

SECTION III

Design and Implementation



CHAPTER FIVE

Archival Principles and the Electronic Office



CHAPTER SIX

Managing Electronic Mail

CHAPTER FIVE

Archival Principles and the Electronic Office*

Electronic information systems are capable of capturing substantial amounts of information about the provenance of records they contain. These systems can track specific functions, activities, and transactions and which individuals contributed to or modified each item. As currently implemented, however, records are often stripped of contextual information, thereby greatly reducing their value as evidence. Contextual information is most often lost during data migration as systems are upgraded or data is communicated from one system to another. A cost/benefit plateau determines the degree of contextual evidence that will be retained. Archivists must articulate what evidential detail is required or desired for each transaction. The author examines the specifics of contextual data capture and retention in the principal software applications: word processing, electronic mail, spreadsheets, database management, and graphics.

* Originally published in *Information Handling in Offices and Archives*, ed. Angelika Menne-Haritz (New York: K.G. Saur, 1993), 177-193, which contains papers presented at the Symposium on the Impact of Information Technologies on Information Handling in Offices and Archives in Gladenbach, near Marburg, Germany, in October 1991. An earlier version of this paper was presented at the National Archives of Canada, 26 February 1991. The author wishes to thank Terry Cook, Richard Cox, John McDonald, and Lisa Weber for their thoughtful critiques which contributed to improving the presentation.

ARCHIVES AS EVIDENCE

Responsible corporate management of electronic records, whether for ongoing operational purposes or for long-term retention of corporate memory,¹ depends upon an understanding of the fundamental archival principle of provenance. This principle is central to the concept of archives as evidence of activity and pertains with equal relevance to all forms of documentation. Two derivative archival methods, "respect des fonds" and "respect for original order" which are also often referred to as principles, are in fact only implementations of the principle of provenance which reflect the nature of record-keeping in bureaucratic organizations during much of the era of paper records. Ironically, electronic records systems make it both possible to more fully capture provenance than paper records systems did and, at the same time, make it more likely that provenance will be lost and that archives, even if they are preserved, will therefore lack evidential value.² This chapter explores the relationship between provenance and evidence and its implications for management of paper or electronic information systems.

Archives are recorded transactions created in the course of organizational activities that have continuing evidential value. The criteria which distinguish archives from all of the information ever created or received in an organization are that:

- (1) archives are records of transactions;
- (2) archives document activities or functions reflected in the mission of the organization, not just incidental to it; and
- (3) archives are retained for their continuing value as evidence.³

These three criteria contribute to an implementation guideline which has been central to records management and archival practice but is rarely made explicit: because the

meaning of archives derives from the context in which they were created, and their evidential value is determined by the degree to which that context is still discernible, records management seeks to capture, and archives management to preserve, recorded transactions, their original form, and information about the historical nexus between creation and use.

Archivists select records for their "evidential historicity." Evidential historicity is the sum of all information that can be determined about an accountable transaction, which is defined as the relationship between a record and an activity determined by archivists to require evidence. The information which contributes to evidential historicity is derived from analyzing the data, the structure, and the context of records, each of which testifies explicitly and implicitly.

- The **data** of the record are the words, numbers, images, and sounds actually made by the creator of the record.

- The **structure** of the record is the relationships among these data as employed by the record creator to convey meaning. One kind of structure is the stylistic formalisms which we use to recognize the "address," "salutation," or "body" of written documents. Another kind of structure is the pointers between physically or logically distinct groupings of information as is the case with forms or databases where one aggregation of data elements is related to another aggregation in a separate record, but kept together in the same case file or in a "relation" in the database definition. Often structural information will be both indicated structurally and recorded explicitly in the data content of a record, as it is in the standard memorandum with the headings "To:," "From:," and "Re:."

- The **context** of the record is the testimony it provides about the nexus of activity out of which it arose and in which it was used and about how it appeared and behaved in that setting. Archivists recognize that a body of records that has lost its provenancial link has little or no value as evidence, but they have not analyzed the sources of information that reveal the relationship between the record and the activity out of

which it arose. The most unimpeachable contextual information from which records derive their evidential value resides in the record system, not the individual records. Although this is the acknowledged reason why archivists retain record systems (which in paper-based systems means "original order"), there is little analysis of how to read the evidence of records systems in the archival literature. As with structural information, it is also possible for contextual information to be carried as data in the record, as it commonly is in the case of dates on correspondence or reference codes in registry systems.

It is important to note that, when information which purports to provide structure or context is carried as data, it can be purposefully or accidentally misleading. The date on a document (data) may not be the date the communication was written or sent (context), and the distribution list may or may not be the same as the people who actually received a memorandum. Authors are free to invent data that purports to be context, and may even do so after the fact.

EVIDENCE AND BUSINESS APPLICATION REQUIREMENTS

Businesses maintain record systems to meet the requirements of ongoing operations for evidence. Archivists retain those records required for ongoing accountability. Archivists recognize the importance of contextual information to the quality of evidence in their appraisal practices when they consider the "evidential value" of the information in deciding about its disposition in their practices of arrangement and description. Archivists reflect the belief that they can say significantly less about the meaning of records that are no longer in their "original order" than of those whose connection to the organization and activity which generated them. But archival theory has not articulated the link between the information conveyed by original order, which suggests how the records were used in the course of business, as a contextual information requirement of evidence even though courts have set such

standards. For example, archivists and records managers know that only the systematic microfilming of records in the normal course of business, in which each record retains its contextual relationship to other records received and sent the same day, can assure the admissibility of a microfilm surrogate as evidence in a court.⁴ Similar structural and contextual information requirements for evidential historicity are present also in all records systems. Here we argue that they can be made explicit and that doing so is essential to defining the requirements for records management.

Different business applications will have differing requirements for evidence. The character, or degree, of requirements for creation and management of contextual information is directly related to the particular functions of each business. Once created to carry out a transaction, records are managed by organizations according to procedures dictated by ongoing needs of the application. In paper-based records systems the primary need of an office can be implied from their filing sequence or "original order" of its records; secondary needs can be implied from the existence of indexes to those records. Thus contextual evidence tells us whether the critical organizing principles of daily work were time (if so we will find chronological files), responsibility (where we find project files), client (case files), or intellectual context (subject files). It is instructive to examine both the actual filing sequences required by given functions and the practices of recordkeeping surrounding some functions in which a greater degree of contextual information is kept in order to understand better the importance of contextual data to evidence. This will also help us to appreciate why contextual data are of such great importance to electronic records management.

In an office conducting a single routine function, paper records are often filed according to a single salient feature such as the license number, plot identifier, or date of application. "Line" offices, which are responsible for carrying out a single function of the organization, will have procedures that

assure the capture and retention of necessary evidence of their activity. If the service they perform is based on priority, for example, the date on which an application was submitted or a fee was paid will very likely also be explicitly maintained by the filing system. In such an office, incoming material will be subject to a transaction (receipt) which involves stamping the documents with the date of receipt in the office performing the application review function. Subsequent steps in the procedure will be subject to similar documentation, which could be recorded or stamped on the original document, or on the folder containing the case, or on an accompanying transaction slip. The consequences of transactions may be recorded by filing the documents in a distinctive file based on the outcome (approved, rejected, continuing under review, etc.) or such information may be reflected in data on the transaction slip and all records interfiled. In this hypothetical case, we can see the accumulation of data from a number of transactions (the original letter of application, each entry acquired along the way), along with structural information (the linking of an application to a license number), and contextual information (the form of the record revealing its provenance, the stamping on the record or file revealing the time of its receipt, and the folder in which the document is found) suggesting, but not explicitly revealing, the activity locus and its relationship to other documents with which it is filed.

Since the late nineteenth century, client-oriented functions have maintained case files for each client. These will typically contain both information about the client and information about the handling of the clients' case by the organization. Much of the latter, contextual, information about handling the case may take the form of dates and initials of various people who collected information, evaluated it, and made recommendations, approved courses of action, and communicated actions to clients. It may also include notes indicating who has viewed the folder. These contextual data are not the text of documents in the folder, but may be marked on the folder it-

self or attached to documents in a variety of routing or transaction slips. They are reflections, in a "paper trail," of the procedures employed by the agency, and hence provide evidence of what procedures were followed.

For example, in the Office of the President of the United States, dossiers containing many briefing papers were routinely marked with their drafting history, the history of their review and approval, and indications of their dissemination.⁵ In such application environments, we must presume that the requirements of the function dictated that detailed contextual information about the use of records be maintained. The fact that we do not keep every draft of paper documents or every copy of a document sent for information in most offices reflects the fact that organizational and economic costs of implementing such a degree of contextual data capture exceeds the business requirement for it.

RECORDS SYSTEMS AND EVIDENTIAL HISTORICITY

Evidential historicity is captured in records systems which reflect the social and mechanical technology of the time. It is then retained and retrieved by archivists whose selection of records shapes the archives. Explicit recording of contextual evidence is rare in paper-based environments because it is costly, inefficient, and, given the amount of implicit evidence provided by paper-based systems, often unnecessary.

A cost/benefits plateau determines the degree of contextual evidence that will be retained. The contextual data capture plateau in paper-based environments typically excludes data on the intermediate transformations of documents; the bulk of paper that would be required to retain "version control," and the lack of means to link changes made within individual drafts to the individuals or offices making them, have generally prevented the long-term retention of drafts of all but the most important of documents. A different equation operates in electronic records environments where a multiplicity of

versions will require incrementally more storage if the software environment stores only changes as it creates subsequent drafts.

In paper-based systems, evidence may be explicitly recorded for accountability as a matter of regular business practice, but still have very limited utility due to the limitations of retrieval. Such information is often retained only until an audit is performed. For example, in the licensing application discussed earlier, dates of application or final approval may be retrievable while intervening procedural states will probably not be. Thus, in this system, the question of who worked on a given case may be answered by the record system (retrieve by case, examine initials), but the question of what cases an individual member of the staff worked on in a given week may not. As a result, archivists may decide to discard contextual data about transaction handling history which are recorded on the file folders when refoldering the files themselves to better preserve their contents for archival retention.

Archival cost/benefits are not isolated, but rather reflect the cost/benefits of the implementation context. In the examples of version control and work analysis, an important consideration in guiding archival retention of evidence is the degree to which it was available, and used, in the course of the work of those who created the records. In the paper-based environment such data was also not used regularly in the course of decision making while in an electronic office it might well be.

But moving from paper-based to electronic systems is not a one-way street towards greater ease in using and retaining contextual data. It is often not necessary to explicitly record much contextual evidence in paper-based record systems because they provide such a wealth of evidential historicity from implicit clues. Scholars working on pre-twentieth century records are frequently able to distinguish entries made by different hands and estimate the date by the ink used. In twentieth century paper-based case files, the context of origin of

records can often be inferred from the forms provided by different offices or functions. No one consciously designed the nineteenth century ledger to take advantage of the changes in handwriting and ink that serve the scholar in the construction of evidence, but these are nonetheless features of its design. The evidential historicity of these records derives from the immutability of the structural connection between entries on a page and the identifiability of handwriting and ink. Probably the handwriting and ink were used by the makers of the ledger, just as they are by the scholar, to retrieve records entered by a particular person. Although such evidential historicity could be captured in electronic systems where the invisibility of the evidence prevents us distinguishing the writer's hand or ink, and tools may be developed to allow authorship to be used as an attribute for retrieval, it is unlikely to happen unless we are conscious of the evidential historicity that will be lost without such design intervention.

Whether explicitly recorded or not, evidential historicity may be subject to loss if the record system is disrupted. In the case of the nineteenth century ledger, a decision to microfilm these records might result in losing the identifiability of the ink. In the case files of license applications, refolding will dispose of much of the contextual data now recorded on the folder itself, while almost any unsupervised use of the materials will disrupt the order of the documents which were originally filed according to a consistent filing procedure (such as "last to front"). In the case of electronic systems, the meaning of the pointer between a transaction record from 1980 and a client record updated in 1982 is obscured when both are written to a flat file for archival retention in 1985.

The costs of explicitly recording evidential historicity, the limitations in paper-based systems on retrieving information based on any recorded data, and the "eye-readability" of the residual evidence of activity in traditional office settings have collectively minimized the attention given to evidential documentation in the design of paper-based records systems. Be-

cause paper-based records systems lack methods to automatically capture, easily keep, or systematically analyze evidential information, business requirements have accepted a low plateau of evidential data capture, retention, and access.

Electronic records systems do not have these limitations. They provide the opportunity to capture larger amounts of evidential historicity, and they create a need to do so. Greater amounts of contextual and structural information must be retained from electronic systems in order to assure that electronic records have evidence of equivalent value to that which was obtained from paper systems. In part, this reflects a greater degree of need for contextual information when records are not eye-readable. If we cannot see the documents then we need to have an external description of them. Not surprisingly, such descriptions, or indexes to records, were most often prepared in paper-based systems only when the records themselves could not be directly inspected for reasons of remote storage, security, etc. But in the electronic systems that serve today's offices, we cannot inspect the records directly so we must rely on information about them which is captured and retained by the system. As a consequence we must ask what evidential historicity should be captured by an appropriately designed system and how it should be retained.

An electronic system can track what function, activity, or transaction, and what individual, contributed to or modified each item of information if only archivists articulated what evidential detail is required or desired regarding any given type of transaction. For example, in the electronic office we can easily track every modification made to a record during its drafting life, every recipient of an electronic message and when they read it, and even every time a document was consulted. However, the potential for automatically capturing large amounts of contextual information from electronic records systems is not simply a boon to archives. It is accompanied by a host of new problems, of which the most obvious is to determine what amount of evidential historicity is re-

quired and how to capture and retain it. Less obvious is the problem that much of the evidential historicity in electronic systems, just like the arrangement of records in a paper ledger or the use of different handwritings and inks, is hardware and software dependent but the systems in which it is implemented are less easily preserved than paper-based systems were. Retaining evidential historicity also increases the complexity of the information system design and the amount of data that must be retained and hence raises the cost of preservation both for storage and migration.⁶

To design records systems that meet organizational and archival needs for accountability requires that we understand not only that evidential historicity is found in data, structure, and context, and that its character and degree reflect business application requirements, but also how records systems capture this information and the threats to its preservation. In the emerging electronic office environments that serve as the vehicles for creating and communicating the recorded knowledge of organizations today, the same relationship between raw data, structural information, and contextual information can be identified.

The raw data of today's electronic office is the text of its documents. This text does not reliably tell us anything about the provenance of the information or the order in which it was kept or used. Like paper documents, electronic records may have data provided by their creators which are intended to tell us about context, such as the date of a letter, but which does not actually assure us that the letter was written on the date indicated or that it was sent. Indeed, in their ASCII form the data does not even tell us what the documents looked like to those who read them. This kind of structural information, corresponding to that provided by the form and filing of paper records in the past, comes from the documentation of the system in which the data reside. Contextual information about the actual use of data in the system and its communication to others (e.g., transactions) may or may not be kept by the soft-

ware application based on whether there is a technical requirement to do so in order for the software to function and whether there was a business functional requirement dictated by the client(s).

APPLICATION SOFTWARE AND INTERCHANGE STANDARDS

To provide concrete guidance on the capture and retention of evidential historicity in electronic records environments requires that we distinguish between the way in which evidential historicity resides in data, structure, and context information in a variety of different applications. Specifically, we need to identify that evidential historicity which is a minimum technical requirement to support each application and that which is the maximum observed evidential historicity implemented to support client functional requirements. This will enable us to develop strategies for capture and retention of evidential historicity and to identify that evidential information which could be interchanged between software systems using existing standards and that which is not covered by existing interchange standards and must therefore be represented in proprietary ways. This in turn enables us to estimate the cost and effort involved in achieving certain plateaus of evidential historicity within a range of applications, thereby making it possible for archivists to return the decision about how much to capture and retain to the organization where the risks and requirements for accountability can best be addressed.

Electronic records present one management challenge less frequently encountered in paper-based systems: retaining records for whatever period of time they have continuing value necessarily means moving them from one software environment to another.⁷ Because the data must be moved, and because their value as evidence depends upon their evidential historicity, the migration of records to future implementations including, but not limited, to archival implementations re-

quires that structural and contextual information be moved as well. Interchange standards which accommodate necessary evidential historicity are essential to the survival of archives.

Interchange standards for data and for structural information are critical for organizations conducting their business electronically. Such standards have therefore received significant attention from information managers and the software industry. The case for migrating contextual information is most evident to those who need to keep information for a very long time, and who are concerned for its evidential value. Unfortunately, archivists have not yet articulated clearly the requirements for evidential historicity associated with particular types of information systems applications. As a consequence, such requirements have been inadequately addressed in the definition of data interchange standards. At the present time, as the requirement for evidential historicity increases, the standards governing the representation of the data become less adequate for migrating records and the degree of proprietary information structuring is increasing.⁸ Archivists, records managers, other information managers, and, to a great extent, program managers dependent on electronic information systems must become aware of the gap between the evidential information available to electronic application systems and the information that can be removed from the proprietary environment and maintained over time.

The degree of difficulty associated with contextual data capture and retention, and the types of information that could be captured in addition to "raw data," differ from one application software system to another. The way to begin to understand the implications is to analyze some generic characteristics of a variety of software application environments. Ultimately, specific software applications and implementations will need to be studied in order to derive the data required to define concrete cost/benefit plateaus.⁹

Word Processing

The most common office application of electronic information systems is word processing. The word processing file as created by the user (e.g., the data of the record) consists of ASCII text interspersed with display instructions such as punctuation, type size and font, tabs, and paragraph breaks. If users save the file, they must typically enter the document name into the appropriate component of the header. The display instructions indicate the logical structure of the file using typographical conventions which are familiar to readers of paper documents. The header consists of contextual information which is mostly supplied by the software system including, for example, the date and time of creation, the file size, and location of the first physical data segment. In some software systems, especially in multi-user systems, the header might also include update history of the file, access rules, output format definitions, and links to distribution lists.

An office system need not maintain a great deal of evidential historicity to manage word processing files if the word processing system is simply a means of making paper output products. But from an information management perspective, if the system serves as a repository of textual records created and maintained in the course of business, it is essential to retain evidence of the structure and context of the data, including information about the creation and use of the record.

Contextual information must be retained for operational as well as archival reasons. In electronic networked office environments that are now becoming the norm, there is a technical requirement to preserve evidence of what changed between versions of a document. This responds both to the business requirement to be able to return to prior drafts and to the audit requirement to identify who was responsible for all changes. The system may also have a technical requirement to identify the authors of each change in order to support security (permission) controls. Finally, the client may impose functional requirements on the software designers to create and

retain information on the users and uses of documents. This kind of information is especially critical for records to serve as evidence in a shared filing system where many individuals have access to records and where "sending" the document in a transaction actually involves only making it accessible to another.

It is possible to design electronic office systems which record the evidential historicity associated with each transaction that creates, communicates, uses, or modifies a document. For example, each transaction that is recorded and communicated beyond the boundaries of an individual workstation, a workgroup, an organizational unit itself or the organization (depending on what boundary the institutional policy decides is evidentially significant) could be captured in an audit trail and written out as a data file. But in order for the system to use data of this sort, the information must be active with respect to the application software, as it is in a document header, rather than passively recorded as data of the document itself. Unfortunately, existing standards for office systems do not provide for interchangeability of header information regarding permissions, or file locations, of such documents. Existing standards do not even define methods of explicitly writing such header data to the file being kept as evidence.

Structural information must also be retained if we are to have adequate evidence and conveying such structural information beyond the confines of the creating software is becoming more, rather than less, problematic. As the means of exchanging information created in electronic systems on media other than paper become widespread, we are witnessing a rapid evolution in "forms of material" which reflect new structural relations between data. Forms of material are socially or culturally constructed information containers with which we communicate.¹⁰ It is as a result of these structural signals that literate people in our culture can immediately recognize the difference between a job application, a greeting card, or a legal summons without reading the words which

appear on each. We are currently witnessing a rate of change in forms of material that has not been seen since the advent of writing and the subsequent introduction of printing.¹¹ Because the appearance of electronic information is governed by software control of the application in which the information is made to appear, common conventions for rendering structural information are essential for the full meaning of records to be transferred.¹²

Electronic technologies are now being used to generate an entirely new form of material called hypermedia or compound documents. These non-linear "documents" are composed of data objects in the form of text, image, and sound that are linked by pathways defined by the author. The pathways are made manifest on being "played back" through software which is at this point still non-standard. Hypertexts and compound documents may, like the graphics generated from spreadsheets contained within word processing documents, point to dynamic data objects which take on characteristics of a changing environment, thus they may be different each time they are viewed. The implications for archivists are obvious: with these forms of material the transaction, and not the data of the document, is the archival record. Standards efforts for hypermedia have not yet succeeded in defining all the terms with which they must deal (the second NIST conference on hypermedia standards in the summer of 1990 had to be canceled because of this), to say nothing of dealing with the contextual data questions raised by archivists.¹³

The challenge of retaining the evidential historicity of dynamic forms of material goes beyond what has been recognized by the standards efforts of today. These efforts have focused on the interchangeability of the data content of documents and to a lesser extent on interoperability based on structural information. They have not had adequate regard for the need to capture and interchange contextual data about the particular transaction in which a user encounters a dynamic document. To do so, the standards efforts would need to ad-

dress documents comprised of data objects that are brought together under user control through the functionality provided by one application, but which might themselves be created by other applications. In the moment of being brought together they have an evidential historicity which the systems designer could have designed to adhere to them. But the methods of representing the evidential historicity of the document would, by necessity, have to be non-standard because no such standards have yet been developed. Yet the new forms of material with which we are increasingly dealing require precisely such standards, operating above the application level, in order to survive with their evidential historicity intact.

Electronic Mail

Electronic mail is a deceptively non-traditional technology which can illustrate some of the problems and potentials of electronic environments. The speed with which electronic mail is delivered seems to reduce the length of individual messages within the communication. With electronic mail, numerous messages will be exchanged as a dialogue where previously the participants would have written a finished argument in a single exchange of letters. The velocity of electronic mail exchanges (the time between a message and the response) in many organizations is 2-3 times per day, thus the equivalent of a full complex exchange can take place incrementally over several days (in the length of time it would have taken to send a letter).¹⁴ All the messages in the exchange can be sent to numerous individuals who respond asynchronously. It is necessary to know what messages have been received by whom and at what point in a discussion in order to fully understand the responses. In other words, the evidential historicity of the documents is essential to the reconstruction of the transaction.

In contemporary implementations of electronic mail, the "envelopes" of the messages are defined by the ANSI standards X.400/X.500 to carry the address to which they will be sent, whether acknowledgment is required, and other infor-

mation relating to the context of the exchange. Archivists require that additional information about the communication context be preserved for the record to retain its evidential historicity but they have yet to examine envelope headers to determine specifically what additional information would be required to establish the provenance, or originating context, of electronic messages. Systems designers have, to date, paid little attention to capturing the envelope data in their attempts to archive electronic mail. Thus we find ourselves in a world in which we have the technological capability to capture a level of context-based information surrounding records that surpasses greatly any data we had in traditional archival provenance, but have settled for a plateau of data interchange that does not satisfy archival requirements. Higher plateaus can be easily envisioned, and the economic and technological costs (including the potential explosion of the size of information bases as they carry evidential historicity data concerning all their objects) can be calculated. Archivists need to participate in such efforts in order to articulate the functional requirements of archivally acceptable electronic mail environments.

Spreadsheets

If we look briefly at the spreadsheet, another common office application, we encounter these and other issues. The cells of electronic spreadsheets contain formulas relating to the nature of work in the organization. These formulas embody a great deal of knowledge about the organization, as anyone who has ever tried to use a spreadsheet developed for a specific office accounting function will recognize. However the nature of the data being cited and the history of the expression or formula are not documented in the application itself unless individuals make a concrete (and rare) effort to document it.

Transactions involving spreadsheets are even more complex. In addition to hiding their operative formulas, spreadsheets may draw their data from numerous active databases and they may be displayed to those making decisions as a series of graphs rather than as textual data points. Evidence of

this transaction is not currently represented in application software except as the resulting graph, which is a print record surrogate, but which disguises the calculations as well as the raw data on which it was based. In short, the spreadsheet is rife with contextual data without which the "raw data" is hardly meaningful. In this and other many layered application software environments, standards for information interchange barely address the transfer of the data content and the structural information that assigns a calculation to the appropriate cell from one system to another, and none have addressed the contextual information interchange issues.

A spreadsheet displayed as a graph, drawing data from a variety of contemporaneous database states, may, furthermore, be displayed in a word-processed document in which it resides as a pointer, rather than as a fixed manifestation. If such a document is used for decisionmaking, archivists will want to know what it said at the time the decision was made, because if they save the document with its pointer, even if they could succeed at the task of keeping the associated information systems intact, they would view a different image when the document is next displayed.

In the spreadsheet we are beginning to encounter a new kind of data -- intelligent data -- which could be major contributors to the productivity of the office in the next decade and which pose a serious threat to archives.¹⁵ "Intelligent data" are aggregates of raw data created in a system which imbues them with contextual information that instructs them how to behave in that system; for example, a text given the "intelligence" assigned to a memorandum would "know" when to send itself and to whom to send itself. A word or two assigned the intelligence of a future date would "know" to remind the creator on or before that date. Creators need not assign this intelligence themselves, because in intelligent systems some of this work can be done automatically. Thus a program could determine the keywords under which to file a document, build a table of contents and an index, and screen readers for security re-

quirements which it determined from reading the text. In the spreadsheet, each cell contains the formulas which it uses to analyze the raw data provided to it.

Database Management Systems

Database management systems are the next most common office applications. When we examine them from an archival perspective, some contextual information retention requirements present themselves immediately. The raw data in a database has some informational value, but it has little evidential meaning. The transactions in a database environment are input records, queries, and output instructions. Without the documentation of the system, we do not know what data any given individual or office had included in their "view." Without update records, we cannot tell the state of a given piece of data at a given time. Without links to other systems that might have been provided to active users of the system, we cannot tell what use might have been made of the data. This kind of documentation of the "potentialities" of the data is recorded in metadata systems called data dictionary/directory systems and Information Resource Directory Systems to which archivists need to pay greater attention. In addition, we could capture documentation of the actual uses, the actual updates, and the actual data participating in specific transactions. In principle a transactional audit trail could document changes so completely as to make possible a "time-travel" database, but the practical implications of maintaining a system so that it can back itself up in time are formidable. Therefore, archivists will need to work with systems designers to assure that database systems create certain transactional records (for particular types of transactions initiated by certain functions within the organization) that otherwise would not be created.

To get a complete picture of what the database does we need to document also whether it acts under human control at all times or if the database management system sometimes takes action (and creates records) on its own. For example, does it have a "tickler" report, or does it generate warning no-

tices when someone tries to breach its security, or does it generate weekly, monthly, and quarterly reports at the appropriate time without specific human request? Increasingly, software is being implemented in organizations so that the database serves each office differently and so that distinct facilities are provided for different functions. The trend in information processing is towards more and more "intelligent" systems. These include systems using "object-oriented languages" and "artificial intelligence" but very traditional programming is also embedding growing amounts of intelligence in software. Unfortunately, the methods of representing this "intelligent" information surrounding the raw data, like those for representing contextual information, are non-standard. Nevertheless, the system rules are context to the transactions, and without them we can only poorly comprehend the evidential significance of a record.

Graphical Software

Beyond the office, a class of applications which include Geographic Information Systems (GIS), Business Graphics and Statistics and Computer-Aided Design (CAD) systems are growing in importance. These applications are characterized by a graphical presentation of data. They enable users to envision the relationships between large numbers of discrete data points by displaying them visually. As such they enable effective communication of complex information, and their outputs are records in such communicated transactions. But in order to employ these outputs as evidence, the underlying data are required in addition to the visualization.

Geographic Information Systems (GIS) can serve to illustrate this. GIS environments are one of the most rapidly growing applications created by and used in government today because so much of the function of government involves delivery of services to the population of a geographic region. That population, and the infrastructures that serve it, must be envisioned by those delivering and receiving services.¹⁶ Geographic Information Systems are databases whose records

are data points in a spatial linking structure (e.g., a map). They constitute a powerful application environment precisely because data points derived from one source can be retrieved in conjunction with data points from other sources, thus building a comprehensive view of what is known about any space on the face of the earth (or under water). For government, the organization of databases by geographical locus is an important tool for management. However, the data points contributed to a geographic information system will have been established at different times, by different agencies, and for different purposes. From an archival perspective, this means that they will have different provenances. Archivists need to decide both what data to retain about the context of creation and use of each discrete data point with respect to its original relation (which becomes simply another retrieval set within the operational GIS application) and what contextual data needs to be preserved as evidence of functions that subsequently use that data point in creating records (transactions) in other applications employing the GIS. Even if there were standards for the interchange of GIS data, or the interoperability of GIS systems, they would not address the retention of the evidential historicity of these new forms of material.

Another increasingly important software-dependent data object is being created by Computer-Aided Design and Computer Assisted Manufacturing (CAD/CAM) systems. These systems employ vectorized information (formulas for lines) to generate displays of three-dimensional objects such as buildings or machine components. CAD/CAM files are used to make decisions about construction and maintenance of increasingly large portions of the capital assets of governments, but they are highly software dependent. Changes to drawings in CAD/CAM systems may or may not be documented and the contextual significance of the source of the data are not generally considered important by the architects, electricians, manufacturers, etc. who make and use the latest version of this data. Hence the implementation plateaus in these systems are

set far below the levels regarded by archivists as important to record for evidential historicity. Meanwhile standards, such as IGES (Initial Graphics Exchange Environment), are focused on interchange and interoperability below the application level.

ELECTRONIC RECORDS MANAGEMENT GUIDELINES

Electronic information systems, therefore, present at least two challenges to archivists. The first is that the designers of these systems may have chosen to document less contextual information than may be of interest to archivists when they designed the system. The second is that the data recorded in any given information system will, someday, need to be transferred to another system. As long as the information created in the course of work in an electronic environment remains in the software and hardware system in which it was created, it loses none of the contextual information which is critical to its meaning, but the transition, or "migration," of data to a new environment threatens to change the way the information looks, feels, or operates, and hence what it means.

Archivists have always been dependent on the information systems implemented in working offices for the amount of contextual information that would be available as evidence. What has changed with electronic records is both that a greater degree of evidential granularity is possible, and that it may not be retained because program managers are less aware of its value, or of their ability to require its retention, than they were in paper-based office systems.¹⁷ Archival retention of contextual data at appropriate levels of granularity is likely to depend, in the future, on program managers insisting on their requirements for this contextual data for information management within their functions. For this to happen, archivists need to articulate what contextual data are required as evidence, how it might be captured, and at what cost.

Although archivists accession records at the level of the record series, we found very little contextual documentation at

the series level in paper-based records environments. As a consequence, documentation of the paper-based records systems by procedures manuals rarely answers questions such as:

- What functions and transactions within functions created records in the series?
- What functions filed into the series and how did they exercise their judgment?
- Which offices had access to information in the series and could they alter it?

Instead, in paper-based environments we find evidence of creation and use at the individual document or transactional entry, or at the level of file folders which detail who saw their contents, the time at which they received them, and what they did with them (forwarding, taking action, etc.). In electronic information systems, metadata systems should explicitly document the activity context out of which records arise, and policy should be implemented to define forms of documents created in such contexts and the organizational retention requirements for each. Systems can then be implemented so that when a record is identified by its origin and date, appropriate retention policy can be automatically executed. Records which need to be retained for their continuing value can be protected from alteration or deletion.

The second threat is more insidious. It is expensive to transfer data from one system to another, especially as the information we are trying to capture in these transfers is the most system-dependent and the least standard. Each system migration is accompanied by extensive planning. Often it must include making modifications to both systems in order to preserve functionality associated with the old system and to retain contextual data, including that upon which systems functionality depends. In an operational environment, the decision is usually made to "migrate" as little data as possible into the new system and to preserve a minimum of contextual data. Archivists must realize however that data loss in migration,

like data loss in copying, is a one-way street. As decisions are made to leave some contextual data behind, the evidence provided by the information that is transferred will be lost.

The current state of standards does not permit functional equivalency (by and large) to be communicated across systems. Only the "data" content of texts, images, and sounds are accommodated by existing standards. New standards will eventually be developed (ideally with archival input) to bring more and more contextual functionality across systems barriers. However, new methods of data organization and new concepts in computing will also continue to emerge and will always lack standards when they are new so the problem will not go away. Archivists will have to continue to face the questions of how much functional equivalency they will try to transfer and how software dependent they will permit the data they manage to be. In other words, archivists will need to confront, on an application-by-application basis, the potential for capture of data with various degrees of evidential historicity.¹⁸

In each case, archivists will need to return to fundamental archival principles to determine just what they really wanted to save anyway. And they will need to look forward into the management of the current technologies to determine how practical it is, or will be, to save data of evidential value. It may be that archivists will be satisfied with the degree of evidential historicity they were able to achieve in paper-based record systems, in which case there are very few barriers to implementing successful electronic-based archival environments. Or archivists may decide that the fuller capability of tracking the actual participation of electronic data objects in organizational activities needs to be documented by archivally satisfactory information systems, in which case they will need to define those levels of evidential historicity that must be attained and specify the systems requirements for such environments.

At a meeting on electronic records management research issues sponsored by the National Historical Publications and Records Commission in January 1991, participants identified the concept of technological and economic plateaus in electronic data capture and archiving as an important arena for research. They proposed research efforts to identify such plateaus in existing applications and to define the technical challenges associated with the retention of greater degrees of evidential historicity within applications as well as the problems and prospects for long-term preservation of data representing different plateaus.¹⁹ Their proposal would, for example, bring sociological evidence about transactions in organizations together with technical specifications of different degrees of evidential historicity of records to generate preliminary estimates of the increases in data volume required to satisfy different levels of requirements imposed by the archival principle of provenance. It is hoped that this research will produce information to help archivists make decisions regarding the amount of contextual information they can afford to capture and the requirements of systems designed to document context along with managing data content. In any case, the analysis to date has enriched the concept of provenance and reinforced its direct link to missions, functions, and ultimately the activities or transactions of an organization rather than to organizational units, as was predicted by Richard Lytle and me in an earlier article.²⁰ I will not be surprised as we refine our concepts of evidential historicity to discover that the concept of provenance takes on even greater granularity and is associated with the particular actions.

NOTES

¹ The term "corporate memory" is used extensively in the writings of John McDonald and the policies of the National Archives of Canada to refer to that evidence which it is important for an organization to retain for its ongoing operations and legitimacy. See for example, National Archives of Canada, "Strategic Framework for the Information Management Standards and Practices Division" (unpublished manuscript, April 1991).

² These concepts were originally developed in a discussion with Richard Cox, Margaret Hedstrom, John McDonald, and Lisa Weber in January 1991 following the NHPRC-sponsored Working Meeting on Research Issues in Electronic Records Management.

³ Note that *archives* as used throughout this chapter refers to a kind of records not to an institutional setting. The institutional setting called an *archives* may contain materials that do not fall within the scope of this definition of archives, such as personal papers, materials kept for their "informational value," and information generated by archivists. The narrow sense of *archives* as evidence is explained further in Chapters 2 and 3 of this volume. See also Sue McKemmish and Frank Upward, "The Archival Document," a submission to the Public Inquiry into Australia as an Information Society, *Archives and Manuscripts* 19:1 (May 1991): 17-31.

⁴ *Digital Image Application and Optical Media Systems: Management Issues, Technical Trends, User Experience: Guidelines for State and Local Agencies*, A Joint Report by the National Archives and Records Administration and the National Association of Government Archives and Records Administrators (July 1991). See also Association for Information and Image Management, *The Use of Optical Disks for Public Records*, Technical Report #25 (Washington, D.C.: AIIM, 1990).

⁵ These paperwork practices in the office of the president were common in the Kennedy Administration as reflected in records at the Kennedy Presidential Library. Similar practices probably were not followed in the Reagan administration use of the IBM Profs electronic messaging system because the system provided some of these features as long as the records were active within it. Chapter 4 in this volume further describes the implications of the suit by Scott Armstrong, et al., to prevent destruction of the Reagan administration White House Profs files.

⁶ Even though storage costs may prove to be insignificant in the future, it is evident that migration costs will not be trivial in the absence of complete interoperability standards.

⁷ The half-life of software products is less than three years and less than one year between releases. In an environment in which several layers of software will run between the operating system and the application, maintaining all these functional capabilities "as they were" would be technically impossible even if it were not the case that the software licenses do not transfer with the data if it comes to the archives, thus making it legally impossible to maintain commercially developed application software.

⁸ As the user's mental model departs from traditional paper models, the complexity increases. Hence databases and GIS systems require greater contextual data capture to represent transactions that word processing or electronic mail where the transaction is essentially like that conducted in the course of business in a paper-based office.

⁹ At present, I know of no reliable studies of the costs of these migrations that can be used for archival cost-benefits analysis.

¹⁰ The definition I wrote for the NISTF data element dictionary has been used in the archives community but few have followed its implications. See David Bearman and Peter Sigmond, "Explorations of Form of Material Authority Files by Dutch Archivists," *American Archivist* 50 (Spring 1987): 249-253; and Luciana Duranti, "Diplomatics: New Uses for an Old Science," *Archivaria* (1988-1992), Part I, 28:7-27; Part II, 29:4-17; Part III, 30:4-20; Part IV, 31:10-25; Part V, 32:6-24; Part VI, 33:6-24.

¹¹ Forms of electronic records are discussed in David M. Levy, Daniel C. Brotsky, and Kenneth R. Olson, "Formalizing the Figural: Aspects of a Foundation for Document Manipulation" (Systems Sciences Laboratory, Xerox Palo Alto Research Center, 1988).

¹² The Text Encoding Initiative, an undertaking of numerous text-oriented humanities disciplines, is developing methods of indicating the structural aspects of texts going back to the middle ages using Standard Generalized Markup Language (SGML). Twentieth century documentation practices are generally covered by a smaller set of declarations developed by the American Association of Publishers. More complex modern Document Type Definitions (DTDs) are being implemented for multimedia documentation for airplanes, battleships and like systems following the specifications developed by

the Department of Defense Computer-aided Acquisitions and Logistical Support (CALs) initiative.

¹³ The National Institute of Standards and Technology (NIST) called a conference on hypertext standards in the early spring 1990. A follow-up meeting to have been held in the summer 1990 was canceled because there was too little agreement about how to proceed to make progress possible. Currently two efforts are underway which emphasize different aspects of hypermedia interchange: HyTime, an extension of the SGML approach, declares the logical functions and links of objects being interchanged while MHEG defines the technical specifications for data representation in those objects.

¹⁴ Tora Bikson and J.D. Eveland, "The Interplay of Work Group Structures and Computer Support," in *Intellectual Teamwork*, ed. Kraut, Galagher, and Egido (Hillsdale, New Jersey: Erlbaum, 1990). Observations on university campuses where electronic mail is widespread confirm that the medium is leading to the evolution of a truncated form of communication, sometimes resulting in exchanges analogous to the nod or grunt in face-to-face communications.

¹⁵ Timothy J. Heintz, "Object-oriented Databases and Their Impact on Future Database Applications," *Information & Management* 20:2 (1991): 95-103.

¹⁶ Kentucky Information Systems Commission, *Current Issues in Government Information Policy Conference Proceedings* (Frankfort, Kentucky: KISC, June 1991).

¹⁷ The practice of associating provenance with the governmental agency responsible for the creation of an entire body of records, which led to the establishment in the United States of the "Record Group" concept, has largely been superseded in the U.S. and Australia anyway by the link between the record series and its functional provenance, or the activity of the agency that gave rise to it. The seeds for this move are found in the writings of Peter Scott in Australia in the 1960s and 1970s; of David Bearman, Max Evans, Richard Szary, and others in the U.S. in the 1980s; and in the practices surrounding the implementation of USMARC AMC and APPM. The increasing granularity of provenance may be a trend that will carry the assignment of provenance down to the contextual information in particular documents in electronic information systems.

¹⁸ The identification of the business application as the locus of intervention by archivists is based on the realization that requirements for data retention derive from the combination of requirements by

ongoing offices for information regarding their functions and the requirements of the organization for evidence of its activity over time. Both of these requirements are focused at the program level and must be achieved through the implementation of software to support concrete functions of the organization. For further discussion, see Chapters 2 and 3 in this volume.

¹⁹ The National Historical Publications and Records Commission, *Research Issues in Electronic Records* (St. Paul: Minnesota Historical Society, 1991).

²⁰ David Bearman and Richard H. Lytle, "The Power of the Principle of Provenance," *Archivaria* 21 (Winter 1985-86): 14-27.