DEFINING ELECTRONIC RECORDS, DOCUMENTS AND DATA

David Roberts

David Roberts is the Manager of the Records Management Office of New South Wales, one of the operational arms of the Archives Authority of New South Wales. From 1980 to 1993, he worked in a variety of positions with Australian Archives and convened the ACA/ASA Standing Committee on Machine-readable Records from 1986 to 1990. He currently represents the ASA on the Standards Australia Committee (‘IT/21’) preparing an Australian standard for records management.

Basic concepts have a profound effect on the approaches taken by archivists and records managers to their work. They are particularly important for the Records Management Office because our work depends on communicating with our clients. We have adopted a concept of records based on their role as evidence of business transactions, which is essential to support accountability in the New South Wales public sector and which leads to a particular set of approaches to electronic records management. We distinguish electronic documents from electronic records by the latter’s transactional origins and evidential qualities and explore the appropriate use of electronic document management tools within this conceptual framework as part of broader electronic records management strategies. We distinguish data management and data administration from the management of electronic records and electronic documents and recognise their role in electronic records management strategies.
The purpose of this article is to describe work being done in the Records Management Office of New South Wales to refine some of the key concepts involved in the management of electronic records. As one of the operational arms of the Archives Authority of New South Wales, the Records Management Office is taking part in the work of the Authority to develop a range of policy and procedural approaches to the management of electronic records in the New South Wales public sector. The focus in this article, however, is on the concepts of *records, documents* and *data* as they are used in the electronic environment. Inevitably, we will also need to look at these and related concepts in archives and records work generally.

Why is it necessary to examine these concepts? Surely archivists and records managers know what they mean when they talk about records, and everyone knows what documents and data are! However, there is ample evidence to indicate that there is little common understanding of these fundamental concepts in the Australian or international archives and records management communities. In addition to the obvious need to improve the theoretical basis of the work of our professions, there are specific reasons arising in the context of electronic records management for reviewing and refining our concepts.

Archivists and records managers are increasingly aware of the way in which the definitions which we use for our basic concepts affect how we approach our work. For example, if we accept that it is the evidential quality of records which distinguishes them from other kinds of recorded information, then the focus on capturing evidence which is a hallmark of the approach to electronic records management being explored by David Bearman and his colleagues makes a great deal of sense. If we regard records as merely an information resource, our approach to electronic records management will take a different path.

Moreover, records managers involved in electronic records management are increasingly recognising the need to augment tools developed for conventional records management with tools for managing data and electronic documents which have been developed entirely outside the records management industry. For example, document management software is growing in popularity as a means for individuals or organisations to manage their electronic documents: word processing documents, spreadsheets, databases and so on. If our concepts of *records* and *documents* are indistinguishable, as much of the traditional records management literature would have us believe, how do we explain the difference between document management and records management software or decide how to use these different kinds of tools in an integrated strategy for the management of electronic records?
Ensuring that these fundamental concepts are satisfactory and complementary is particularly important for the Records Management Office. In our consultancy, training, publications and standards work, we need to be able to communicate these concepts clearly to people who, for the most part, are understandably little interested in the subtleties of archives and records management theory, but who need to use a wide variety of tools to manage their conventional and electronic records. If we cannot get our thinking straight on these fundamental matters, we cannot expect our clients or readers to do so. In the Records Management Office, our concern in this regard has been partly driven by our involvement in the development of means of managing electronic records, in company with many of our colleagues in the archives and records management communities. It has also been driven by our own need for a fundamental review of many aspects of our doctrine, leading to such specific tasks as rewriting the glossaries which we use in our training and other materials.

Electronic records

For the Records Management Office, our review of what we mean by records and, consequently, electronic records started with such simple questions as the then names of our training courses. Was Effective Information Management really a course on information management or on records management? If the latter, why call it the former? The problem cropped up throughout our publications, training and promotional materials and in our consultancy work.

The work of the Standards Australia ‘IT/21’ Committee which is developing an Australian standard on records management has provided, among many benefits and challenges, an opportunity to examine what is meant by records management and, of necessity, by records in the Australian and international archives and records management communities. A study of definitions of records in the literature, prepared as part of the Committee’s preliminary work, showed that there is little agreement about what characterises a record. The Association of Records Managers and Administrators Inc. (ARMA International) defines a record as ‘recorded information, regardless of medium or characteristics, made or received by an organization that is useful in the operation of the organization’. Many of the American records management texts simply equate records and recorded information or follow the ARMA definition. Other definitions define records in terms of the physical formats in which they appear, along with the information which they contain.

A further group of definitions identified in the study give prominence to the role of records in providing evidence to document the transaction of business. This prominence is characteristic of the concept of the archival document, developed in Australia by Frank
Upward and Sue McKemmish and of the definition of a record developed in October 1993 by the Electronic Records Committee of the International Council on Archives. This emphasis provided the basis for the working definition of records adopted by the Committee:

For the purposes of the Standard, transactional records are defined as:

- recorded information
- in any form, including data in computer systems,
- created or received and maintained
- by an organisation or person
- in the transaction of business or the conduct of affairs
- and kept as evidence of such activity.

Records in this definition are described as transactional in reflection of the Committee’s terms of reference, emphasising that it is records in this sense which are the subject of the Committee’s work. The debate about the relationship between archives and records management and information management has been canvassed in many places elsewhere. The ‘IT/21’ Committee has recognised the tension between the ‘unitary’ and ‘pluralist’ views, to use Sue McKemmish’s distinction, without attempting the daunting task of resolving it or imposing a particular position on the issues of the relationship of records and information.

The evidential quality of records is particularly important for the Records Management Office. The crucial role of records in supporting accountability has been capably demonstrated in recent Australian writings. New South Wales has by no means been immune from the crisis of accountability described in these places. The Independent Commission Against Corruption’s examination of poor recordkeeping practices in Corrective Services is only one of a series of instances in this State, albeit the most spectacular, which reveal the nexus between records and accountability. It is particularly in this context that we have confirmed that the business of the Records Management Office is the management of records, rather than of information generally. We have adopted the ‘IT/21’ Committee’s working definition and use it as a vehicle for placing a strong emphasis on evidence, accountability and documenting the transaction of business in our consultancy and training work in the management of conventional and electronic records.

Accepting this notion of a record has a significant impact on what we understand as an electronic record. We can, of course, start with the observation that an electronic record is a record, however we understand it, in electronic form. Thus, for the ‘records equals information’ school, electronic records are ‘records which contain machine-readable, as opposed to human-readable, information’. Similarly, where the transactional and evidential qualities of records are accepted, electronic records can be described as ‘Recorded
information that is communicated and maintained by means of electronic equipment in the course of conducting a transaction'.

Much of the archival practice in the management of electronic records until the last few years, and of the literature describing this practice, has been concerned with managing databases as machine-readable or electronic records. Applying the transaction/evidence test to many databases indicates that they should not be regarded as records in this sense, operating as electronic information systems rather than as electronic recordkeeping systems, to use the neat distinction developed by Bearman and his colleagues. This can be expected to pose a dilemma for many archival institutions, which may accept the logic of the transaction/evidence argument — and indeed adopt the approaches to electronic records management which flow from it — but which have custody of, or legislative or jurisdictional responsibility for, databases and electronic information which do not function as records.

The problem cannot be avoided but perhaps it can be accommodated, not by diluting the concept of a record, but by recognising the informational value of such databases as information resources, not as records providing evidence of transactions. This means exploring and adopting the techniques which have been developed by the 'data archive' community, which better suit these kinds of materials.

In the Records Management Office, the importance of accountability in our jurisdiction, coupled with the role of records in providing documentary evidence to support accountability, forces us to take a strict line on what we mean by records and electronic records. As noted above, we have adopted the 'IT/21' working definition of a record; we discuss the concepts of records, information, transactions and evidence, and establish the distinction between records and information, in our courses; and we are exploring the approaches developed by Bearman and his colleagues for electronic records management as providing the potential for meeting important aspects of the accountability requirements of New South Wales public sector organisations.

Electronic documents
If we distinguish records from recorded information by their evidential quality and their role in documenting the transaction of business, where do documents fit into the picture? Is there any difference between uses of the term in the paper and electronic environments, or within and outside the archives and records management communities?

In the archives and records management literature, we can identify
two strands of meaning. Firstly a document is defined as recorded information, regardless of medium or form. In the context of the discussion above, then, a document may be distinguished from a record by the latter’s evidential quality in documenting the transaction of business. This distinction is most clearly drawn in the Keeping Archives definition of records, where records are recognised as a class of documents distinguished from other kinds of documents by their transactional origin and evidential qualities. Where the essentially evidential quality of a record is not accepted, that is, where records are simply equated with recorded information, the distinction between records and documents tends to disappear.

Secondly, a document is defined as a physical record item or unit, perhaps in terms of a specific physical format. This conforms more closely to the usage of the word in the world at large. It is also used in some records management software products, usually as an alternative level of control to the file, while other products use folio for the same purpose.

In computing, a document is a ‘named, structural unit of text that can be stored, retrieved, and exchanged among systems and users as a separate unit’. Each of these strands of meaning is evident when we look at electronic documents in recent records management literature. Richard Jones has described electronic documents as: ‘analognes of paper documents. In other words they are multi-media forms of information composed of text, numbers, images, figures: perhaps extended to include sound ... Electronic documents may be completely representable on paper, but this will not always be the case.” For the Australian Government’s Information Exchange Steering Committee (IESC) an electronic document is:

A collection of electronic data which may be produced by the creation of original data (typically a text document, small database, spreadsheet, graphic created within the electronic office environment) or by the combination of existing data (which may include data extracted from data files or databases). It should be managed as a unique entity by means of a standard set of descriptors.

We can identify several characteristics of electronic documents in these definitions and their implications. Firstly, an electronic document is discrete and identifiable from other electronic documents, including other versions of the document. This means that it can be managed as a unit. Secondly, it possesses a logical structure of relationships between the smaller collections of data — paragraphs, tables, cells, images, headers — which comprise the document. This distinguishes an electronic document from other kinds of data files. Thirdly it is increasingly likely to be a multimedia or compound document, perhaps incorporating sound or moving images. This
means that it is likely to be stored as more than one data file; it will require the use of more than one application program to use or present it, especially in object-oriented programming environments; and it is increasingly unlikely to be capable of satisfactory representation on paper. All these characteristics are shared by electronic documents and electronic records.

Prima facie, the IESC definition conforms to a model which sees electronic records as a class of electronic documents, distinguished from other kinds of electronic documents by their transactional origin and evidential qualities. However, the IESC Guidelines define records as ‘Papers, correspondence, forms, books, photographs, films, sound recordings, maps, drawings or other documents, regardless of physical form or characteristics, either originated or received by an agency’. Again we see a definition of a record based on a list of physical formats (which, unintentionally, one hopes, does not include data storage media like disks or tapes) and distinguished from documents, apparently, by being originated or received by an agency. Presumably document is used here in the physical sense which we noted earlier. Yet are not documents in this context also either originated or received by an agency? What difference is there between a document and a record on this basis?

Moreover, the Guidelines consistently refer to provisions of the Commonwealth’s Archives Act (1983) as relating to documents, when the Act itself talks of records, albeit in the sense defined by the Act. Electronic records appear nowhere in the Guidelines, despite the extensive use of the term by the Australian Archives. Clearly, then, the IESC either does not realise or does not accept the transaction/evidence distinction between records and documents. That it does not is not intended as a criticism of the Guidelines. Rather, the point is the effect of inconsistent terminology here and elsewhere. Much of the advice in the Guidelines is of great practical value, which we recommend to our course participants and clients as one of a number of sources for developing their electronic records management strategies. However, we need to make it clear that the practices advocated in the Guidelines are tools for the management of electronic documents which can be applied to electronic records as part of a broader strategy. The Guidelines do not, and were not intended to, provide answers to the problems of keeping electronic records which will serve as evidence of transactions.

Software tools for the management of electronic documents are an increasingly important part of the office environment, their popularity fuelled by the rapid growth of networks and distributed computing, work organisation methods based on work groups, document image processing, large data storage devices and a variety of related developments. As such they form part of the armoury available to
archivists and records managers for the management of electronic records, in addition to their value for the management of an organisation’s broader electronic information resources.

A growing number of electronic document management products are available off-the-shelf. How closely do they fit with the conceptual view of electronic documents developed above? In the world of IBM-compatible personal computers and networks, a product such as PageKeeper is distinguished by its strongly textual orientation: the basis of the program is text-retrieval software, analysing the syntax of each document to identify its relevance to a search, and supported by optical character recognition, data compression and network sharing. However valuable such a product is, it is worthwhile being aware that the concept of electronic document on which it is based is that of a discrete text-based document. As we have seen, this is only part of the story. A product such as Keyfile is closer to an enhanced (computer) file manager, but with a user interface based on a filing cabinet and folders and in and out trays. It is designed to manage more broadly defined electronic documents than the text-oriented products, and supports the Microsoft WAV format, which enables sound, such as dictated comments, to be embedded in documents. Similarly PC-DOCS Open produces a SQL database of information about electronic documents in any media. It combines retrieval tools for searching the database with text-retrieval software.24

Despite some limitations in their current scope, all three products are essentially consistent with the concept of an electronic document described above. This is fortunate for us, because it makes it easier to fit them as useful tools into a strategy for electronic records management founded on a sound conceptual basis and without expecting users to adopt different vocabularies for different tools.

In the Records Management Office, we now specifically distinguish electronic documents and electronic records by the transactional origins and evidential qualities of the latter. We encourage our training course participants and clients to investigate the appropriate use of electronic document management tools, whether software tools like those noted above or practices such as those described in the IESC Guidelines, as part of a broad electronic records management strategy.

Electronic data
If electronic records are a class of electronic documents possessing certain characteristics, then we may regard electronic documents as a class of electronic data files possessing certain characteristics. When we come to consider the management of data files25 in electronic information systems, we enter the world of data management and data administration.
By data management, we mean ‘the function of controlling the acquisition, analysis, storage, retrieval, and distribution of data’. Consequently, it can involve a large range of functions, such as protecting the physical security of data through adequate backup procedures and recovery arrangements, protection of confidentiality and privacy in data, establishing and enforcing users’ responsibility for data accuracy, reducing data redundancy or duplication, organising data in rational and consistent ways, and ensuring the retention of data for the required periods of time. Beyond these bread and butter functions, data management can involve promoting the use of common and consistent data across the organisation, establishing a corporate data architecture and promoting the design of electronic information systems which are consistent with the data architecture and strategic IT plans. Anyone who uses a personal computer regularly performs data management tasks, such as backing up our data files and rationalising our directory structures, whether we think of it in these terms or not. In organisations, data management is normally regarded as a part of the management of information technology resources, although it is not necessarily recognised organisationally with a specific position or unit in an IT branch responsible for the function.

By data administration, we mean ‘the corporate service which assists the provision of information systems by controlling and/or coordinating the definitions (format and characteristics) and usage of reliable and relevant data’. Thus it is a more specific discipline which primarily involves the management of metadata, using such tools as data dictionaries, Information Resource Directory Systems and a range of types of system documentation.

Data management and data administration are increasingly important for archivists and records managers for two reasons. Firstly, developments in the last few years have demonstrated that both functions have an essential role to play in approaches to the management of electronic records. Thus data management and data administration techniques appear among the methods advocated by David Bearman and his colleagues, while the data management principles identified in Australian Government agencies are harnessed and enhanced for the purposes of electronic records management by the Australian Archives. If archivists and records managers expect to exploit data management and data administration techniques and to forge alliances with IT personnel as part of their electronic records management strategies, it is essential to understand the nature and scope of these functions and their relationship with the management of electronic documents and electronic records.

Secondly, the growth in end user computing, with the proliferation of personal computers and small networks, coupled with the reduction or break-up of many organisations’ IT branches, has meant that there
are increasing numbers of computer systems without formal or practical data management arrangements. In these circumstances, records managers increasingly find themselves taking on a data management role by default or perhaps by design, because a work group or business unit's electronic information systems contain data files functioning as records.

Promoting and performing data management at a practical level is not new to archivists and records managers and useful advice is to be found in the archives and records management literature. However, we need to recognise that data management and data administration are disciplines in their own right, part of the larger discipline of computing. This is important for two reasons. Firstly it leads us to literature and other sources of advice about their practice which we will not find if we restrict ourselves to what is available in the archives and records management communities. Secondly it makes us realise our limitations in this area: few archivists or records managers can expect to develop the levels and range of expertise or responsibilities expected of a professional data manager or administrator, and few organisations would justify the cost of our developing such expertise. Thus it is a pity that Saffady's otherwise valuable practical advice does not place the data management practices which he advocates into context, show how the records manager's role in data management might relate to those of other people in the organisation with IT responsibilities, or recognise data management for what it is.

In our training and other work at the Records Management Office, we have found it useful to make a clear distinction between data management functions — ensuring that our course participants and clients understand the full meaning of the term — and the management of electronic documents and electronic records. This enables us to discuss options for a range of electronic records management strategies involving data management and data administration, ranging from teaching basic practical data management for individuals and work groups to the tactical use of data management and data administration as part of the design of electronic recordkeeping systems.

Conclusion
For the Records Management Office, refining our concepts of electronic records, documents and data seems a small thing (and indeed it is only one part of our work to review our doctrine) but we have found it to be an essential and stimulating exercise. It has helped us to impose some structure on the confusing world of information; to focus on what we need to achieve in electronic records management; to communicate the fundamentals of records management to our clients, laying a sound foundation for what follows; and even to define our very
business. The issues take us well beyond electronic records management and pervade more and more of our work.

ENDNOTES


4. Records Management Association of Australia leaflet, undated; Archives Act, 1983 (Commonwealth), section 3.


7. ‘A record is information created, collected or received in the initiation, conduct or completion of an institutional or personal activity. Records have the following requirements:

- provide evidence (which is a by-product)
- comprise content, context and structure
- have integrity and immutability
- are unique
- exist regardless of physical format
- lead to an outcome.

All these attributes exist with electronic records, but for these context is more important, to establish a relationship in a system of records.’

The definition is unpublished at the time of writing. The author understands that the Committee does not regard this definition as definitive and expects to refine it further. The author’s thanks are due to Steve Stuckey of Australian Archives for bringing this work to his attention.


9. Particularly in Sue McKemmish and Frank Upward, editors, Archival Documents: Providing Accountability Through Recordkeeping, Monash Occasional Papers in
Librarianship, Recordkeeping and Bibliography, No. 3. Ancora Press, Melbourne, 1993 and in articles on the accountability theme in the May 1993 issue of Archives and Manuscripts.


11. Our governing legislation, the Archives Act, 1960 (NSW), has a definition of a public record which does not, of course, reflect these recent developments. At the time of writing, we are working towards a definition which takes account of them for inclusion in draft state records legislation which will replace the 1960 Act.


13. Charles Dollar, Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods, Odo Bucci, editor, Ancona, University of Macerata, 1992, p. 85. A variation is to describe electronic records as records, defined elsewhere in evidence/transaction terms, but in electronic form, as in Acland, op. cit., p. 469: ‘Records [i.e. defined as we noted above] capable of being processed in a computer system and/or stored at any instant in a medium which requires electronic or computer equipment to retrieve them’; or in State Archives of Western Australia, Electronic Records: an investigation into retention, storage and transfer options, LISWA Research Series No. 4. The Library and Information Service of Western Australia, 1993, p. 67: ‘Electronic record: Machine-readable information in electronic storage media judged to be an organisational record in accordance with the definition of what constitutes a record adopted as a records management policy’.


15. Acland, op. cit., p. 477: ‘Records: Documents containing data or information of any kind and in any form, created or received and accumulated by an organisation or person in the transaction of business or the conduct of affairs and subsequently kept as evidence of such activity through incorporation into the recordkeeping system of the organisation or person. Records are the information by-products of organisational activity.’

16. ARMA Glossary, p. 10: ‘Document: Recorded information regardless of medium or characteristics. Frequently used interchangeably with the word record...’ Compare with the Glossary’s definition of a record noted earlier in this article. Also Violet Thomas, Dexter Schubert and Jo Ann Lee, Records Management: Systems and Administration, John Wiley & Sons, New York, 1983, p. 393: ‘Record: Any document valuable enough to be retained’.


19. Richard Jones, ‘Directions in Intelligent Electronic Document Management’, Informaa Quarterly, vol. 9, no. 1, May 1993, p. 4. Jones is also careful to distinguish electronic documents from other kinds of multimedia: ‘A key distinguishing feature of electronic documents compared with general multimedia is that text is the predominant medium, acting metaphorically as the glue holding the other media forms together. This may be in the form of intra or inter-document references.’

21. Compare with word processing programs, where a document has traditionally been synonymous with a file.

22. ibid., p. 55.

23. Under Section 3, a record means 'a document (including any written or printed material) or object (including a sound recording, coded storage device, magnetic tape or disc, microform, photograph, film, map, plan or model or a painting or other pictorial or graphic work) that is, or has been, kept by reason of any information or matter that it contains or can be obtained from it or by reason of its connection with any event, person, circumstance or thing'.

24. This brief analysis is not intended to imply anything about the relative value of the products. It is only concerned with the scope of the concept of electronic documents underlying the products.

25. Most archivists and records managers are familiar with the exotic use in computing of such terms as archive, record and file, which presents another set of terminological challenges beyond the scope of this article.


28. See, for example, the use of data administration as part of a generic framework for managing electronic records in David Bearman's article 'Managing Electronic Mail' elsewhere in this issue of Archives and Manuscripts.


30. We should note that the distinction between data management and data administration is not necessarily well recognised in the computing industry or in its literature, with the border between the two often ill defined and with significant areas of overlap when the two functions are described. Thus a useful recent text covering the full range of data management functions uses data administration in its title (Brian Horrocks and Judy Moss, Practical Data Administration, Prentice Hall, Hemel Hempstead, 1993). However, the distinction is important for archivists and records managers because both provide useful, though different, tools for electronic records management. Again we need to impose a degree of terminological precision which does not necessarily exist au naturel.