



Information Overload

Issue 19, March 2004

Welcome:

This month's issue of Information Overload is a little different from previous editions. For starters it's considerably longer, but that is because we have chosen the subject of electronic archiving as the topic under discussion.

As this subject is likely to generate a lot of comment, we will endeavour to publish some or all of the discussion threads in later editions.

We hope you enjoy reading.

Electronic Archiving: A Discussion

Lorraine Bradshaw
Marketing and Training Coordinator
Information Enterprises Australia

"It is a fundamental tenet of our democratic society that evidence in the form of records, be created, kept, preserved and be accessible into the future. With the growing diversity of electronic records, we face a major challenge; that is, developing strategies, standards and processes to ensure electronic records are accessible for as long as they are needed." (1)

In the 9 years since the then Director General, Mr Nichols made that statement; there has been a move towards that goal. However, despite many tens of millions of dollars, thousands of "person" hours and countless projects to try and solve the many issues surrounding the longevity of electronic records, there is still not a single, long term, tried and tested solution to the problem of what on earth do we do with electronic records when they are no longer needed on a day to day basis, but still need to be kept for legal and/or other reasons.

Whilst there have been a few notable exceptions, few organisations have yet been brave enough to attempt an electronic archiving implementation strategy based entirely on current thinking. (2) It is not hard to see why. Jeff Rothenberg said way back in 1995 "it is only slightly facetious to say that digital information lasts forever – or five years, whichever comes first." (3)

This month's issue of Information Enterprises Newsletter "Information Overload" looks at the problems associated with keeping electronic records, and perhaps more importantly, what work is being done to ensure that an organisation's electronic archive is not a black hole into which documents are sent, but nothing ever seems to come back out again.

Please note, it appears that the terms "digital" and "electronic" are seemingly interchangeable in the literature, therefore we will be using the term "electronic" to denote both terms, unless specified otherwise.

But first, a bit of background.

© IEA 2004. All rights reserved. You are free to use material from the Information Overload newsletter in whole or in part, as long as you include acknowledgement of source.

Information Enterprises Australia Pty Ltd
Unit 4, Upper Level, 201 High Street, FREMANTLE WA 6160
Tel: 08 9335 2533 Fax: 08 9335 2544 e-mail: training@iea.com.au

For the un-initiated an electronic record is not just a wav file that you've acquired, or an MP3 music/audio file that you happened to have lurking on your home computer, just as a record is not simply a piece of vinyl that has music recorded on it.

A record can be defined as follows:

"Information created, received, and maintained as evidence and information by an organisation or person, in pursuance of legal obligations or in the transaction of business." (4), whilst this definition is considered to be the definitive statement, the West Australian State Records Act 2000 goes a little further and describes a record as:

"Information however recorded and includes –

- (a) any thing on which there is writing or Braille;
- (b) a map, plan, diagram or graph;
- (c) a drawing, pictorial or graphic work, or photograph;
- (d) any thing in which there are figures, marks, perforations, or symbols, having a meaning for persons qualified to interpret them;
- (e) any thing from which images, sounds or writing can be reproduced with or without the aid of anything else; and
- (f) any thing on which information has been stored or recorded, either mechanically, magnetically, or electronically." (5)

If that is the case, then electronic records must be treated in exactly the same way as their paper based (or other) equivalents, and should be preserved and maintained for as long as the legislation requires you to do so.

Easier said than done? Well, yes it is. One of the problems with trying to maintain access to electronic records is obsolescence of the hardware and the software on which the material is created and stored. It is unlikely, for instance, that records created using the very first computer driven technology can still be accessed, read and acted upon in today's electronic environment. It is perhaps one of the biggest ironies of the modern age, that whilst we have more information generated on a day to day basis, by more and more people, we are unable to utilise this valuable resource simply because we can't find it, or if we can find it, we can't access it because we no longer run the operating system that it was generated upon.

To give you a couple of examples, the National Space Agency (NASA) sent satellite probes into our galaxy to record information. The information from the Viking probes was recorded onto digital tape. The tape recorders are now obsolete and most of the data has been lost. (6)

And during the 1980's an ambitious project was undertaken to celebrate the 900th anniversary of the original Domesday Book of 1086. Entitled the BBC Domesday project, the idea was to produce a modern day equivalent of the famous book. Pupils from over half the schools in the United Kingdom were recruited to work on the project and information was gathered and data input using BBC Micros (the common pc in use at the time). Because of the number of photographs and video footage, a decision was made to produce the information on a videodisc. Developing the hardware took two years and was undertaken by Philips, the only manufacturer of the videodisc players in Europe. Logica wrote the software using BCPL, a forerunner of C. In all, over 70,000 lines of custom code were written.

It's not hard to see where this is heading is it? Once the project was completed, the price of the discs was a massive £4,000 and far too expensive for either organisations or individuals

to purchase, especially given that they could only be run using specialised machinery. A copy was given to the keeper of records at the Public Record Office to be placed alongside the original Domesday Book. The project ended, the resources that had been ploughed into it were re-assigned and the hardware and the software quickly became obsolete. The irony of course is that the original Domesday Book can still be read. The videodisc cannot. (7)

I know that neither example I have just given to you equates to a major catastrophe. So the data cannot be read, so what if the media is now obsolete, it doesn't matter does it? Well imagine if those projects contained information relating to patients suffering from the effects of asbestosis, or if the data contained information on the research you did, for a patent that you currently hold, that someone is now challenging. All of a sudden, the problem of hardware and software obsolescence takes on a higher degree of urgency and concern than a school project doesn't it.

I can hear a degree of muttering about emulation software, and that will solve all those problems. Well yes and no. To carry on with the example of the BBC Domesday project, another project was embarked upon in order to try and rescue the material held on the discs. The CAMiLEON project (Creative Archiving at Michigan and Leeds Emulating the Old on the New) managed to acquire a semi-working system from the School of Geography at the University of Leeds. Whilst the project team wanted the system to run in an "open-environment" the software used to emulate the original code (BeebEm) was programmed in a Windows environment, and therefore the whole project is now tied to the Windows platform. Does anyone else have a sense of Déjà vu?

We seem to have a very short memory when it comes to computer technology, we know that the operating system that we are running today will be replaced with a more modern equivalent in the not too distant future, and we will happily install it (or have it installed for us – whether we want it or not) with little thought or regard as to how much data we will lose as a direct result of the upgrade. Unfortunately, until more people "give a damn" problems like these will continue. Until then, we will continue to use proprietary software and let someone else worry what will happen if the company we have chosen goes out of business, or withdraws the technological support for the maintenance of older versions of software that we are still running. Why should we care, after all – we only work here!

That may sound cynical, but unfortunately for the records managers, librarians and archivists who are desperately trying to keep up with the plethora of documents that are created on a daily basis, using technology they know will be obsolete in a few short years, the problems of what to do with records, and electronic ones in particular is all too real.

Whilst we've touched on some of the problems associated with keeping electronic records, we need to get a little more specific.

An electronic record encompasses more than those records that have been "born electronic" – for example – generated on a computer, using a software program such as Word, but encompass such things as E-mails, MSN Chat and Mobile phone text messages too. However, in today's discussion we should not forget those records that began life in our organisations as paper records, and have been scanned to an electronic format (digital object), and form part of our paper archive, including those very early handwritten documents such as letters, faxes and memos, as well as research results written into a laboratory notebook (still current practice for most research departments world wide), as well as those documents that were typed, using carbon inserts to make multiple copies.

However, an organisations archive can also include copies of computer-generated documents from other organisations – for example tax invoices that have been supplied against goods or services.

For example, the originating organisation has a need to keep the electronically born invoice to prove that the item was supplied at a specific cost. The receiving organisation has a need to keep a copy of the invoice to prove when they received it, and when it was paid. However, to make matters even more interesting - If the receiving organisation adds the data to an electronic database such as MYOB or QuickBooks, then the evidence relating to the transaction will then exist in more than one place in more than one format (paper and electronic), but both have the same legal requirements attached to it.

Whilst most organisations seem to be quite happy with the concept of what to do with their paper based records, the same cannot be said of electronic ones. Paper based records are a tangible entity, they can be actioned, indexed and put into an “archive box” and sent off to storage. We are quite happy, safe in the knowledge, that barring disasters, the records in the archive box will be accessible by our organisation for as long as we need to keep it. We can retrieve it from the archive and be able to read whatever it is that is contained in its pages, and be satisfied that the record has not been altered or tampered with in any way. If we have assigned a disposal date to the box of records, we can then destroy the record (or have it destroyed for us) safe in the knowledge that we have acted within the current boundaries of the record keeping legislation. As a matter of interest, the 2003 edition of the Australian Record Retention Manual (ARRM) contains references to over 1660 pieces of legislation that directly relates to record keeping. Whilst this figure is quite staggering by itself, the 2003 edition amended almost 1,000 of the references listed. (8)

Whilst the archiving of paper based records appears (at least on the surface) to be a manageable task. And before everyone in the records management industry jumps onto their soapboxes and berates me for making it sound like it’s an easy job. I know it’s not as simple as it sounds. I am well aware that trying to keep pace with the glut of records generated by an organisation each year is staggering. But the point is – they are a very visible entity. Electronic records on the other hand exist in cyberspace, and therein lies part of the problem.

Bill Gates said in 1981 that “640k ought to be enough for anybody”, as we move through the beginning of the 21st century we are faced with terabytes of data being generated each year, all of which has to be managed in some way, shape or form. *Note: A byte holds the equivalent of a single character, such as the letter A, a dollar sign or decimal point. For numbers, a byte can hold a single decimal digit (0 to 9), two numeric digits (packed decimal) or a number from 0 to 255 (binary numbers). A Terabyte is a trillion characters of instructions or data.*
<http://www.techweb.com/encyclopedia/defineterm?term=byte>

If managing the electronic environment, should be akin to the paper-based environment, then it would make sense for records that are not being used on a day-to-day basis to be archived, thus allowing space for new documents. However, that does not seem to be happening in an organised or consistent manner. Because of the relative cheapness of “server” space, it appears that most organisations have adopted the policy of “keeping it live.” Why worry about archiving stuff when server space is cheap and getting cheaper. Whilst this is not a bad approach to take, there are a number of problems that have been encountered by proponents of this method.

Disasters:

© IEA 2004. All rights reserved. You are free to use material from the Information Overload newsletter in whole or in part, as long as you include acknowledgement of source.

I've often heard the saying "technology is great, when it works" and yes, when everything is humming along nicely, most people don't have a bad word to say about the systems they are using. But when it breaks down, the curses can be heard a mile away. But being serious for just a moment. If I were to hazard a guess, I would say that we have all suffered from some sort of computer disaster at some point during our professional lives. Lost documents due to power failures, system crashes, hard drive failures, back up tapes refusing to back up, viruses, hacking, industrial accidents, terrorism, the list is endless.

It is said that every organisation will face some sort of disaster at least once during its operational lifetime. Disasters can and do strike with impunity, usually at the least convenient time; it is how you manage the event that ultimately makes the difference.

So what can you do to counter this type of problem?

- Mirror sites in alternative locations;
- Multiple copies of vital records in alternative locations;
- The use of different media types, preferably in alternative locations. For example back up tapes held in safe storage;
- Rely on IT to rescue the data from back up tapes; Assuming your back up tapes were stored away from the main seat of operations, back up tapes may ensure that you are back up and running very quickly. However, if your back up tapes are stored in the same building and the building is destroyed – Earthquakes, Industrial accidents and terrorist activities - then chances are you will not be able to recover your vital data, and therefore reconstruct your business model.
- Keep your fingers crossed and hope it never happens to you.

Whilst the last comment may sound a little facetious, disaster planning is like having insurance. You decide on the policy and the level of cover that you want, pay the premium, and then hope you never have to use it. Of course some organisations don't bother buying "insurance".

Migration:

By keeping all the documents "live" you need to migrate all documents in all formats across upgrades and platforms as they occur. As we have already discussed, the cycle time between software versions has shrunk dramatically. Migration of documents across software upgrades can be problematic, and software enhancements and upgrades can render old files and documents unreadable by the later versions. "Migration is essentially a translation. With migration, as with all translations, some information is lost, no matter how skilled the interpreter. In migration, it is usually the context, rather than the data, that drops out or is improperly reconstructed in the new code. This can be crippling in dynamic formats, in relational databases, and even in simple spreadsheets." (9) Unfortunately, with the speed to obsolescence of digital media, organisations cannot wait for the optimal solution, and must risk migrating their information, or not be able to read it at all in a few short years.

Maintaining Document Integrity:

Whilst the cost of additional server space is relatively cheap, the use of embedded objects such as images have increased the size of individual files dramatically. What happens if the record that you are planning to archive, contains links to other objects, documents, or web sites (hypertext links), how do you maintain the objects, the links and their integrity once you have decided to archive the item? How will this affect your ability to produce a document in its entirety should you need to, especially in the face of litigation, patent challenges and so on?

As with all these things it usually takes some sort of precedent in order for others to sit up and take notice. For example, the Sarbanes-Oxley Act 2002 (USA) was introduced by the American government in response to the collapse of companies such as Enron, Worldcom and GlobalCrossing, and now seeks to tighten regulation through five main areas of corporate governance. These areas are:

- Auditors;
- Disclosure;
- Compensation;
- Board of directors; and
- Ethics.

The Sarbanes Oxley Act 2002 states that non-compliance with the rules applying to the maintenance of records is a federal crime in America and can result in a jail term of up to 20 years and large fines. The Act also governs accounting practices and specifies mandatory retention periods of five years for all audit and review work papers. Failure to keep records (in whatever format) for the specified term can result in jail terms of up to 10 years.

Complying with the Act requires that an organisation has the ability to produce, on request, authentic and reliable records and all supporting documentation. Section 1102 of the act is concerned with tampering with records or impeding official proceedings and states that:

“Whoever corruptly – (1) alters, destroys, mutilates, or conceals a record, document, or other object, or attempts to do so, with the intent to impair the object’s integrity or availability for use in an official proceeding; or
(2) otherwise obstructs, influences, or impedes any official proceeding, or attempts to do so, shall be fined under this title or imprisoned not more than 20 years, or both.” (10)

If you think that’s all very well, but what has the Sarbanes Oxley Act got to do with us. Well any Australian company that is an SEC (the United States Government’s Securities and Exchange Commission) registrant, as well as those Australian subsidiaries of US or European parent companies that are SEC registrants MUST comply in full with Sarbanes Oxley. Does your organisation need to comply?

Filing Systems:

The electronic filing system should mirror your paper-based system, so that all records, regardless of format can be actioned (e.g. produced if required by a court of law, or destroyed as per an organisations retention and disposal schedule) at the same time. However, users tend to regard electronic documents as their “own property” rather than belonging to the organisation. This is especially true of E-mail. However, it should be noted that, as with paper based records - not all documents that are created need to be kept in perpetuity, and more than one copy of an electronic document may exist in more than one place on the server, making it very difficult to action the item in the first place. Whilst user education may help to solve some of the issues, such as the keeping of multiple copies of documents, it is unlikely to eradicate the problem completely. There will always be someone, somewhere who keeps a copy “just in case”.

Please note: when marking electronic records for destruction you should also ensure that you have destroyed copies on servers and backup tapes. However, as some people have found to their cost. A record can be recovered from an electronic system, months, if not years after it has been deleted

from the system. Why? Well until the space on the server is overwritten with new data, the "old" data will still be there, even though it is not visible to you or I.

Security:

It is very easy for anyone to alter or delete electronic records, even with the tightest security measures in place; you just have to have the correct access. For those of you who are interested in the subject of computer crime, and hacking in particular, I would recommend reading a book by Clifford Stoll, entitled "The Cuckoo's Egg" for information on how to break into secure sites and what system managers are doing to rectify the problem. (11)

Whilst user education may prevent inadvertent use of old documents, (version control) some organisations may feel it essential to use encryption and digital signaturing technology as a way of ensuring data security.

Media Fragility:

We've mentioned some of the problems with media fragility, for example, failed back up tapes, and unreadable discs. However, one of the biggest problems with media fragility is human intervention. How many times have you inadvertently hit the cancel button instead of save and watch your morning's work go into the ether never to be seen again? How many of you wish you could have a "go back button?"

Whilst hitting the delete button rather than the save button can be classed as a minor inconvenience, the loss of data through misuse or age of the storage medium cannot be discounted. Whilst it is recognised that electronic media will outlast the hardware and the software needed to read the data, the media itself is not immune from obsolescence either. The problem with media obsolescence is that the bits and bytes contained on the tapes and the discs, slowly decay over time, and the only way that you know you have a problem is when the media "corrupts" and cannot be read when you try and re-load the data. The question is how do you know when it's going to happen? Unfortunately we have no way of knowing, and therefore we have to rely on media refreshing to ensure the data's integrity.

Media Refreshing simply means copying the original data from one media source to another media source of the same type. For example tape to tape. (As opposed to migration which transfers data from one media source to an entirely different media source for example from floppy disc to CD-R).

Libraries have not been immune to the problems of media obsolescence either. With a mandate to provide access to information in whatever form, this is not a simple task. In a paper presented in 1998, Deborah Woodyard of the National Archives of Australia explains the problems they encountered trying to maintain access to their collection of electronic material. (12) An additional problem facing libraries and organisations who use the Internet/Intranet for the storage of information, it has been said that the "average lifespan for Web sites is just 44 days, according to James Billington, Librarian of Congress." (13).

The National Library of Australia has looked at the problems of providing long-term access to electronic information via the Internet. Entitled PANDORA (Preserving and Accessing Networked Documentary Resources of Australia), the National Library of Australia uses selectively chooses web pages to archive, based on a pre-determined harvesting schedule. The National Library acknowledges that there are disadvantages with the selective approach., as they are making subjective judgments about the value of resources and what

researchers of the future may find useful. Librarians have always made these decisions but the dissemination of information online is still in its infancy and the way that researchers will want to access, use and apply the potential of the Web is still developing. Selective archiving takes a resource out of context and often does not include other resources to which it is linked. (Harvests individual web pages rather than entire sites), therefore some contextual meaning is lost. (14)

What about PDF?

We have mentioned that keeping documents in whatever format “live” may be one of the simplest answers to the electronic archiving problem, in other words – don’t bother to archive documents at all, there are some organisations who choose to “archive” their documents to PDF.

Unfortunately Adobe Acrobat’s Portable Document Format (PDF) is not the answer either. Whilst there are some people who might think that converting documents to PDF will solve some or all of their organisations archival woes, as a stand-alone strategy it is flawed on several counts.

Adobe Acrobat, like Windows is proprietary software. When Adobe first launched the software they were keen to tell people who bought the software that “free” readers would be available so that anyone who downloaded the documents could read them today and well into the future. They were also keen to stress that they would maintain their software for upwards of 25 years and many generations of the software. For those of you who remember BBC Micro Computers, AMSTRAD and Atari operating systems, we know that maintaining software through multiple generations is not a small task. Adobe also emphasised that their software had both forwards and backwards capability. This ensured that documents which were created in the earliest versions being readable in a later one, and vice versa. This was a major advantage over other software programs such as Microsoft who relied on people upgrading their systems to maintain readability of their electronic media.

There is also a degree of security possible with using PDF’s as an archival medium, as security can be assigned at document level to prevent unauthorised copying, saving and printing.

So why is PDF not a good stand alone archiving strategy?

- A PDF is a snapshot of how a document “looked” at a particular point in time. Like its other electronic counterparts, a PDF document may also be overwritten or deleted (accidentally or otherwise). Even with security measures in place, the person who attaches the security to the record can remove it. Throwing into doubt the records reliability, authenticity and accuracy.
- A PDF does not contain the ‘metadata’ of the original document. For instance, in a “Word” document you can track changes, and know which machine was used to make them. If an organisation insists on strict password control over individual pc’s, then the changes may also be linked back to an individual.
- Multiple copies of a record may exist in more than one place, and different versions of the same document may also exist.
- Who determines what should be converted to PDF and treated as part of the organisations archive? And who manages the conversion process? It is unlikely that an organisation will be able to afford a PDF writer for individual PC’s, so determining which documents are “harvested” can be problematic.

- What happens if Adobe withdraws its free reader from the market? Whilst there are some “open source” versions available, readability and compatibility may be a problem.
- Where are the documents stored once they are converted to PDF? Will they be kept “live” or will they be “archived” onto a different part of the server, or an entirely new server. Or will they be archived onto a different media entirely, for instance CD-R or DVD-R. What happens when the archive server becomes obsolete and needs replacing? Who will check to make sure that all the existing PDF documents can be read with the new software and hardware?

Whilst Adobe made an early undertaking to ensure the formats would be compatible for the foreseeable future, they may ultimately decide that it is no longer in their commercial interests to do so.

To counter some of the problems using PDF as a stand-alone archiving solution, some organisations are using a format called XML to capture the metadata associated with the document, usually into a web based environment. XML or Extensible Markup Language is designed to improve the functionality of the Web by providing more flexible and adaptable information identification. It is called extensible because it is not a fixed format like HTML (a single, predefined markup language). Instead, XML is actually a ‘metalanguage’ — a language for describing other languages—which lets you design your own customized markup languages for limitless different types of documents. XML can do this because it’s written in [SGML](#), the international standard metalanguage for text markup systems (ISO 8879). <http://www.ucc.ie/xml/>

Open-Source Software:

Why is XML so important? Well XML is an “open-source specification” it is not reliant on any specific software or platform to operate it, does not cost anything to acquire, and can be modified and redistributed without fear of breaking any licensing agreements. For a complete definition of what constitutes “open source software” please go to <http://www.opensource.org/docs/definition.php>

The National Archives of Australia uses Open Source Software for the archiving of the records within its care. At a recent breakfast seminar presented by the NAA’s Simon Davis, Assistant Director, Digital Preservation and the Records Management Association of Australasia, Simon explained why open source software was a vital strategy for preserving access to an organisations electronic history.

Most people agree that if you have a paper document – you can preserve the object and you preserve the record. With E-records, people experience the record through a performance (by using appropriate software/hardware). Therefore with e-records if you preserve the performance you can preserve the record. However, there is the issue of data migration – if the record has been migrated through various versions – questions you need to ask yourself are:

- Is the version that I am viewing the version that the originator wanted me to see?
- Is it in the correct format?
- Can I see the object in the same way as the original creator saw?

The National Archives of Australia uses the “Open Office” suite of products. These are not linked to any proprietary software or company, but are cross industry based, and can be

adapted and updated within the open forum, by anyone with the knowledge and skill to do so.

The NAA say that the key to preservation is:

- Actively determining what it is you want to keep;
- Standards;
- Full documentation is vital (especially in the open environment);
- Active involvement in technology decisions – at an organisational level; (Simon is a member of the Organisation of the Advancement of Structured Information Standards (OASIS) Open Office XML Format Technical Committee).
- There is no silver bullet;
- No product-driven solution.

In addition to creating its own open source code, the NAA uses existing formats where possible as well and includes:

- XHTML
- PNG (web based graphics format) similar to GIF and JPEG
- Open Office XML Format – (www.openoffice.org)

When the NAA receive electronic records, they convert the data from “normal” programs through a process they term “normalisation”. This is a one-off migration process, ensuring that the essence of a document is not lost through multiple conversions/upgrades, as we have discussed earlier.

For more information on the NAA’s approach to archiving electronic records, please go to www.naa.gov.au/recordkeeping/er/digital_preservation/summary.html and www.naa.gov.au/recordkeeping/preservation/digital/xml_data_formats.html

As a solution to the problem of what to do with electronic records it is certainly one of the more viable options currently open to us at this point in time. Whilst time will tell how successful or not the strategy is, there will be some organisations who may prefer to adopt alternative short to medium term options until such time as a definitive answer is found.

Converting electronic records to paper:

I know this sounds like a backwards step, but as we have mentioned, people seem comfortable with what do with their paper based records. As with other options such as converting the document to PDF, this does not preserve the “metadata” of the record, but does ensure that it can be read by future generations.

As we all know, under the correct conditions, paper based records have a life span of hundreds if not thousands of years (For example, the Dead Sea Scrolls and the Domesday Book); There are major drawbacks to this of course. The costs of storage will go up, the personnel needed to manage the increasing workload will go up, and it is reliant on people doing the right thing and printing out a copy for file.

However, as an archival strategy (as with all strategies) there are problems associated with this methodology. In today’s dynamic working environments, there are some things that cannot simply be printed out and stored. Take for example dynamic databases created using java script, rather than the static HTML code, or videos and audio sound files. Whilst keeping a paper archive may sound like the answer to your archival problems, in reality for some information it simply cannot be done.

© IEA 2004. All rights reserved. You are free to use material from the Information Overload newsletter in whole or in part, as long as you include acknowledgement of source.

Extending the life of originating computer technology:

This has the short-term advantage in that the data retains its integrity and links that migration cannot ensure. It is not an option for long-term preservation strategies due to the fact that spare parts for the hardware become less available as the technology ages. A United Kingdom pharmaceutical company (15) took this idea one step further. Instead of migrating their data to a new server, the organisation took the drastic and interesting step of archiving the server itself. They loaded the server along with a number of computers and all software needed to access the data onto a flatbed lorry (sorry truck) and sent it to their repository. The organisation assumed that rather than trying to sort out a mess of old records, they could adopt strategies that would manage their new information properly, whether or not they have succeeded with this rather interesting strategy remains to be seen.

Use of Microforms (Microfiche, Microfilm etc):

This technology has a tried and testing longevity that newer technology does not yet have. Like the paper or PDF equivalent, a microfiche cannot capture the metadata of the record. However, even with hardware obsolescence the material can still be read, all you need is a torch and a magnifying glass (not to mention a steady hand!). Some organisations still use microfiche as a "back up" of their archives, usually at a separate location to satisfy vital records and disaster planning strategies.

Copying of electronic archives onto CD-R and DVD-R technology:

Instead of purchasing additional server space (which may not be cost effective for small to medium enterprises), organisations have been burning their electronic archives onto CD and more recently onto DVD instead. The popularity of the media appears to be the ease of use of the technology, the amount of data that can be stored and the relatively low cost of the media. Whilst CD-R seems to have been the technology of choice, (due to the fact that the technology has been around longer), the use of DVD is increasing, and part of this is due to the fact that you can store an awful lot more on a DVD than you can a CD. (4.7Gb as opposed to 700Mb), you can also store moving images and sound, essential if you need to store presentations and video footage alongside the documents you presented.

However, CD's and more recently DVD's are not immune to lost data or unreadable discs either, however this appears to be largely dependent on how the material was transferred, the burn rate and how much testing of the final product was done prior to storage, as well as how the media was stored after completion. CD's and DVD's can be broken just as easily as tapes. The difference with a break in the CD is that it cannot be repaired.

When embarking on an archiving project using CD or DVD it is wise to buy the best media that you can afford. Both Kodak and Mitsui make special "archive" discs, which come with their own jewel cases. The bulk buy, spindle type available from every good supermarket and are manufactured to cheaper standards do not.

Do not "fast burn" your CD/DVD's. Use a machine dedicated to the task and do not attempt to do anything else whilst the disc is being burnt. Unless of course you want more "coasters" to add to your collection. If you are planning on outsourcing the process, then make sure that the company you choose uses archival quality discs, has a slow burn rate, can do the entire process in-house and tests every disc for accuracy rather than undertaking a random test to save time. If you are planning on copying your archive to DVD rather than CD then it is advisable to use a company with advanced DVD authoring technology as used

by the TV and Film industry, rather than domestic standard as used by most organisations.
(16) (17)

As we have seen, electronic archiving is not without its risks. But it is a problem that is here to stay. It is how we choose to manage the problem that will ensure whether our electronic archive will survive for as long as we need it, or not as the case may be.

So what of the future?

- More money will be spent on digitising paper-based records, with little or no thought as to the long term consequences;
- Emulation software will play a huge part in giving access to previously unreadable material;
- Technology will continue to change at an alarming rate, migration and technology refreshing will become a normal part of an archivists and records managers archiving strategy;
- Information will continue to be lost, altered, deleted or damaged through ignorance, malicious intent and disasters; (18)
- There will be lots of changes, but everything will stay the same.

Notes:

(1) Nichols, G; Director-General, "The Electronic Challenge" page 3. In the booklet entitled Managing Electronic Records: A Shared Responsibility. Written by Greg O'Shea, National Archives of Australia, March 1995. The Electronic Challenge

(2) For example: National Archives of Australia; Public Record Office of Victoria VERS project, Indiana University Electronic Records Project.

(3) Rothenberg, Jeff. Ensuring the longevity of digital documents; Scientific American, January 1995, p42.

(4) Definition taken from ISO 15489-1: Information and documentation – Records management – Part 1: General

(5) Definition taken from State Records Act 2000: State Records Principles and Standards 2002. Western Australian Government Gazette, Perth, Tuesday, 5 March 2002 No.38.

(6) Smith, Thomas E, Issues inherent with implementing multiple technologies from a records manager's perspective p4. Paper presented at the ARMA-Madison Spring Seminar April 10, 2001. "Millennium Challenge: Managing Electronic Records.

(7) Mellor, Phil. CAMiLEON: Emulation and BBC Domesday; RLG DigiNews, April 15, 2003, Volume 7, Number 2, http://www.rlg.org/preserv/diginews/v7_n2_feature3.html

(8) The Australian Record Retention Manual is published annually by Information Enterprises Australia Pty Ltd. For more information please visit www.iea.com.au or telephone 08 9335 2533.

(9) Lawrence, Gregory W et al. Risk Management of Digital Information: A File Format Investigation. Council on Library and Information Resources, June 2000, vi (www.clir.org)

(10) The Act can be viewed in its entirety at:

<http://news.findlaw.com/hdocs/docs/gwbush/sarbanesoxley072302.pdf>

(11) Stoll, Clifford: The Cuckoo's Egg: Tracking a spy through the maze of computer espionage.

(12) Coughlin, Kevin; Archivists say computers have no sense of history. NJ.com Thursday 19 June 2003. <http://www.nj.com>

(13) Woodyard, Deborah, Farewell my Floppy: A strategy for migration of digital information. Presented at the VALA Conference 28th - 30th January 1998 in Melbourne on Electronic Preservation. <http://www.nla.gov.au/nla/staffpaper/valadw.html>

(14) Preserving And Accessing Networked Documentary Resources Of Australia (Pandora): <http://pandora.nla.gov.au/background.html>.

(15) Details withheld due to the sensitive nature of the information.

(16) Promote Media Group – www.procopy.com.au

(17) Technical Advisory Service for Images, Using CD-R and DVD-R for Digital Preservation; www.tasi.ac.uk/advice/delivering/cdr-dvdr.html

(18) Microsoft Office 2003 has an expiration date that can be added to emails and documents, allowing them to “self destruct” once that date has been passed. Finlayson, Stuart; This message will self-destruct. Image and Data Manager, January/February 2004p 12-15

Further Reading:

Collaborative Electronic Notebook Systems Association (Censa): www.censa.org

Friedberg, Errol C et al, How E-mail raises the spectre of a digital dark age. Nature 19 June 2003, Vol 423 P801

Information Enterprises Australia Pty Ltd, Australian Record Retention Manual: Section 5: Records Management Standards and Benchmarks abstracts all the major work being carried out by organisations in the UK, America and Australia.

InterPARES (International Research on Permanent Authentic Records in Electronic Systems) Project: <http://www.interpares.org>

National Archives of Australia: Design & Implement Recordkeeping systems (DIRKS): [DIRKS: A strategic approach to managing business information.](http://www.naa.gov.au/recordkeeping/er/biblio/er_biblio.html)

National Archives of Australia: Digital Records Bibliography: http://www.naa.gov.au/recordkeeping/er/biblio/er_biblio.html

Preserving Access to Digital Information (PADI): <http://www.nla.gov.au/padi/>

Records Continuum Research Group: <http://rcrg.dstc.edu.au/index.html>.

Rowe, Richard; Digital Archives: How we can provide access to ‘old’ biomedical information. Published in Nature’s Web Debates. <http://www.nature.com/nature/debates/e-access/Articles/rowe.html>

United Kingdom’s Data Protection Act 1998. The implications of the Hutton Inquiry on the privacy of emails as records. <http://www.hms.gov.uk/acts/acts1998/19980029.htm>

United Kingdom’s Electronic Records In Office Systems (Eros): <http://www.pro.gov.uk/recordsmanagement/erecords/default.htm>.

University of British Columbia Project: <http://www.interpares.org/UBCProject/>.

University of Pittsburgh Electronic Records Project. Note: Due to a technical glitch at the School the Web site with the working files of this project was destroyed. The Web site has not been updated since 1996 when the Project ended. Individuals interested in the project can access it through the Internet Archive. Go to the Internet Archive site (<http://www.archive.org/>) and use the “Wayback Machine” by entering the URL of the Pittsburgh Project (www.sis.pitt.edu/~nhprc).

Hope you enjoyed reading, have a great week.

A Thought to Ponder:

“Wisdom is not knowing what to do now, but what to do next.”

W.G.P

Your comments and suggestions on the subject of this newsletter are most welcome. Or if you would like to see other issues covered in future editions, please email training@iea.com.au

Please feel free to pass on this newsletter to your colleagues' friends and associates. To subscribe they should send an e-mail to training@iea.com.au with “subscribe newsletter” in the subject line.

If you would prefer not to receive this newsletter, please send an email to training@iea.com.au with “unsubscribe newsletter” in the subject line. If you have any suggestions as to what should be included in future editions, then please send an email to training@iea.com.au.

© IEA 2004. All rights reserved. You are free to use material from the Information Overload newsletter in whole or in part, as long as you include acknowledgement of source.

Information Enterprises Australia Pty Ltd
Unit 4, Upper Level, 201 High Street, FREMANTLE WA 6160
Tel: 08 9335 2533 Fax: 08 9335 2544 e-mail: training@iea.com.au