, RTF, Word…) ⏩
- – graphic (BMP, GIF, TIFF, PNG) ⏩
- – sound (WAV, RDF, AIFF) ⏩
- – moving image (QTM, MPG)

Yes, this was a trick question – they're all exactly the same. I hope this shows in a graphic way that in the digital domain the context is at least as important as the data itself ... neither has any meaning without the other. Each of these bitstream samples lifted out of its context is absolutely identical with all of the others even though they each represent parts of very different digital objects. To prevent a digital bitstream (including error correction and redundancy) from degenerating into nothing more than a random string of ones and zeroes, a digital preservation process must specify the file's context – in other words, the hardware and software used to create and derive meaningful output from it.

This is something quite outside our experience in the world of tangible objects. Even a small fragment of papyrus can have intrinsic meaning and can often be related to other fragments to enhance their collective meaning. A digital fragment divested of its context is in stark contrast completely meaningless. Digital Preservation encompasses therefore not only the preservation of a digital data stream, but of all that is necessary to extract meaning from it.

⏩

[*Hardware obsolescence*]

Hardware in general changes at an ever-increasing pace. In terms of the familiar area of personal computers, we know how quickly they go out of date – usually the week after one has bought a new one! The providers of hardware and associated software of course have a commercial interest in rapid but managed obsolescence. Backward compatibility is provided to some extent but is certainly not forever.

We already need to consider relatively recent computer equipment as 'historic' hardware. Machines such as ⏩

the BBC micro (made by Acorn), ⏩

the Acorn Archimedes, ⏩

the Amiga, ⏩

the NeXT box have all come and gone and each had specific software written perhaps solely for that platform. 'Historic' in this context actually means well within the last twenty years. ⏩

Over the same period IBM-compatible PCs ⏩

and Apple Macs have also developed rapidly. ⏩

The BBC Microcomputer is a particularly apposite example in this context because of its use in the BBC Domesday project, a technically complex project completed in 1986 for the 900th anniversary of the original Domesday Book. For those who've never heard of the Domesday Book I should explain that it was compiled in 1086 at the instigation of King William I of England (known as William the Conqueror) as a complete inventory of land ownership in the 20th year of his reign. It is now held at the Public Record Office in Kew near London.

For the BBC project, schools across the length and breadth of the United Kingdom were organized to collect data and images relating to their local areas and communities to enable the creation of a modern equivalent in multimedia form. The BBC Domesday project was designed to run on the BBC Micro, which was at that time by far the most widely available computer in UK schools. The large quantity of data collected necessitated the use of two laser discs as storage media. The discs were read by a specially-modified Philips laserdisc player that could interface with the BBC Micro in order, for instance, to overlay maps with data. [By the way, when did any of you last use or even see a laserdisc player? – another moribund if not completely dead carrier from quite recent times!] The cost of a complete BBC Domesday system was around £2,500 ($3,750) which was unfortunately too high for many schools and libraries to afford and the take up was not large. The NSA has two or three sets of the discs in its stock, but no working means to play them – there is apparently a working system at the Science Museum in London. The loss of capability for using this data happened in little more than ten years due to technology change.

⏩

[*Changing digital media formats*]

Looking briefly at changing media formats, who remembers and still has equipment that could read ⏩ the 8" (200mm) diskette, or floppy disk? – that was common up to the mid 1980s (less than twenty years ago) ⏩

Who has equipment that could read the 5.25" (130mm) floppy disc? – it was common up to the mid 1990s (less than ten years ago) ⏩

How long do you think the 3.5" (90mm) floppy disk will be with us? PC users might still have a use for them, but Apple stopped putting floppy drives into its machines with the appearance of the first iMacs about five years ago. ⏩

There are many other more specialized formats: Zip disks, Jaz disks (which I read recently are already heading towards oblivion), portable hard drives with SCSI or Firewire connections, formatted for specific computer file systems and therefore not readable by other file systems ... ⏩

There are also those same DAT tapes that can be used for digital audio, and high capacity formats such as DLT cartridges. These have huge data capacities but very uncertain long-term archival reliability.

⏩

[*Software obsolescence*]

Computer software is mostly under continuous development and goes out of date even more rapidly than the hardware on which it runs. Most companies recognize a need for backward compatibility, but there always comes a point where sustaining this without compromising further development of the product simply becomes uneconomic and users are forced to change or stagnate. If you've ever tried to extract data from obsolete but not particularly old file formats you will certainly understand this problem at first hand. ⏩

IBM-compatible PCs, started with DOS, moved through Windows 3.1, and then rushed on through Windows 95 and 98, NT, 2000, Me and now XP, in less than seven years. ⏩

Meanwhile Apple Macs, have moved through System 6, OS 7, 8 and 9 to the radically new OS X built on Unix. Using Unix is a potential benefit because it has been a more stable operating system over the same period than any of the others, though there are of course different 'flavours'. The Linux operating system – effectively an open-source form of Unix – is becoming ever more widely used. It has the advantage of being cross-platform and maintained and developed by the Open Source movement, which is not driven by the need to innovate simply for the sake of innovation or profit. ⏩

During this time periodic changes to connector specifications have taken place. These are hardware changes that one might expect only to cause problems in connecting older peripherals – though that might be crucial if they happen to be storage media. The changes though can have unexpected effects on software usability. I have some specialized software that I cannot run unless I buy a very expensive upgrade, even though what I have would actually run under the current operating system. The problem lies in the fact that the software will not run unless it sees a dongle – a hardware protection key – connected through a now obsolete kind of port. The dongle is not recognized by the software even when accessed through a fully-functional hardware port adapter. Copy-protection strategies such as this are clearly essential for small producers of high value software, but they can add to the headaches for anyone attempting to preserve working copies for archival purposes.

⏩

[Degradation of storage media]

Let's look at some of the physical degradations that can occur to media used in the digital domain – for the tape media most of the points apply equally in the analogue domain. ⏩

Magnetic media can suffer from: ⏩

- drop-outs – momentary loss of signal due either to inconsistencies in the magnetic coating or to spot demagnetization from external causes. If of short duration they might go almost unnoticed in analogue use, but digital signals will have to rely heavily on error-correction algorithms to supply the missing signal. Longer drop-outs might well be beyond the capacity of the error-correction algorithm. ⏩
- substrate problems include long-term loss of plasticizer with consequent loss of flexibility and possibly even cracking of tape; ⏩
- binder problems include sticky tape syndrome where layers of tape become stuck to one another and lose the magnetized layer in the process. In the opposite state the binder can lose its adhesiveness allowing the tape to shed signal-carrying oxide or metal particles whenever the tape passes over tape heads and guides; ⏩
- old tape edits made by cutting and sticking together with splicing tape can suffer from loss of adhesion – tape breaks or adhesive creep – sticky tape. Tapes used for digital storage are never edited in this way, but it is a very real problem for early analogue tape masters, especially those for the first electronic or musique concrète scores.

Floppy disks, Zip disks and Jaz disks are made of essentially the same material as tapes, so they are to some extent susceptible to similar problems in the longer term. ⏩

Current optical media (all of them digital except for the "old" 30 cm laserdisc) can suffer from problems of: ⏩

- physical abrasion of the playing surface leading possibly to irrecoverable errors. ⏩
- oxidation of the aluminium reflectant layer as a result of some kind of breach of the upper lacquer layer. This can result from physical damage or chemical attack and can make a disc completely unplayable. The risk is raised whenever discs are stored at high levels of humidity or are taken across steep and extreme temperature gradients. Gold is not subject to oxidation and it is generally accepted at present that archival copies should be made only to gold CDRs from reputable manufacturers; trusting to cheap discs is false economy in these circumstances.

Some early CDs are already unplayable: in some cases because of unthinkingly slipshod fabrication techniques, and in others because of chemical interaction with inserts that were printed on acidic paper or used acidic ink.

⏩

[*Born-digital*]

Born-digital materials include data files representing any or all of audio, graphics, video and text, possibly in complex interaction with one another. In most respects there is no effective difference from the preservation point of view between born-digital and digitized analogue data. ⏩

The one real difference is that we usually still have access to the original for digitized analogue materials, though it might of course be in an increasingly fragile state. It has been suggested that digitized surrogates have a symbiotic relationship with their analogue or graphic originals. The original can be preserved untouched but remains as the ultimate backup while the digital object enables uninhibited access. ⏩

In the case of born-digital materials, what we are actually preserving are the original files, or digital clones (which should be bitwise-identical). If these become unreadable there is no fallback alternative. Because of the sometimes complex nature of born-digital packages they will be particularly reliant on preservation of the full hardware/software context.

⏩

[Digital preservation]

Digital preservation has been defined by Jones and Beagrie in the Digital Preservation Coalition handbook as "the series of managed activities necessary to ensure continued access to digital materials for as long as necessary … beyond the limits of media failure and technological change." This involves "the need actively to manage the resource at each stage of its life-cycle and to recognise the inter-dependencies between each stage and commence preservation activities *as early as practicable*. This represents a major difference with most traditional preservation, where management is largely passive until detailed conservation work is required, typically, many years after creation and rarely, if ever, involving the creator."

⏩

Several strategies have been proposed for preservation of digital data: ⏩

- Technology preservation involves preserving a working collection of obsolete hardware and software to enable playback whenever required. It's not difficult to see that this is an almost impossible strategy to maintain in the long term. Apart from the sheer variety of equipment that would need to be preserved, at some stage all equipment will inevitably fail. Modern equipment, based as it is on integrated circuits that at some stage will no longer be manufactured, will ultimately be reduced, at best, to a partially-functioning cultural artefact. ⏩

- Refreshing of data by copying it to other media works well as long as the means to interpret the bitstream still exists. However, unless careful verification is done, it is quite possible for the process itself to introduce and replicate fatal data errors that might only be discovered many years later. ⏩

- Data migration involves transfer to a current file format possibly under a different operating system. The term would be appropriate for transfer of data from one database to another. As with data refreshing, careful verification needs to be done to ensure that data integrity is preserved. ⏩

- Emulation is probably one of the more promising ways forward but is potentially very expensive and certainly not a once only strategy. It involves writing a complex software environment to run on current hardware that enables preserved software to behave exactly as it had on the original hardware and operating system combination. It's not difficult to see that the hardware/software combination for which an emulator has been written will itself get out of date in a relatively short time so research is examining ways to make the process modular. This should obviate the need either to rewrite an emulator for each new generation of hardware and software, or to run multiple layers of emulation. As part of a research project known as CAMiLEON, based jointly at the Universities of Michigan and Leeds, a freelance programmer is working on reverse-engineering the data files stored on the BBC Domesday videodiscs in order to enable the data to be displayed on a 'modern' system. This is effectively a combination of both emulation and data archaeology. The CAMiLEON project is also investigating other combinations such as emulation with migration. However, as Holdsworth and Wheatley of the CAMiLEON project say: "If you really want to preserve things for which emulation is the most meaningful method, then it is important to get in early and at least achieve a proof of concept before relevant information (including human memory) is lost." ⏩

- In a 1995 *Scientific American* article Jeff Rothenberg proposed that original data files should be preserved in their original state but encapsulated in a virtual envelope that would contain all the information necessary for retrieval and use of the data through an appropriate hardware and software emulator. Alternatively, an envelope might simply contain instructions for the construction of such an emulator. ⏩

- Data archaeology is literally the last-ditch option. It is the use of sophisticated techniques to attempt to recover meaning from data for which the original context has already been lost. As multimedia formats and software become increasingly complex this approach will become more and more difficult and prohibitively expensive for all but the most important data. ⏩

- Wherever possible, open standards or those licensed across multiple platforms should be adopted in preference to formats that are proprietary to a single manufacturer. Linux and the Darwin flavour of Unix on which Apple's Mac OS X sits are both examples of open standards in contrast with Microsoft's closed and proprietary operating systems. The 'red book' CD format is an example of a proprietary but widely-licensed format.

Over a period of time it seems likely that an archive might need to have recourse to more than one of these strategies, though having to resort to data archaeology after any of the others would effectively be an admission of earlier failure.

⏩

- Some digital content such as print or straightforward graphic files can simply be printed or output to microfilm since paper and microfilm both have well understood ageing characteristics and storage requirements. ⏩

- Another option for text is to reduce richly-formatted data to a basic standard such as plain ASCII text. If XML or SGML markup can be done prior to reduction to ASCII there might not necessarily be any loss of layout and style. ⏩

For audio materials the equivalent to this last approach would be to save to analogue format. However, if we wanted to preserve a recording that had been created at the DVD-Audio standard of 24 bits 192 kHz it would violate a fundamental archiving principle to regard an analogue copy as a preservation master because it would actually contain less information than the original. In any case no analogue medium is without its preservation problems and equipment for replaying analogue originals is also becoming less commonly available. For the preservation of digital audio there is no option but to stay within the digital domain and to manage the digital preservation process effectively. ⏩

- In terms of media longevity there are interesting developments from Norsam, a Los Alamos-based company that in collaboration with IBM has developed a High Density Read-Only Memory or HD-ROM that uses ion beams to engrave up to 196 Gb of data on hard corrosion-resistant metal discs the size of a CD or DVD. The company can also use ion beams to engrave up to 196,000 pages of human-readable text onto a piece of nickel only 50mm square and 6mm thick (2″ x 2″ x 1/4″). It is not beyond the bounds of possibility therefore that some kind of long-life digital storage medium might become available that could contain human-readable instructions on how to decode a recorded digital data stream. [In complete contrast though you might appreciate the thought that Hansard, the record of British parliamentary business apparently spends approximately £30,000 ($45,000) per annum on producing one archival copy on ... vellum!]

⏩

[Digital stores]

Digital stores were initially conceived more from the access than the preservation point of view. They provide a consistent managed environment in which to hold and preserve digital data together with such ancillary information as will be needed to maintain its integrity and use over time. ⏩

Metadata associated with a digital store needs to record ⏩

not only the more traditional kind of data that we would expect to find in a catalogue record, but also ⏩

preservation data that describes the origins and characteristics of the data file. This includes the file format, ⏩

the hardware and software needed to derive meaning from the data, ⏩

dates of origination, copying, etc. ⏩

The system should also record data about the mean error-free life of each of the media types in use and about the viability of file formats in use, based always on conservative estimates. With all of this information the system should be able to issue timely prompts for migration of data to other formats or file systems whenever any of these factors reaches a predefined level of risk.

⏩

[*Trusted Digital Repositories*]

This is a very specialized field and the concept has arisen recently of a Trusted Digital Repository. Such a repository has been defined as "… one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future." (RLG-OCLC report entitled 'Trusted digital repositories: attributes and responsibilities', May 2002 .)

⏩

[*Conclusion*]

In conclusion, I can do no better than quote from a paper called "The archive as an ecosystem" by Julia Martin and David Coleman published in the April 2002 issue of the Journal of Electronic Publishing. "As technology continues to develop apace, those seeking to preserve and to update are faced with ever-expanding options of how to do this. It is an exercise in futility to try to pick the technological winner, for 'the latest' will be hopelessly outdated in ten years' time. What is necessary is a permanent strategy for handling perpetual change. This will involve periodic reassessment, the migration of material from one platform to the next, and the allocation of resources. Conceptualizing the archive as an ecosystem may give some succor to those involved with the complexities of the task."

⏩

[*Links and references*]

http://www.rlg.org/longterm/oais.html

The Reference Model for an Open Archival Information System (OAIS) is a conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term. RLG and OCLC and several groundbreaking international projects have played key roles in shaping the reference model and in adapting it for use in libraries, archives and research repositories, which represent only a section of those with interest in the OAIS reference model.

http://www.nla.gov.au/padi/

The National Library of Australia's PADI site is a subject gateway to digital preservation resources. PADI is supported by the Digital Preservation Coalition of which the BL is a member.

http://www.jisc.ac.uk/dner/preservation/workbook/

The Digital Preservation Coalition handbook is available on the web and as PDF, or can be ordered in hard copy form.

http://www.atsf.co.uk/dottext/domesday.html

This is a page describing the aims, technical achievements and subsequent demise of the BBC Domesday Project

http://www.si.umich.edu/CAMILEON/index.htm

The project has three main objectives:

1.) To explore the options for long-term retention of the original functionality and 'look and feel' of digital objects.

2.) To investigate technology emulation as a strategy for long-term preservation and access to digital objects.

3.) To consider where and how emulation fits into a suite of digital preservation strategies.

http://129.11.152.25/CAMiLEON/dh/ep5.html

Emulation (Holdsworth and Wheatley)