The archivist in the electronic age

Anita Hollier ([*](#))

Abstract

The article looks at the changing role of the archivist in the electronic age. It considers the new challenges posed, and some of the ways in which they are being met. Physical threats to records include technology obsolescence and physical deterioration of media. But a more fundamental issue that archivists and records managers have had to face is to reconsider what it is that constitutes a genuine 'record'. Only then can policies be put in place to ensure that they are preserved for as long as   they   are   needed.

Introduction

Over the last 50 - 100 years technological advances have changed the way we handle data to such an extent that we can now speak of a real 'information revolution'. Simple and efficient methods now exist to create, retrieve and share information. However, there are some corresponding headaches for those in professions dedicated to the preservation of records. As Jeff Rothenberg famously said "digital information lasts forever - or five years, whichever comes first" [1]. This article aims to summarise some of these issues, and to consider the ways in which archivists and records managers are having to rethink their profession in order to meet the new challenges. While these changes are not painless, they bring many benefits in new ways of working and closer relationships with colleagues in related professions. The way forward in this field is certainly through international and cross-disciplinary cooperation, and a number of models are already emerging.

The challenge

Preserving access to the record

Increasingly librarians and archivists are finding their roles converging as both take on responsibility for the long-term preservation of digital materials. So although I will speak mainly about archives it is fitting to begin by referring to a study commissioned by the Research Libraries Group (RLG), which stated "digital materials, regardless of whether they are created initially in digital form or converted to digital form, are threatened by technology obsolescence and physical deterioration.[2]

" The first of these threats arises from the fact that electronic records, unlike their counterparts on paper, are not directly readable by humans. They are machine-dependent, and if the hardware and software required to read them is no longer available they cannot be accessed. The classic example is that of the 1960 US Census records. In 1976 the National Archives identified some of this data, which the Census Bureau had retained in what it regarded as 'permanent' storage, as having long-term historical value. However, by that time the UNIVAC type II-A tape drive needed to read the tapes was long obsolete. Often some sort of data rescue can be achieved (as was finally the case for most of the 1960 Census records), but it may be very expensive and there can be no guarantee that all the information will be recovered. The media themselves (tapes, disks, etc) are inherently less stable than traditional materials. Whilst paper or parchment can survive 'benign neglect' for many years, magnetic media deteriorate more quickly and need to be regularly checked and refreshed. Electronic documents are also vulnerable to accidental damage; they can be accidentally deleted in a keystroke.

Preserving a genuine record

The International Council on Archives defines a record as "a specific piece of recorded information generated, collected or received in the initiation, conduct or completion of an activity, and that comprises sufficient content, context and structure to provide proof or evidence of that activity.[3]"This rather long definition was developed precisely because of the new challenges posed by electronic media. With the advent of computers it is very easy to create and store information and documents that do not have the characteristics of a genuine 'record'. In this definition, the content refers to the actual data, words, numbers, sounds, etc, produced by the creator of the record. The structure refers to the relationships between these data, for example which words represent a heading, or an address, or the body of the text. It also includes the links between different files that may make up one record, for example in a relational database. The context provides information about the activity and circumstances which gave rise to the record, who produced it, why it was produced, how it was used, how it relates to other records, etc. Of these, the context is often the most difficult to preserve in an electronic document.

Recognition of the need for contextual information has long been at the heart of archival principles. Most published items, such as library books, contain a certain amount of explanatory information as well as their main content, for example there will be the name of the author, a publication date, and probably some sort of general introduction which gives some idea of why the book was produced and for what sort of audience it is intended. An individual archival record usually lacks this information; it is simply a piece of recorded information that was produced in the course of some activity. Hopefully an office memo made sense in the context in which it was created - the creator and other users understood what it was about and what role it played in their business. But how much sense is it going to make, by itself, to a historian consulting it in an archive 100 years, or even 10 years, later? For this reason archivists try to catalogue each record thoroughly, giving particular importance to its provenance. For this reason also, the original order of records is respected when they enter an archive. The way the records were arranged during their active life (e.g. the office filing system) may not be the most convenient way of arranging the data for future historians but it does give invaluable information on the context in which the different items were created and how they relate to each other. This would be lost if the records were rearranged.

So, for paper records, preserving record systems can provide valuable contextual information that will help the records to be interpreted in the future. But one of the great advantages of electronic documents is that they do not need to be stored in one physical place. They are easily shared between work groups, and virtual documents can be created 'on the fly' to satisfy each individual's

requirements. This means that the sort of contextual information that was intrinsic to most paper records is often lacking in the electronic environment. If it is not actively provided, for example by the provision of adequate metadata, the documents can be difficult to use for current business and may lose their value as archival records.

Once a valid record has been created it must be preserved as a valid record for as long as it is required, and here the issue of authenticity is crucial. Another great advantage of electronic records for business purposes is that they can easily be edited and updated, but it is important to keep track of these changes in order to guarantee that the record really is the genuine article. It is very easy to alter electronic documents without anyone being aware of the fact. Of course, paper records can be forged too, but this is harder to do and easier to detect. This concern is clearly shown in the uncertainly surrounding the legal admissibility of electronic records, but in fact this is only one instance of the general rule that a genuine record must provide valid evidence.

Meeting   the   challenge

• The most obvious challenge of recordkeeping in the electronic environment is the technical one; it was the affair of the unreadable 1960 Census records that led the Committee on the Records of Government to say "the United States is in danger of losing its memory."[4]But technical solutions alone are not sufficient; archivists and records managers have also been obliged to rethink some of their assumptions about what constitutes the 'record' that they are trying to preserve. This provides the basis for policies which will ensure that correct procedures are followed and the appropriate technical   solutions   correctly   applied.

Nature of the record

A number of projects have addressed the issue of what constitutes a record. In 1993 the University of Pittsburgh School of Information Sciences began a research project to examine variables that affect the integration of recordkeeping requirements in electronic information systems. [5] The major objectives of this research project were to develop a set of well-defined recordkeeping functional requirements, satisfying all the various legal, administrative and other needs of a particular organization, which can be used in the design and implementation of electronic information systems. In September 1996 the project published a Framework for Business Acceptable Communications, including Functional Requirements for Evidence in Recordkeeping, and Metadata Specifications based on them. Its specifications have been adopted as the basis for a number of archival electronic recordkeeping implementations, for example at Indiana University.[6]

At around the same time, a research project by the University of British Columbia's School of Library, Archival and Information Studies aimed to identify and define the requirements for creating, handling and preserving reliable and authentic electronic records.[7] The UBC Project researchers worked in close collaboration with the U.S. Department of Defense Records Management Task Force to identify requirements for Records Management Applications (RMA). The resulting standard (DOD 5015.2) [8] is now in use by the U.S. Defense Information Systems Agency to certify RMA vendors. InterPARES (International Research on Permanent Authentic Records in Electronic Systems) is the second phase of the UBC Project, intended to address the long-term preservation of inactive electronic records (i.e. records which are no longer needed for day-to-day business but which must be preserved for operational, legal, or historical reasons)[9] .

Technical issues

This is intended only as a very brief overview. More detailed information is available elsewhere [10]; in particular, guidelines for storage and physical conservation are well covered [11]. There are three main technical strategies to address digital preservation, and each has its supporters and critics and its own advantages and disadvantages (Jean Marie Deken gives a good introduction to this and other matters concerning archives and records management in the electronic environment) [12].

- Technology preservation - preserve the original software (and the operating system and hardware on which to run it) that was used to create the record.

- Technology emulation - program future computer systems to emulate obsolete computer platforms and operating systems.

- Information migration - ensure that the digital information is re-encoded in new formats before the old format becomes obsolete.

Encapsulation is a technique that involves grouping a digital object (e.g. the record) with anything else necessary to provide access to that object and containing them within physical or logical structures called 'containers' or 'wrappers'. It is useful both in implementing emulation and also in attaching metadata to a record. In order to demonstrate that none of the information has been altered the encapsulated 'package' may be secured by digital signatures (see, for example, the Victorian Electronic Records Strategy, described below).

Policy issues

Since good information management is fundamental to all organisations it is important for them to ensure that their corporate policy is clear and well communicated, and that there is adequate funding and expertise available to implement it. Where many organisations are facing the same challenges it makes sense for them to try to work together, and this is particularly true in the archive profession where the individual units can be quite small. In the UK, funding from Resource (the recently formed Council for Museums, Archives and Libraries) has supported the production of the Preservation Management of Digital Materials Workbook [13]. This is currently available as a pre-publication draft, and aims to identify good practice in creating, managing and preserving digital materials, and to provide practical tools to assist in that process. Amongst other things it describes the sort of issues that should be addressed in formal corporate policies, such as the status and definition of the electronic record, and relevant roles and responsibilities within the organisation. It then treats corporate strategies in more detail:

"The following issues should be addressed in corporate strategies and may well require other supporting documents setting out in more details how the strategies can be achieved:

- Authenticity - What organisational and technical strategies will ensure that the electronic record is reliable and legally admissible?

- Appraisal and retention periods - How will they be developed and applied?

- Migrating records worthy of permanent preservation to new systems - How will they remain accessible and usable for as long as they are needed?

- Selection of media and formats - What standards should be used for various categories of electronic documents across the organisation?

- Metadata - How will essential metadata be defined and how can it be ensured that it will remain linked to the corporate record?

- Training and awareness raising for staff - What ongoing training requirements and supporting guidelines are required to support good records management practice?

- Compliance with the corporate policy - How will compliance with the corporate policy be monitored?"

Some codes of practice have appeared regarding the legal admissibility of electronic records[14]. Other important legal issues to be addressed relate to relevant national or international laws on freedom of information (access) and data protection (privacy), and copyright.

Changing approaches

What's different now?

From a preservation point of view a major difference between paper and digital documents is the time scale in which decisions and action must be taken. Actions and decisions that used to be taken towards the end of a document's life, where the choice was between 'destroy' or 'send to the archive', now typically need to be taken at the moment when the document is created so that it can be actively managed throughout its life-cycle. Indeed the whole concept of a 'life-cycle' has been called into question as inappropriate to the electronic environment, where records move into and out of active use much more often [15]. Likewise it is no longer clear that the best way to preserve a record is for it to enter the custody of the archive. Sometimes it may be more appropriate for the archive to have 'intellectual' control of the record, but for it to be physically located and preserved elsewhere [16].

Another change is that many more stakeholders are now involved in the long-term preservation of a record. The creators of the documents, the IT specialists who manage them, as well as archivists, records managers and librarians all have a role to play and their immediate concerns and priorities may be very different. Further confusion may arise from differences in terminology which can make it difficult for them to cooperate effectively; for example, words such as 'record' and 'archive' have clear but quite different meanings to archivists and IT specialists. Some misunderstandings are more subtle, for example 'metadata' is similarly understood by both, but tends to have a broader meaning to archivists than in the IT world where the emphasis is more on automatically captured data to aid resource discovery. It is worthwhile for collaborations to spend some time at the outset establishing a common vocabulary.

What has to change?

The most important step to meet these changes is for archivists and records managers to get more closely involved with the other professionals involved in electronic recordkeeping. One way to facilitate the involvement of all relevant sectors is for an organisation to establish a cross-disciplinary committee to develop its electronic recordkeeping policies and objectives. This committee can then allocate implementation tasks wherever the relevant skills exist. This

cooperation can function in-house, between organisations or internationally (e.g. the multidisciplinary DLM Forum organised by the European Commission) [17] .

Such cooperation will also allow archivists and records managers to be involved at the point of creation of the record (which has long been their aim in theory, but too seldom put into practice) rather than the point of destruction. Records appraisal now needs to be built into recordkeeping systems at the design stage. This goes hand in hand with a move towards functional appraisal, which uses tools such as business process modelling which were previously more familiar to systems designers than to archivists. This method makes explicit the business processes and information flow in the unit, then identifies the points at which a record (i.e. evidence) is required. The National Archives of Australia have adopted this as the basis of good recordkeeping [18]. The traditional records management approach has been to conduct a record survey to see what documents exist then to draw up retention schedules to specify for how long each of these series of records must be kept. The schedules are usually organised by record series and often a different one is produced for each organisational unit. This means that they quickly become out of date when any of these factors change, so some records managers have long been advocating functional appraisal [19]. The new type of retention schedule for electronic records will reflect a more 'top down' approach and will contain mainly policy decisions about what activities and transactions the organisation needs to keep evidence of. These requirements will normally remain valid for some time. They will be supported by implementation guidelines, i.e. the detail of exactly how this will be achieved, and these will change as technology or organisational structure changes. The records inventory is still useful in drawing up these guidelines, and acts as a reality check on the business process model.

Archivists and records managers are never going to be able to drive business processes, but they are able to contribute to process improvement. Information is a valuable resource and any organisation will benefit from using it more efficiently. Conversely, if valid records are not properly managed, not only will an archival resource be lost to future historians, but the organisation also faces a number of risks including failure to produce evidence that it has fulfilled its obligations, loss of proof of ownership or rights, inability to find information required for current activities, loss of efficiency if this information is difficult to find, and potential high costs of 'rescuing' it if it is not in a condition to be accessed directly. A good information strategy will ensure that when information is needed the user has access to a single authorised source that is easy to find and is maintained for as long as it is needed, but no longer. If access to the information needs to be restricted, appropriate controls can be put in place, but otherwise the whole organisation will benefit from sharing and exploiting it efficiently. This lies at the heart of the Knowledge Management that is currently in vogue, and also has an important role to play in Quality Assurance procedures. It is not only big business that is taking this on board, for example, the Joint Information Systems Committee (JISC) promotes the innovative application and use of information systems and information technology in higher and further education across the UK [20].

Best practice

A number of projects are underway which address various aspects of the electronic information challenge. There have been efforts to agree acceptable practices in the creation of digital material, and these guidelines will hopefully become even more widely observed. These usually include the observance of appropriate standards, the use of open, non-proprietary data formats, provision of documentation and metadata in accordance with emerging standards, and assigning permanent names to online digital resources. There is not yet any one accepted solution to long-term preservation, probably there never will be given the rapidly changing environment and the differing needs of different organisations, but some common approaches are emerging. One example is the

Open Archival Information System (OAIS) model described below. Of the various practical implementations of electronic recordkeeping I have chosen to describe briefly that of the Public Record Office Victoria, Australia. It is not necessarily typical of models followed elsewhere, but it is a very good, practical working example.

Open Archival Information System OAIS [21]

An initiative by NASA's Consultative Committee for Space Data Systems has led to the development of a Reference Model for an Open Archival Information System (OAIS) that is now being reviewed as a draft International Standardization Organization (ISO) standard and is expected to become a fully fledged standard in future [22] . It addresses the issue of what constitutes a genuine archive in the electronic era (as opposed to the sort of 'archiving' that really equates to moving data offline). But it is not limited to traditional archives, indeed it has more following in the library community where it has influenced (or been adopted by) a number of digital initiatives such as CEDARS, PANDORA and NEDLIB. By their definition an OAIS is an archive (consisting of an organisation of people and systems) that has accepted the responsibility to preserve information and make it available for a designated community. It lists the mandatory responsibilities of an OAIS archive as follows, it must:

- "Negotiate and accept appropriate information from information Producers.

- Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.

- Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and therefore should be able to understand the information provided.

- Ensure the information to be preserved is independently understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.

- Follow documented policies and procedures which ensures the information is preserved against all reasonable contingencies and enables the information to be disseminated as authenticated copies of the original or as traceable to the original.

- Make the preserved information available to the Designated Community."

These terms are fully explained in the model, as is the basic concept which involves packaging the 'data object' with enough supporting information to make it understandable in the long-term to a designated user community. It specifies the types of information that should be captured and gives a functional model describing five functions of an archive. An observation by a couple of the participating bodies who were invited to comment on the draft is that it needs a preservation function as well as its archival storage function, and that it currently neglects this area and focuses more on access. But the model is an excellent conceptual framework that can (and already does) serve as a meeting place for the diverse community of institutions involved in long-term digital preservation, many of whom may never have thought of themselves as 'archives'.

Victorian Electronic Records Strategy (VERS) [23]

The report, *Keeping Electronic Records Forever* commissioned by Public Record Office Victoria (PROV) in 1995-96, advocates that instead of taking a system oriented approach to electronic records, a data driven approach is more appropriate as the records will outlast any system developed to manage them. In order to achieve this the VERS project team first examined how records were created within Government and how they were used and archived, and made this explicit via process maps. They also looked at developments elsewhere in the world. They concluded that an electronic record had to be a fully self-documenting object, and chose to describe these objects in XML. They also determined that an electronic record was made up of one or more documents, contextual information relating this record with other records, and evidential integrity checks. They then built an archival system prototype and a retrieval system prototype, and created several sample working environments. They adopted a standard record format, the 'layered' (onion) model which encapsulates the documents, the context, and authentication in a single object, and they also adopted standard metadata sets (based on the Pittsburgh project model). The Final Report details the VERS project's findings and includes a general description of their prototype system, functional descriptions of electronic archiving, details of the VERS long term electronic record format, the metadata schema used by the project and costings of possible VERS compliant system implementations.

Conclusion

Electronic records are just records. Archivists have long agreed that they should not be accorded greater or lesser importance because of their format. However, their format does require a change of approach. The policy may not change but its implementation does. Established principles of archives and records management can form a sound basis for new approaches, for example by clarifying the essential requirements of a genuine record, but they should not be allowed to limit the view. Archivists urgently need to learn how to work effectively with their fellow professionals who now share the responsibility for the preservation of records. In future good recordkeeping needs to be built into systems, not applied retrospectively at the end of a record's life-cycle, and archivists and records managers alone are unlikely to achieve this. Positive action is needed, but sometimes we all need a little encouragement that it is achievable. Help is at hand! Our professional societies and training bodies offer guidelines. Most of the projects I have mentioned here have excellent Web pages, which include bibliographies citing many other related studies and projects [24]. Standards are emerging and many practical implementations are helping to establish and share best practice. The Victorian Electronic Records Strategy offers the following heartening words: "As a result of this project we have concluded that the:

- capture of electronic records into long term format, with much of the contextual information captured automatically, is possible and achievable now
- archiving of electronic records is possible and achievable now.[25] "

References

[1] Jeff Rothenberg, Senior Researcher, RAND Corporation. See also his article "Ensuring the Longevity of Digital Documents" Scientific American (January 1995)

[2] Margaret Hedstrom and Sheon Montgomery   Digital Preservation Needs and Requirements in RLG Member Institutions" (December 1998)
URL:<http://www.rlg.org/preserv/digpres.html>
For more general information, see URL:<http://www.rlg.org/longterm/>

[3] International Council on Archives. Guide For Managing Electronic Records From an Archival Perspective Committee on Electronic Records, ICA Studies/Etudes CIA 8, (February 1997)

[4] Committee on the Records of Government (1985) Report. Reprinted, Malabar, Florida: Robert E. Kieger
Publishing Company, Inc. (1988).

[5] URL:<http://www.lis.pitt.edu/~nhprc/>

[6] As part of a project to deal with its own electronic records, the Indiana University Electronic Records Project developed a methodology to evaluate information systems against the University of Pittsburgh Functional Requirements for Evidence in Recordkeeping. Several phases and tasks for the methodology have been identified. These include:
1) a description of the business through functional decomposition and identification of transactions and
associated evidence,
2) a description of the associated information system in terms of the identified transactions,
3) evaluation of the information system against the Functional Requirements in the context of the identified
transactions of the business, and
4) recommendations for intervention to satisfy the Functional Requirements.
URL:<http://www.indiana.edu/~libarche/index.html>

[7] URL:<http://www.interpares.org/UBCProject/ >

[8] Design Criteria Standard for Electronic Records Management Software Applications. DOD 5015.2-STD
(Department of Defense, USA). This standard is endorsed by NARA (National Archives and Records Administration)
and is the first example of a federal agency developing formal criteria for electronic records management.
In addition to the standard, a software test procedure has been developed along with a register of records management software applications products that have passed the test.
URL:<http://jitc.fhu.disa.mil/recmgt/index.htm>

[9] URL:<http://www.interpares.org/>

[10] See, for example: Preserving Access to Digital Information (PADI)
<http://www.nla.gov.au/padi/>

[11] Conservation OnLine (CoOL), produced by the Preservation Department of Stanford University Libraries, is a full text library of conservation information covering the conservation of library, archives and museum materials. It includes information on electronic media and records.
URL:<http://palimpsest.stanford.edu/>
See also Standards Published on Permanence of Electronic Materials
URL:<http://www.cd-info.com/CDIC/Industry/news/ansi.html>

[12] Jean Marie Deken "Electronic Recordkeeping: An Introduction" Invited talk presented at the US Department of Energy Records Management Conference (5/17/99 - 5/20/99)
URL:<http://www.slac.stanford.edu/~jmdeken/papers/SLAC-PUB8152.html>

[13] Neil Beagrie and Maggie Jones Preservation Management of Digital Materials Workbook: a pre-publication draft (October 2000)
URL:<http://www.jisc.ac.uk/dner/preservation/workbook/>

[14] For example, The British Standards Institute's Code of practice on legal admissibility and evidential weight of information stored electronically, BSI DISC PD0008:1999

[15] See, for example: G. O'Shea "Keeping electronic records: issues and strategies" Provenance: the electronic magazine, Vol 1, No. 2 (1996)
URL:<http://www.netpac.com/provenance/>

[16] Barbara Reed, "Appraisal and disposal", in Judith Ellis (ed.), Keeping Archives, 2nd edition (Port Melbourne, 1993)

[17] This is a multidisciplinary forum on the problems of the management, storage, conservation and retrieval of
machine-readable data, organised by the European Commission. It takes quite a high-level view to solving
problems, formulating a "DLM-message" to the Information and Communication Technology Industry,
and trying to influence funding of projects. But its 1996 conference also included some interesting case studies such as the description by David Bowen of Pfizer's Central Electronic Archive.
URL:<http://europa.eu.int/ISPO/dlm/>

[18] "Australian Standard AS 4390–1996, Records Management" provides a methodological framework for organisations to develop recordkeeping strategies that satisfy business needs, accountability requirements and community expectations. The Designing and Implementing Recordkeeping Systems (DIRKS) methodology is an eight-step process which agencies can use to design and implement AS 4390 compliant recordkeeping systems.
URL:<http://www.naa.gov.au/recordkeeping/dirks/summary.html>

[19] See, for example:
Jeff Morelli "Process-driven retention scheduling" Records Management Bulletin Issue 94 (1999);
G. O'Shea "Keeping electronic records: issues and strategies" Provenance: the electronic magazine, Vol 1, No. 2 (1996) URL:<http://www.netpac.com/provenance/>; and
Tom Ruller URL:<http://www.truller.com//>

[20] URL:<http://www.jisc.ac.uk/>

[21] OAIS Consultative Committee for Space Data Systems "Reference Model for an Open Archival Information System (OAIS)" CCSDS 650.0-R-l. Red Book (May 1999)
URL:<http://ftp.ccsds.org/ccsds/documents/pdf/CCSDS-650.0-R-1%20.pdf>

[22] For a good overview of this initiative, see: Brian Lavoie "Meeting the challenges of digital preservation: the OAIS reference model". OCLC Newsletter. No. 243. (January/February 2000)
URL:<http://www2.oclc.org/oclc/pdf/news243.pdf>

[23] URL:<http://www.prov.vic.gov.au/vers/welcome.htm>

[24] See also, URL:<http://www.records.nsw.gov.au/publicsector/erk/websites/websiteguide.htm>

[25] The Victorian Electronic Records Strategy Final Report (31 March 1999)
URL:<http://www.prov.vic.gov.au/vers/final.htm>

Author Details

Anita Hollier - CERN Archivist
CERN: http://www.cern.ch
Tel.: +41 22 767 49 53
Fax: +41 22 782 86 11
Address: CERN, CH 1211 Geneva 23, Switzerland

E-mail: http://library.cern.ch/HEPLW/3/papers/5/Anita.Hollier@cern.ch