DIGITAL CONTINUITY

An introduction to the wider context

Draft for consultation March 2009

The National Archives

Will a public inquiry still be able to look at vital CCTV footage in ten years time?
digital information matters

This introduction is for anyone in government who needs a broad understanding of what managing information risk involves. It sets out the wider context for one particular issue: digital continuity.

Digital continuity is about ensuring that government information survives and is useable for as long as it is needed for ongoing business purposes.

Anyone who needs the information should be able to identify it, retrieve a copy, read, or otherwise use the content as required. They should be confident that the information is complete and authentic, and understand its context and significance.

What this introduction covers

1. Continuity: a universal problem

Increasingly, information is recognised as an asset that is valuable and needs to be managed and protected as carefully as other property – money or equipment.

Virtually all information is now digital and most transactions with government are conducted online. From citizens reviewing their car tax to the soldier on the battlefield, we all rely on digital information. But such data is radically different from paper-based information and this means that new and radical approaches are needed to manage it. Anyone who has a digital camera has a digital preservation problem and there is even increasing academic interest in the preservation of computer games with major projects in the UK and USA.

Digital information is fragile

Digital information is more fragile than paper-based information and is susceptible to a range of threats. These include:

- The availability of technology to provide access to information, following technological change and obsolescence.
- The quality of metadata and effectiveness of information management processes in providing meaningful context for finding and interpreting information.
- Governance processes and the management of organisational changes to provide appropriate risk management and continuity strategies.

- Malicious damage or accidental loss.

Although all organisations have rigorous back-up procedures in place, these will not ensure the long-term survival of digital information because they cannot protect against changes in format or software, and may not necessarily protect against loss of metadata.

Positive action needed

Ensuring the survival of digital information requires positive action, which might include changing the storage medium or migrating the software in which the information is written. Expert opinion varies as to how long digital information will survive without active intervention. The most pessimistic assumption is that unless something is done within 5 – 7 years, it will be at risk. Even the trade organisation which represents storage manufacturers does not believe that it is possible to keep digital information for more than about 10 years without taking some form of action.

Serious consequences

In some cases, the issues surrounding the survival of digital information are potentially hugely serious and can pose threats to life.

For example

Writing in 2005, Gavan McCarthy and Ian Upshall described the threats to the survival of information relating to nuclear waste. They said that: ‘Where information has survived and remained accessible for many years, it has often been the result of accidental circumstances rather than well-planned and adequately-resourced processes’.

They went on to make the point that: ‘The information required to manage the long-term safety of radioactive waste is so important that the present generation cannot afford to ignore information preservation failures of the past. It must also be prepared to invest the necessary resources now not just to remedy the inadequacies of past practice but prepare more robust systems for the future’.

Government requirements

More and more, government is recognising the need to ensure the continued availability of digital information and to deal with the threats posed by the risk of losing that continuity.

Consequently CESG (the National Technical Authority for Information Assurance) has included a minimum requirement in their Information Assurance Maturity Model that government departments recognise the risk to continuity of access to their business information assets arising from digital obsolescence in their strategic risk registers.

Higher levels of the Model now include the requirements to carry out risk assessment processes and to implement mitigation exercises. See: http://www.cesg.gov.uk/products_services/iacs/iamm/index.shtml

2. What information is at risk?

Government uses a diverse range of systems to create digital information and this adds to its complexity and vulnerability, as each has its own particular continuity issues.

Formats can be broadly categorised as:

- Office systems
- Business systems
- Websites
- Images
- Design software

- Non-official delivery and file sharing mechanisms

Office systems

This category includes:

- Communications systems – emails.
- Policy and other documents – Office software.
- Presentations – Microsoft PowerPoint.
- Financial information – spreadsheets.

Some of this information is stored in structured corporate systems such as Electronic Document and Records Management Systems (EDRMS), while some is stored on shared drives, and some organisations have dedicated email archiving systems.

Business systems

These are normally based on database technology and include business-specific database systems and ones covering HR, Finance, Case Management and so on. These systems are normally updated when new versions of software are introduced but might need to be migrated to another application at a break point when one system, for example DB2, is replaced by another, for example Oracle.

Websites

These contain a wide mixture of types of information, including official publications, news items, and statements of policy and guidance. Web browser-based technology is used for a wide range of other applications across government, including departmental intranets.

Images

Increasingly used for surveillance and traffic management, video is by far the fastest growing type of electronic information worldwide. There appear to be no systematic approaches for storing this material, significant quantities of which are of potential forensic value. Moreover, the images themselves are of no value without metadata which identifies them.

Design software

CAD systems are increasingly used in a range of building- or technology-related applications, while most departments use Geographical Information Systems. Again, there is little evidence of systematic approaches to the storage of this material.

Non-official delivery and file sharing mechanisms

Increasingly, government departments are using non-official applications as places to work on or deliver information. A number of departments now deliver videos using YouTube, while some documents are produced using file sharing technology, such as BaseCamp. The survival of this material depends on the whim, or financial stability, of the body which is hosting it.

3. Why preserve digital information?

Business- critical digital information needs to be preserved for as long as it is required, to support ongoing business activities or to provide a record for audit or governance procedures. It should be seen as an asset – as valuable as cash, buildings or human resources.

Organisations need to have processes in place to identify what information is key to the management of their businesses, to have assessed the risks to such information and developed plans to mitigate identified risks.

Digital information needs to be preserved for a variety of reasons, as follows.

Legal compliance

Government must be able to access information it is legally required to retain and/or provide, for example under Health and Safety, Freedom of Information and Data Protection legislation, and the Public Records Act.

Cost avoidance and saving

Acting to ensure the survival of digital information lowers the likelihood of three types of cost:

- The business cost of being unable to use information, for example for service delivery or policy formulation.
- The financial cost of any necessary data recovery.
- The reputational cost of losing it.

Action to mitigate risk will also help departments better understand what information they have, what they need to keep, and what therefore they can discard. Most departments are currently paying increasing amounts for energy and management to store data they don't need to keep any more.

Support for policy making and service delivery

Government cannot do 'business as usual' unless it has reliable access to all the information it needs, for as long as it needs it; continuity of access protects the investment made in creating the information, enhances government's ability to make sound, evidence-based decisions, and underpins good service delivery.

Maximising efficiency and effectiveness of public services

Beyond the 'business as usual' requirement, government aims to use technology and information to develop its capability and improve service delivery, for example through the Transformational Government agenda. Ensuring continued access to information is a key enabler for efficient and effective government.

Accountability, accessibility and good business practice

There are policy and public expectations for the way government carries out its accounting and audit functions, and an obligation to remain open to any future public or parliamentary scrutiny. The risk of information loss compromises government's ability and reputation in these areas. Enquiries can take place a significant time after the events themselves. For example, the BSE enquiry, which reported in 2000, looked at government information going back to 1970.

Reuse

There is increasing emphasis on the potential reuse of information and on creating value by allowing information assets to be shared more widely. The 2007 Power of Information Review (http://www.opsi.gov.uk/advice/poi/power-of-information-review.pdf) states that government should now grasp the opportunities that are emerging in terms of the creation, consumption and re-use of information. The Review states that the government should:

- Welcome and engage with users and operators of user-generated sites in pursuit of common social and economic objectives.
- Supply innovators that are re-using government-held

information with the information they need, when they need it, in a way that maximises the long-term benefits for all citizens.

- Protect the public interest by preparing citizens for a world of plentiful (and sometimes unreliable) information, and helps excluded groups take advantage.

Supporting openness, accountability and democracy

Government needs to be able to preserve its key information for the future with authenticity and integrity. If this information were consistently lost, public confidence would be undermined, and society would lose the ability to make historical judgements about government. That would compromise openness and the long-term robustness of the democratic process.

4. What is digital continuity

The aim of digital continuity is not to create an archival object which will remain readable for hundreds of years. Rather, it is to make sure that a person seeking government information would, for as long as is necessary, be able to:

- Read, or otherwise use and understand the content as its producers intended.
- Decide whether the information received is trustworthy.
- Understand the information context and exploit its value – by being able to exploit embedded references (links), read or use the metadata or other contextual information, including location.
- Exercise all this functionality conveniently.

Essentially, continuity of information involves sustaining its **completeness**, **availability** and **usability**.

It is therefore important that not only the information, but also the associated metadata, is kept and that it remains useable.

For example

One of the biggest data loss crises in recent years happened because the Japanese government was unfortunate enough to lose the connection between the data about pensions and the names of 50 million people to whom the records related. This led to a major political crisis in Japan. In January 2008, the new Prime Minister announced: 'The careless management of public documents, such as pension records, is absolutely unacceptable. We will conduct a fundamental review... for managing administrative records and will consider their legislation, and furthermore, we will improve the system for preserving public records, including expanding the national archives.'

Information with ongoing business value

Unlike paper, which can be kept on a shelf for lengthy periods, digital data needs active intervention; it cannot simply be left in a computer or stored on tape or disk. However, active intervention costs money and consequently it is important that an informed decision is taken early in the life of a digital object as to how long it will be needed for ongoing business purposes.

According to Hewlett Packard, when one English county council moved to an electronic records management system, they found that 75 percent of their records were no longer needed for business purposes.

The National Archives is currently leading a programme to develop guidance on identifying the value of information to the business so that it can decide what needs to be kept for the long-term.

5. Threats to the continuity of digital information

Digital information is at risk because of its very nature – the media on which it is stored and the software in which it is written is fragile and can easily decay. It is widely reported in the digital preservation community that vast quantities of data are at risk because of technological obsolescence, which might affect the software, operating systems or hardware used to create documents or the media used to store them.

For example

Over ten years ago, Terry Kuny warned of apocalyptic loss of information in his article *A Digital Dark Ages? Challenges in the Preservation of Electronic Information*. In the UK, the National Council on Archives has warned that if we do not manage our digital assets well, we are heading for a black hole in our and our organisation's and our country's collective memory.

The problem of continuity is not just technical. Perhaps more losses occur because of managerial and human issues. Our view in the Digital Continuity team is that it is often more effective to intervene at the strategic and information management levels, rather than just treating the problem as a technical one.

Managerial risks

Probably the greatest threat to the survival of digital information is its sheer volume. Computer systems, email, scientific data sets and surveillance cameras are capable of creating volumes of data that were unimaginable in the paper world.

For example

The United States National Archives have estimated that in 2010 they will have 10,000 terabytes of data to be preserved forever. In 2020 they will have 230,000 terabytes, and in 2022, 350,000 terabytes.

Such huge volumes of information pose great difficulties for the future. There are considerable costs associated with storing data – maintaining disks or tape libraries and cataloguing or managing it.

The complexity of digital information adds to the problems of managing it. Paper records can exist only in one location at one time and it is relatively easy to identify which is an original paper record. As a result, they are relatively easy to manage.

Digital information can exist in multiple locations. Consider email, where an instance of an email might exist on my PC, on your pc and on the servers of my email provider and your email provider. In this context it is hard to say which is the original, and the concept of an original document is hard to sustain in the digital world.

For example

In June 2008, the US General Accountability Office looked at four US government agencies. It made a detailed analysis of the record keeping practices of 15 senior officials and found that 8 of them kept their emails in systems which were not recordkeeping systems.

Slightly earlier in the year, a pressure group called Citizens for Responsible Ethics in Washington conducted an in-depth investigation of record keeping in the US government and found significant failures.

Fundamentally, digital information will only survive if there is a clear decision by the organisation – backed up by the provision of sufficient resources – to take steps to ensure its survival.

Consequently senior staff need to recognise the ongoing business value of the information they are creating.

Organisational change

Work by the Australian Science and Technology Heritage Centre has revealed that one of the most serious and likely sources of knowledge loss is institutional or organisational change.

Change can occur at a personal level when a particular expert retires, taking with them valuable implicit expertise in their area of work, knowledge about what information exists and where, and of its significance.

An everyday example of how knowledge can be lost through change is the increasing use of short term contract staff. In order to spread costs, reduce timescales and improve efficiency, many organisations use third-party experts who meticulously build up a knowledge-base which is subsequently lost when the contract ends. Thus, the accumulated knowledge becomes increasingly dispersed and disconnected.

If this knowledge dispersal is then combined with organisational change, the likelihood of losing the information increases dramatically. Although information at all levels faces these risks, it is the critical contextual information, which is not systematically managed, that is most at risk.

For example

There are examples from UK central government where information loss has occurred because of machinery of government changes. In one case it was discovered that information about 30,000 physical folders could not be found and they had to be re-catalogued from scratch.

In another case, a department had to spend £200,000 re-indexing and relocating paper files. In a third case, emails were renamed and exported in a format which meant that their attachments could not be read. While in a fourth, the way in which information was exported meant that electronic files could not be accessed without a complex look-up table.

Perhaps better known are the many examples of information loss that occur at the end of projects. Project managers and their staff, anxious to move onto the next thing, neglect to ensure that proper arrangements are made for their records. While such a problem was recoverable in the paper world, digital information that has not been properly captured and structured is virtually impossible to recover.

Governance and culture

A failure to have comprehensive oversight and control in the form of an effective information management programme means that vital information might be lost or be susceptible to theft or damage.

For example

The coroner at the inquest into the loss of an RAF Hercules in Iraq in 2005 said that: ‘I believe that the ability to retrieve and view documents that record key decisions as not just important, but essential - equally important is the rationale behind them’.

This inquest has been plagued by an inability to retrieve documents. He went on to say that a military policy of ‘shredding documents’ was ‘difficult to come to terms with’.

A failure to have appropriate controls over third party suppliers might result in crucial information being lost, mislaid or used inappropriately. There have been a number of incidents during 2008 where third party contractors have lost USB memory sticks containing crucial information.

Information management policies, processes and systems

Basic information management systems are vital if government departments are to have adequate controls over their information. Unlike the paper world, digital records need special care if they are to survive and be reusable.

For example

In the United States, the White House has been embroiled in a lengthy and complex law suit brought by an independent research institute and library, the National Security Archive, concerning the loss of emails between 2003 and 2005. The White House had dismantled the Clinton administration's Automated Records Management System, but had failed to put in place adequate steps to ensure that emails were kept.

By 2008, a number of the missing emails had been found on back-up tapes, but there was evidence that some had been lost because of the White House's habit of reusing tapes and of wiping the hard drives of obsolete computer tapes.

The human dimension

There are serious risks from users of digital information, including both accidental loss (well documented in the press) and deliberate fraud.

For example

In 2008, three members of staff of a government department stole £390,000 by colluding with external accomplices to perpetrate frauds against the payments system. They supplied information about some of the department's customers to the external parties who used the information to make false claims.

Ensuring authenticity

In order to ensure that the information held by an organisation is authentic and valid, and to guard against the possibility of deliberate loss through staff actions, it is important that organisations have a forensic readiness policy in place.

Such policies are designed to enable organisations to conduct investigations into security incidents or computer-related crime. However, the essence of such policies is that they require the careful collection and preservation of potentially useful data, which could also be used to ensure the authenticity of digital records.

Further information about Forensic Readiness Policies is available at: http://www.cpni.gov.uk/docs/re-20050621-00503.pdf

Technical risks

Loss of continuity of access to digital information can occur in three major ways.

Lack of knowledge
It can happen because of simple ignorance. Without a careful investigation, departments may not know what files they have on their systems and thus may not be able to take steps to ensure their survival.

A recent survey of The National Archives' own business file store revealed a significant volume of unidentified digital files – probably most of these were of little ongoing value, but this needs further investigation to avoid the risk that essential information might be lost.

Metadata failure
Loss of continuity can occur because of a metadata failure, when the original documents survive but the information about them is lost. There are examples from hospitals of x-ray records losing their value because the relevant patient data has been lost.

Complex and interrelated systems also pose difficulties for preservation. Even more problematic are the videos from surveillance cameras, which only have meaning if information is kept about when and where the images were captured.

Technical obsolescence
The ability to access digital information may be lost because:

- The software in which it was written becomes obsolete. Early word processing files such as Wordstar are now very difficult to open in modern word processing systems. There is a considerable volume of important information still in Word 97 format and it is not certain how long this will be fully readable in future Microsoft systems.

- The operating system in which the software runs becomes obsolete. There are numerous examples. In November 2008, Microsoft announced that it was withdrawing support for its Windows 3.x operating system. The Word 2007 equation editor is incompatible with that of Word 2003 and previous versions. And when converting DOCX files to DOC files equations are rendered as graphics. Consequently, Word 2007 cannot be used for any publishing, file-sharing and collaborative endeavour in any mathematics-based fields, including science and technology, in which users may have earlier versions of Word.

- The hardware required to run the operating system no longer functions.

- There are incompatibilities between different generations of the same software.

- Failures in storage media also pose a serious threat to the survival of digital information. Many suppliers make claims that their media can survive for very long periods. Some tape manufacturers claim that their tape can have an archival life of 15 to 30 years. However, some academics have claimed that it is impossible to verify these claims. It is unlikely that, even if the tape itself survived, there would be suitable readers available in 30 years' time.

- Business obsolescence poses a significant threat to digital information. Existing systems are frequently replaced to meet new business objectives (for example a new finance system might be acquired to improve the management of resources). However, unless appropriate steps are taken to identify all data stores and to migrate data to new formats, then there might be significant losses of data.

Websites
Websites pose particular difficulties. On the one hand, they are quite ephemeral with frequent changes in content and even URLs. Some have content management systems that automatically delete material after a set period.

On the other hand, many contain significant material – publications and detailed information that is not available elsewhere. A recent investigation showed that 60 percent of URLs cited in

Hansard between 1997 and 2006 were broken.
The National Archives is carrying out an in-depth harvest of all UK central government websites three times a year, and intends to preserve these and make them available via the European Archive http://www.europarchive.org/

However, increasing numbers of departments are using file sharing sites, including Flickr and YouTube, to display images and video clips. These pose particular preservation issues if copies of the videos and images are not held elsewhere, because the National Archives is not currently able to archive this information.

6. Managing digital information to support continuity

In order to ensure that government has continuity of access to its business-critical digital information, departments need to implement a programme to manage their digital information. This should cover:

- Identifying all digital information held by the organisation, including policy files, email stores, databases, GIS material, video and so on.
- Identifying all material that the business needs to keep for more than 5 years to support its ongoing business. Five years is the critical period because if material is needed for longer, then action will begin to need to be taken to ensure its survival. All metadata relating to the material should also be identified.

- Developing secure processes to dispose of material that is not needed for ongoing business purposes. Secure disposal should include deletion of material from live systems and from back-up tapes or disks. Destruction of unwanted material must be done in a controlled way to meet the requirements of the organisation under Data Protection and Freedom of Information legislation.

7. Implementing a Digital Continuity programme

A Digital Continuity programme should be implemented following the guidance and standards issued by The National Archives, using a framework of evaluated tools and services to manage and control the risks. This will provide a flexible structure which will enable departments to respond appropriately to their own individual digital continuity challenges.

The Service will consist of the following components:

- **Guidance and support**: on how to understand and address digital continuity risks, create an action plan and effectively mitigate identified risks using the right tools and services.

- **Commercial Framework agreement**: a selected range of commercially provided tools and services from multiple providers, assured by TNA to address specific digital continuity technical and other risks.

- **Standards**: standards and requirements for digital continuity tools and services to ensure that they are fit for purpose.

8. What next?

We hope this introduction has given you a good background on why you need to assess your digital continuity risks, and take action to address them. We'll post further digital continuity guidance, as we develop it on www.nationalarchives.gov.uk/digitalcontinuity

9. Case study: NASA Apollo missions to the moon

The Apollo 11, 12 and 14 missions of the late 1960s carried 'dust detectors' that were invented by Australian physicist Brian O'Brien. They gathered vital scientific data that was beamed back to earth and recorded onto 173 tapes held at NASA, with back-up tapes at Sydney University.

Dr O'Brien's preliminary findings failed to spark as much interest from the scientific community as he was hoping for, and his tapes were sent to storage.

It's only recently that their value has been recognised. NASA hopes to return to the moon, and even eventually build a base there. So they need as much information as possible on the impact of moon dust.

However, they were unable to use Dr O'Brien's unique scientific data, because they'd failed to archive the tapes.

"These were the only active measurements of moon dust made during the Apollo missions, and no-one thought it was important," Dr O'Brien says.

Fortunately, Sydney University still had the back up tapes, but there was no real way to unlock the data they contained.

Dr O'Brien contacted Guy Holmes from data recovery company SpectrumData, who offered to help.

Mr Holmes kept the tapes in a climate controlled room, but it was only when he stumbled upon a 1960s IBM729 Mark 5 tape drive at the Australian Computer Museum Society that he was able to unlock the information.

Mr Holmes was concerned about the condition of the recorder which was complex, dirty and needed parts from a 1975 Toyota to get it going again.

Mr Holmes is hopeful of getting the tape recorder working again early in 2009. It should then only take a week to extract information that has been locked away since the early 1970s.