Technology Trends

Introduction

DAVID BEARMAN

About the author: David Bearman is president of Archives & Museum Informatics, a Pittsburgh-based consulting firm. Bearman consults on issues relating to electronic records and archives, integrating multiformat cultural information and museum information systems and edits the quarterly journal Archives and Museum Informatics.

THE FOLLOWING STIMULATING DISCUSSION of technology trends and their archival implications owes much to the themes introduced by Ronald Weissman, the lead speaker at this section of the 2020 Vision session. Weissman is currently the director of strategic marketing for NeXT Computer, Inc., but he also has a background in modern European history. As a historian and a computer company executive, he was asked to forecast the technology trends of the next thirty years. I think the most important lesson we can draw from Weissman's paper is that his forecast is not a forecast at all. He rather cleverly limits himself to exploring probable implications of technologies that currently exist and how they will play out from now until the end of the decade. In fact, he underscores for all of us the lesson that came from all the visionary talks: that managing change may itself be the challenge of the next thirty years because in the technological realm, as in the organizational and ideological realms, prediction is a business frought with risks.

Weissman tells us that computers will become cheaper and faster and more ubiquitous, and he argues that the paradigm within which we operate in the world of computing will shift from one of software applications (centered on the machines and the software in those machines) to a paradigm that is information centered and, ultimately, to one that is centered around work and what we are doing in our lives and in the organizations where we work. This shift will be made possible by technologies that currently exist and are now being implemented in object-oriented systems and that will increasingly be implemented in computing systems in the next six years. The final lesson we can draw from Weissman's article is that technology is not self-implementing. Archivists will have a major role in deciding how it will be implemented.

In her comments on Weissman's paper, Luciana Duranti of the University of British Columbia brings to bear a rich European archival perspective and her insights from her work on diplomatics. Duranti notes that a universe in which technology is essentially invisible to the user is to some extent an archival nightmare. Individuals in such a world would be unaware of the role that technology was playing in creating records and in documenting their activities, and that blissful lack of awareness would endanger the historical record.

John McDonald's commentary draws on his experience as Director of Information Management and Standards Practices at the National Archives in Canada, and he speaks from the viewpoint of someone who has devoted considerable time in the last few years to trying to make bureaucrats and technologists in the Canadian government more conscious of the threats to what he calls "corporate memory." McDonald also sees a threatening trend: the way technology actually operates is increasingly invisible and increasingly different from that imagined by its users, because very clever metaphors in user interfaces present a view that differs from the actual events that are taking place in the machine. But McDonald also sees reason for optimism in the picture Ron Weissman portrays and in the challenges that Luciana Duranti presents. First, institutions and their leaders are becoming aware of the threats that the electronic age poses to corporate memory. Second, archivists alone can understand the kind of functional requirements institutions have—not for preserving information, but for preserving evidence, preserving the record, and preserving their archives. If archivists defined their functional requirements, they could help the organizations in which they work to implement the change and to implement the technology. This returns full circle to Ron Weissman's point—that technologies are not self-implementing and that if they are to work in organizations, either as models to enable people to conduct their day-to- day work or as mechanisms for us to ensure the long-term preservation of memory, we will have to be strategic about how we implement them.

We, as archivists, are caught up in the changing definition of technologies and the changing definitions of our cultures. Our challenge is to implement what we are best at implementing, and that is a formal sense of the record as evidence, the record as bearing on transaction. If there is one technological and ideological concept that informs our work, it is a uniquely twentieth-century notion of systems, networks, and interrelationships. And those systems and networks and interrelationships provide an opportunity for us to capture communications of transactions—to capture evidence, and thereby to focus on that aspect of information which is uniquely archival.

Technology Trends

Archives and the New Information Architecture of the Late 1990s

RONALD F. E. WEISSMAN

Abstract: While hardware capabilities will increase dramatically through the end of this decade, the real change in electronic technology will be the spread of an information architecture. This architecture will change the model of how we interact with computer-based information. The spread of new software technologies, such as object-oriented software, workflow, document-centric computing, and database-centric computing, will change substantially our traditional information-processing model. Of greatest interest to archivists will be the potential of this change to integrate today's diverse information search and retrieval strategies and technologies.

About the author: *Ronald F. E. Weissman is director of marketing at NeXT Computer, Inc., and manages all vertical markets, strategic and business planning, market research, company positioning, and systems integration partnerships. Weissman is NeXT's senior market strategist and a senior corporate spokesperson. Before joining NeXT in 1990, he served as an adviser to NeXT for four years. He has also been an adviser to Microsoft, Apple, IBM, and Sun Microsystems. Weissman has served as an information systems consultant to federal agencies and other ogranizations, and he has advised and written strategic IS plans for such organizations as the U.S. Departments of Defense, Transportation, and Energy; the USDA Food and Nutrition Service; Oak Ridge National Laboratory; and the National Archives and Records Administration. The long-range plan he authored for the Food and Nutrition Service was cited by the White House as a model plan. Most recently, he served on a national advisory panel advising NARA on the future of electronic records in federal agencies. Weissman was assistant vice president for computing at Brown University and associate professor of history, assistant to the president, and director of academic computing at the University of Maryland. A Fulbright Scholar with a Ph.D. from the University of California at Berkeley, Weissman has published extensively on object orientation and desktop computing architectures.*

OUR MODELS OF INFORMATION PROCESSING have not changed fundamentally since the mid-1970s. During the next decade, however, I believe that core models of information technology will change dramatically because we are at the infancy of information-processing technology. I distinguish information processing from data processing in several ways. Information processing seeks to impute meaning to data, to uncover patterns and to wrestle creatively with ambiguity. By contrast, data processing deals with a few regular types of data and seeks to organize these data in regular and predictable ways via standard report generators. The fundamental task of the archivist is to manage meaning-rich information, not simply to store or classify raw bits of data. To a large degree, the reason so little has changed during the past several decades in information processing is that our models of computing have been attuned to the easier problems of data management and our technology has lacked the requisite power to deal with the more difficult and more interesting problems of information processing.

This discussion of the future of information technology seeks to project change out to the end of this decade. To project trends longer than eight to ten years is, I believe, a fruitless exercise, for it will take only one breakthrough during this period to render all speculation about the years following 2000 irrelevant. Advances in optical computing, superconductivity, nanotechnology, or biotechnology will fundamentally change the scale of what is computationally possible. Should biochips become real, for example, any estimate made today would be akin to plotting the kind exploration possible using a pack mule when, in fact, the Concorde was really at one's disposal. Nevertheless, even if technological breakthroughs occur during the next several years, it will still take years to turn research into products. Thus, while I feel relatively secure projecting advances in technology out through the end of the decade, I believe that speculating further is an exercise in ensuring being wrong far more often than being right.

New Computing Platforms

Organizational architecture. The basic hardware capabilities of the computers that we all will use in 2000 will differ from the computers of the early 1990s in form, capacity, and speed of processing (see figure 1). Tomorrow's architecture will be fully client/server based, with personally configurable portable machines having disk drives, a selection of monitor and projection devices, keyboards, flash memory cards, pens, microphones, cellular modems, and portable scanners, all available as snap-together accessories or personalities. These personal devices will be networked—sometimes physically and sometimes via wireless technology—to powerful servers containing trillions of characters and millions of pages of information stored locally. These servers, in turn, will create a distributed information architecture throughout an organization, linked by very high-speed networks. Today's mainframes will be used as tomorrow's highest performance and highest capacity servers, and they will differ from other servers in computational and storage capability but not in basic functionality.

Future computers. Through continued progress in microminiaturization, two fundamental classes of computers will exist by 2000: highly portable component computers and larger, stationary high-capacity servers. Tomorrow's personal systems will consist of components (storage, CPU, input devices, and display devices) that snap together and that allow users to carry with them and use as much or as little as they desire. These personal computers will have revolutionary new form factors, allowing them to be clipped to a working space, attached to a wall, embedded in larger de-

Figure 1. What $5,000 Will Buy in Hardware and Networking Technology

	1992	2000
Form factor	Desktops and notebooks	Modular portability plus very large servers; embedded systems
CPU power	25 MIPS CISC and RISC	1,000 MIPS; multiprocessing RISC
Screens	70 to 100 dots per inch, grayscale; 640 × 480 to 1,024 × 768	150 to 200 dots per inch, color; 1,200 × 1,200; wallboard displays; holographic displays
Memory	4 to 8 megabytes	256 megabytes to 1 gigabyte
Storage	200 megabytes	10 gigabytes local; trillions of bytes on local servers
Media	Rich text and graphics; limited sound and video	Rich text, graphics; CD sound; full-motion video; integrated compression technologies
Networks	2 to 16 million bits per second, using Ethernet and Token Ring	600 million+ bits per second, using asynchronous transfer mode
Local area network/ wide area network	Hardwired, peer-to-peer local area network	Wide area network; client/ server architecture; cellular connectivity; universal office and home ISDN
Network geography	Linked to physical network; optimized for one location	Anyplace, anytime
Primary peripheral	300 × 300 dot-per-inch monochrome printer	1,000 × 1,000 dot-per-inch color with built-in fax and scanning

vices (such as automobiles or appliances), or designed into flexible materials—including materials allowing them to be worn instead of carried—and treated as fashion accessories, available in a range of colors and styles. It was in 1992 that the spring designer fashion show in New York for the first time featured prototype wearable computers from leading manufacturers.

The traditional stationary desktop computer will be made obsolete by this variety of portable, embedded and wearable computing technology. At the consumer end of the technology chain, computers will be so cheap, so much a commodity, that like today's instant portable cameras or calculators, they will become disposable, one-use devices for travellers who, for whatever reason, have failed to bring along a computing device. At least until the biochip revolution, there will always be a place for larger computers, but the CPU and user-interface components—screen, microphone, keyboard, and mouse—will be lightweight and highly transportable.

CPU processing power. For the past decade, raw computer processing power has doubled every eighteen to twenty-four months with little change in price. For this reason, most chip vendors project that today's $5,000 computers running at 20 to 50 million instructions per second (MIPS) will be replaced by computers providing more than 1,000 MIPS in the year 2000.

This is approximately the processing power required to create real-time photorealistic animations on a par with *Star Wars* or *Terminator 2*. Much of the processing power of these computers will come from continued miniaturization of chip technology and from improved clock speeds, multiprocessing and superscalar architectures in which multiple computer instructions are executed during a single computer clock cycle. This enormous growth in processing power will enable whole new classes of applications, and they will increase the effectiveness of technologies now still in their infancy, such as continuous speech recognition or, more relevant to our discussion here, cursive handwriting recognition.

Displays. If information technology is ever to compete with the paperback book, screens will have to become significantly lighter, with greater contrast, better color, and much higher resolution to allow users to read finely detailed pages of text for extended periods of time. By the year 2000, inexpensive, true-color, high-resolution screens exceeding 150 to 200 dots per inch will be standard and will come in a variety of forms, ranging from full-page models for portable computing to two- and three-page models for office use. In conference rooms and meeting areas, wallboard-sized screens will also be increasingly common, allowing interactive computer conferences. New projection technology will make possible holographic displays, allowing the illusion of live, interactive, three-dimensional objects projected in three- dimensional space. By the end of the decade, HDTV-quality screens will begin to integrate with traditional computing technologies for a universal, high- definition multimedia display and imaging platform for video applications.

Memory. With inexpensive 16-megabyte memory chips now available and 64-megabyte memory modules under development, tomorrow's computers will be equipped with hundreds of megabytes of local memory and many more megabytes of reserve "virtual" memory on disk, enough to store hundreds of thousands of typed pages in a computer's main memory and millions of pages on local disk drives.

Storage and media. Tomorrow's multimedia computers will be capable of storing and displaying far more than the typical typed page. Given advances in compression technology, storing both the editable text of a page as well as its scanned facsimile image in grayscale or color will be increasingly common. Multigigabyte local storage devices and trillion-character storage devices on nearby servers will also be commonplace. Advances in compression technology, such as fractal-based algorithms, will allow compression rates to exceed 100:1 without perceived visual loss of quality. Similar rates will be possible for video, voice, and sound. And compression chip technology will be built into the architecture of most computers. Other compression technologies under development will allow video—which currently requires approximately 45 megabytes per second for full-frame, high-quality display—to run effectively at rates as low as 1.5 megabytes per second.

The ubiquity of CD-ROM jukebox technology and the rapidly expanding memory and storage of personal computers, multiplied by the effect of quality compression, will transform the way librarians and archivists think of retention. I expect tomorrow's debates to be as much about what classes of documents to compress at very high rates as about what to preserve or discard. Tomorrow's storage systems, which will, after all, contain their own locally embedded smart controllers, will allow virtually limitless storage of materials of all kinds.

Networks. Just as hardware is experiencing dramatic increases in capability, so also will local area, wide area, national, and international networks. Today's Ethernet and Token Ring networks are capable of transferring two million to sixteen million bits of information per second. Tomorrow's emerging fiber digital data interconnect (FDDI) and ATM (the latter is essentially a high-capacity broadband ISDN) will increase network bandwidth by roughly the same performance factors as other elements of the computing environment. High-speed networks will, for example, vastly exceed today's dedicated T-3 land- line links, providing the capacity for multisession, real-time, full-motion video. And the number of public and private data network carriers and service providers will expand dramatically, offering users a range of services and data rates, from low speed ISDN for traditional data to high-speed ATM (600 megabyte per second) for real-time, multichannel, multisession video.

Today's networks are largely hard-wired, requiring physical interconnection between computers, networks, and servers. Tomorrow's networks will be increasingly based on wireless technology allowing one to connect from anywhere to anywhere, although physical interconnection will still provide the highest and the most reliable data throughput.

Peripherals. Peripheral technology will evolve in ways parallel to the rest of the hardware architecture. Tomorrow's standard peripheral will be a four-color laser printer with fax scanning and photocopying capabilities and will print at resolutions of 1,000 by 1,000 dots per inch, providing photorealistic output and custom publishing and printing on demand. Scanning will encompass a broader range of uses, including handwritten documents. Through the use of fuzzy logic and related artificial intelligence, neural network, and pattern recognition technologies, cursive handwriting recognition will become a useful technology of potentially great significance to historical archives.

These emerging hardware capabilities are interesting for the new functionality they will allow, including highly portable computing anywhere, anytime, in any media. They will allow real-time photographic and HDTV-quality simulation of complex processes—real-time movie making where none of the objects visualized actually exists in the real world. But in terms of relevance for the archival community, these capabilities pale in relation to the new software architecture made possible by the revolution in hardware.

This revolution will allow millions of documents to be stored locally and to be made vastly more accessible over networks throughout an organization in a distributed client/server architecture. In this context, a fundamentally different information-processing model will emerge, one that focuses the user on information, not, as today, on software applications. The new hardware capabilities, particularly the doubling of raw power every eighteen to twenty-four months, will be the enabling force behind a new information paradigm. But it will also in part be the cause of that new paradigm because today's software architecture is incapable of effectively managing the millions of related documents that new hardware capacities will enable.

Toward a New Software Architecture and a New Computing Model

It will take the kind of revolution in hardware capabilities described earlier to make possible the first real revolution in software architecture since the invention of time-sharing, terminal-based computers in the 1960s. Our information-processing model has not changed much during the past two decades, apart from becoming vastly easier to use through the development of graphical user interfaces, developed by Xerox Parc and produced as usable machines by Apple Computer, Inc., originally.

The traditional information processing model is represented in figure 2.

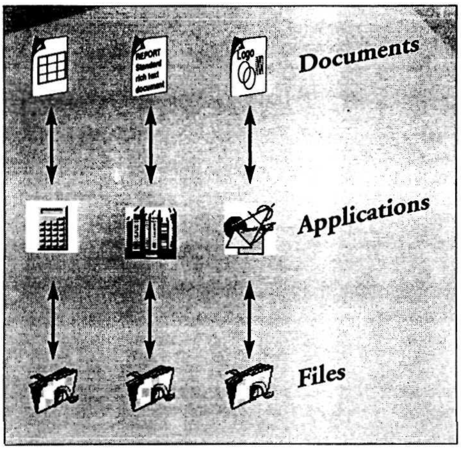


Figure 2: Traditional Application-Oriented Architecture

- The user runs software applications—separate word processors, spreadsheets, databases, and graphics programs.
- Information is created and, once created, is managed by these applications.
- These applications store and retrieve information in specialized kinds of files that are unique to the applications that created them.
- It is the task of the operating system to manage the files created by diverse applications.

In this traditional information model, users launch applications, several of which may be opened simultaneously, each in its own window. Within each software application, a user may open one or more documents, each of which also resides in its own window. Given this flexibility, what is wrong with this model?

First, it does not correspond to the way information-intensive analysts and researchers work. The desktop metaphor—using multiple tools (such as calendars, budget spreadsheets, and word-processed memos) to manage a busy day—aptly describes the work of a mid-level manager performing a variety of functions at once (e.g., a telephone log, an expense account, and a report). It is less well suited to the information professional, analyst, or researcher who must manage different kinds of information, all of which are associated with a common project. Today's computing model forces us to segment these related kinds of information because the tools that work on each kind of data are different. The end products of modern research today involve complex documents that may contain a variety of information from many different sources. Today's sophisticated researcher defines projects that require data tables, statistics, maps, graphs, text, free-form drawings, precision-drawn technical illustrations, and scanned bitmaps, video, and sound. It is inefficient to require users to run many different applications and manage many separate documents when, in fact, the purpose of the project is to integrate different kinds of data for a common analytic purpose.

A second and related problem limiting the usefulness of the current desktop environment is the tight coupling that currently exists between data and the applications that create them. Typically, data and software applications are tightly coupled. They coexist as application islands, generally disconnected from other software except insofar as vendors have provided data translation programs or filters that allow data to be ported from one application's files to another application, or to interchange protocols that allow data from one application to be embedded in another application. Knowledge about the underlying data model is essentially private to the application and not publicly available to the computing environment as a whole. It is therefore entirely up to the application designer to determine to what extent data created by one application may be usable by other applications and to provide a specific data embedding or exchange mechanism to accomplish this.

Third, users want to be able to locate everything in a computer system that is relevant to a given topic, regardless of how or where it is stored. What information exists about last year's budget hearings? About school funding in a given region? About the quality of public health-care services? Users do not care that some information is textual in form, or tabular, or graphical, or multimedia. Users want to find all relevant information stored on line about a given subject.

Today's operating systems provide only simple file names as clues to what is or is not relevant, and large files may contain a variety of information relevant to a variety of potential subjects. Even full-text searching tools may be inadequate for this discovery process. In any case, when using today's file-oriented systems, users must initiate multiple kinds of searches to have even the remotest hope of finding a range of related, relevant materials.

Our desktops and servers will be filled with potentially useful data of all sorts. But our ability to access and manipulate data in meaningful ways is limited, given our traditional model of how information is handled by software. Today, we can only really access such information from within the applications that created those files. Our data are difficult to access, and our sense of the possibilities of data analysis is bounded by our limited sense of the exploration, visualization, and analytical tools available to us. We must rethink our user environments and the ways we can manage, interlink, visualize, and interpret information. A software revolution, which will be complete before the end of the decade, promises to do precisely that.

The New Software Architecture

What are the key changes coming to the traditional information- processing model? There are too many to cover in any detail in a necessarily short paper, but let me outline a number of them. (See figure 3.) Key architectural changes include the following:

- Collaboration tools and workgroup document processing
- More powerful user interfaces
- A transition from monolithic to modular object-oriented applications
- A transition from stand-alone applications to intelligent workflow
- A transition from application-centered to document-centered computing
- A transition from operating systems as file systems to operating systems as databases
- A unification of searching and retrieval tools and paradigms

Collaboration tools and workgroup document processing. Tomorrow's stand-alone single-user applications will become tools for group collaboration, as operating systems gain automated features supporting versioning, concurrence tracking, redlining, and group markup. Local and long-distance electronic conferencing systems will provide electronic blackboards, allowing all users connected to the conference to annotate and illustrate in a common "play space." The history of those conferences, collaborative meetings, and serial document reviews will themselves be able to be replayed, documenting not only documents but the process of their group creation. Thus, a future opportunity for the archivist will not only be to store and document the official record of organizations but, given the versioning tools soon to be at our disposal, to document the process by which records are created.

More powerful user interfaces. Improvements in information-processing capabilities will enrich individual applications by allowing them to become multithreaded—that is, to execute multiple operations at the same time. (A spreadsheet, for example, could update itself from one database at the same time that it formats a report or communicates with a graphics

Figure 3. Advances in Software Architecture

Technology	1992	2000
Application model	Single tasking, stand-alone applications	Parallel processing, multithreaded and integrated
Productivity tools	Stand-alone word processing, spreadsheet, database, e-mail	Integrated object-oriented applications, collaboration tools, workflow
Operating system	File, disk, and process management; traditional operating systems	Information architecture; object-oriented systems
User interface	Application-oriented; "mouse click and shoot": keyboard, mouse, windows	Document-oriented; "say and point": speech, gesture, keyboard, mouse, 3D windows, pen, natural language
Computing model	Application and file	Document and database

package.) The most visible changes to come, however, will be in fundamental user-interface technology. Vast increases in processing power will allow users to command computers via speech and hand gestures. And the two-dimensional windowing model will soon give way to a three-dimensional play space. If research at Xerox Parc is any indication, new user-interface metaphors will focus on the navigation within a simulated geographical space (such as a library), more than the manipulation of two-dimensional windows. How geographic spaces are translated into navigable information spaces is one of the hottest topics in user-interface research today. Of most profound importance will be the integration of natural language as a primary element in all human-computer interaction.

From monolithic to modular object-oriented applications. In an object- oriented environment, all software tools are reusable, highly modular, communicating components, very much like integrated circuits. An application consists of many small modules, each of which sends and receives messages from other modular components. Applications are built from small communicating objects. And applications themselves—word processors, spreadsheets, databases—can themselves be treated as objects, sending and receiving messages from other applications. Thus, through user scripting and workflow management, one can constuct new functionality based on the interaction of existing software components. In the object-oriented environment, we will find such benefits as self-updating maps and charts, self-formatting publications, and spreadsheets that communicate their output with equal facility in terms of voice messages, charts, or updated graphs in a linked document. For developing integrated software environments for analysis and knowledge work, modular object-oriented platforms will become very important. Object-oriented systems are extensible by third parties—these systems encourage others to reuse components and software code. Finally, they are inherently suited to modeling complex problems and processes, such as organizations, budgeting, and negotiations and to modeling complex systems, such as aircraft or electric power plants.

From stand-alone applications to intelligent workflow. In this new world in which applications communicate by sending each other messages and in the world of collaborative document processing, users will determine the flow of events affecting a document. A new category of software, workflow applications, is emerging. Over time, workflow software, based on intelligent user scripting and easy-to-implement decision rules, will automate much of the processing of documents. As workflow tools mature, they will allow users to employ the headers, addressees, and contents of documents, to do the following automatically:

- route documents to users
- gain and track concurrence and approval processes
- update themselves, based on changes to fundamental databases elsewhere in the organization
- format themselves appropriately for different audiences and uses
- based on their latest contents, trigger other processes

Thus, workflow applications will enable the automatic construction of a manager's daily or weekly status report. They will acquire and aggregate data from databases, attach appropriate paragraph summations based on the tracking of numerical goals ("the quality division has met all of its targets; the health division has missed critical path deadlines"), alert selected individuals to individual status items, and trigger more elaborate data analysis and exploration routines to diagnose more fully the early warning indicators tracked in the status report. Workflow engines will control the behavior and sequencing of other software applications, such as word processors and spreadsheets. The combination of workflow engines and object-oriented applications will allow users to create their own meta- applications, which will provide custom functionality and will automate many routine tasks.

From application-centered to document-centered computing. If applications become reusable components that can communicate with each other, then other fundamental changes, particularly changes in our overall information-processing model, become possible. One of the most important changes is the transition from information processing viewed from the perspective of the software application package to information processing viewed from the perspective of a complex document.

In contrast to today's existing desktop model, a compound document- centered view of work (sometimes called a plain paper metaphor or docucentric computing) is more functionally appropriate to integrated information work where the task is bringing different kinds of information to a common project. In our emerging document-centered world of computing, users will launch a document (a collection of different information visualized as a notebook or compound document) to which they will bring different tools as needed.[1] The same complex document will remain open, regardless of which tool is in use. A data-sensitive and data-centered document environment will bring data-appropriate editors to the document, depending on the kind of data that need to be edited—table editors for tables, graphics editors for charts, or cartographic tools for maps.

Contrast a document-centered environment with current desktop environments. In today's environment, different applications must be opened and closed; different documents within those applications must be launched and exited; and data must be cut, pasted, or linked between documents to create one compound document that is not very well integrated. Given the kinds of complex documents users frequently seek to create and the variety of uses to which analysts, scholars, and researchers wish to put existing data, an inversion of today's paradigm is appropriate. With tomorrow's user interface for information-centered work,

[1] For one example of a notebook interface to information and computer applications, see Robert Carr, *The Power of PenPoint* (Reading, Mass.: Addison Wesley, 1991).

using a notebook or document interface will no longer involve the launching of applications through which one opens files. Rather, users will work in a world in which one opens documents or projects to which one brings small, modular tools. The data will stay open, and the tools will change, depending on how one needs to manipulate data.

From operating systems as file systems to operating systems as databases. The transition from application-oriented computing to document- oriented computing will be made possible by another fundamental transition, the replacement of today's file-oriented operating systems by operating systems with object-oriented databases at their core.

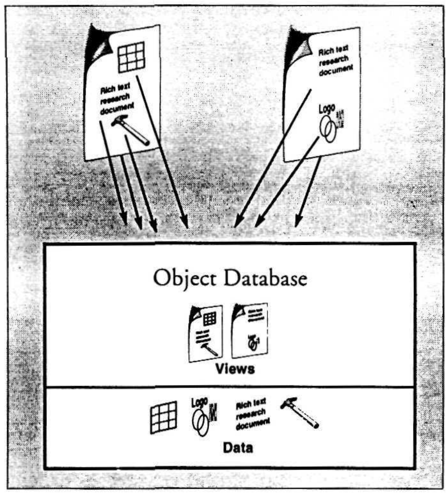


Figure 4: The Emerging Document and Database Model

Databases offer advantages that complement object-oriented applications. In database environments, a fundamental distinction is made between underlying data in the database, applications that can analyze these data, different views of underlying data (such as lists, maps, tables, and bar graphs), and end-user documents containing data views (text and bar graphs drawn from the underlying dataset, for example). Separating data from views allows data to be shared by different applications; it also allows effortless updating of data. As a data object is changed, all views based on those data change, without manual cutting and pasting to maintain currency across data and documents. Figure 4 provides an overview of how documents and databases will interrelate. In this context, documents are essentially different views of objects stored in the same underlying database. Figures 4 and 2 show the contrast between the two types of systems.

The benefits of integrating database tools into system software are many. Data and applications are separated, allowing the same data to be used within different applications and encouraging the creation of general-purpose analysis tools that can work with different datasets. Different documents from different projects can open and manipulate common data files. A standard means of exploring data and datasets is available to end users regardless of the specific application that created the data or is considered the primary owner of the data.

All applications access data through one common interface: the database's data dictionary. The data dictionary creates a standard and simple interface between the application and the data, thereby reducing programmer effort. A clear separation of data, applications, and data views exists; each different way of looking at data is cataloged in the data dictionary, which links data, applications that operate on them, and the data views (e.g., table, block of text, and graph) that those applications generate. Whole documents can be treated as views and stored in the database, too, as figure 4 illustrates. Changes in the data occurring anywhere in the environment are reflected in all subsequent views of those data. And should views relying on the same underlying data be used in several documents, all documents are automatically and concurrently updated whenever the underlying data change.

This new software paradigm will create a whole new way of looking at documents, since in many cases documents will be virtual documents, that is, they will be hot-linked to other documents and datasets, representing an ongoing process of collaborative development rather than a paper document fixed at a single point in time. These virtual documents consist of data sources—text, graphics, images and tables, and other data types—residing elsewhere in the computer network or database. Documents are constructed from them. As underlying data change, the documents that use the data also change. This will create information management challenges, since those who use these documents may not know the history or status of these changing entities.[2]

Unification of searching and retrieval tools and paradigms. As file systems give way to database systems, an even more profound opportunity will emerge for the information management community: the unification of our disparate paradigms for finding information. Today, several distinct paradigms exist for information retrieval and access.[3] Among these (and this list is meant to be illustrative, not exhaustive) are the following (see also figure 5):

- the file attribute model, which finds information based on the attributes of files such as file name and creation date.
- the fixed-field database model, which finds data in specific database fields.
- the relational database model, which finds table-oriented data based on lookup keys.
- the descriptive catalog model, which finds information based on abstracted summary and/or title information.
- the keyword model, in which blocks of information are given descriptive keywords.
- the markup/tagged-text model, which finds blocks of text, graphics, or video based on content-descriptive tags, such as those used in standard generalized markup language (SGML) data description, or database fields associated with the coding of images.
- the explicit hypertext model, in which human coders manually link terms or ideas.
- the implicit hypertext model, in which hypertext links are constructed based on common terms or on the organizational structure of documents using automated linking tools.
- the "bare bones" indexed full-text model, which locates information based on finding user-specified text strings.
- the thesaurus-based full-text model, which finds information based on user-specified text strings and on thesaurus-based synonyms.
- the statistical pattern recognition/relevancy model, also called document clustering, which finds information based on relevancy rankings to other information previously identified as appropriate.
- the semantic network model, which finds information based on related concepts, where relations may be as simple as thesaurus-based synonyms or as complex as artificial intelligence–based inference models.

[2] Ronald F. E. Weissman, "Virtual Documents on an Electronic Desktop: Hypermedia, Emerging Computing Environments and the Future of Electronic Information Management," in *Management of Recorded Information: Converging Disciplines. Proceedings of the International Council on Archives' Symposium on Current Records*, edited by Cynthis J. Durance National Archives of Canada, (Munich: K.G. Saur, 1990), 37–60.

[3] For concise descriptions of several of the more advanced research and retrieval paradigms, particularly those dependent on descriptive markup and on document clustering and relevancy analysis, see the June 1992 issue of *Byte* magazine devoted to "Managing Infoglut." Of particular relevance are articles by Christopher Locke, "Making Knowledge Pay," 245–52; Louis R. Reynolds and Steven J. Derose, "Electronic Books," 263–68; Earlene Busch, "Search and Retrieval," 271–76; and Haviland Wright, "SGML Frees Information," 279–86.

Figure 5. Search and Retrieval Paradigms

Model	Appropriate Use
File system information	Retrieves files based on title and other file system attributes
Fixed-field database	Simple, record-oriented flatfile databases
Relational database	Structured tables of simple text and numerical records
Descriptive catalog	Documents retrieved by author name, titles, or document abstracts
Keyword	Tagged blocks of information, including tags assigned to images, sounds, and video
Descriptive markup	Structured documents
Explicit hypertext	Purposefully linked information
Implicit hypertext	Information linked by automated means analysis of common key terms
Indexed full text	Exact occurrences of words in text documents
Thesaurus-based full text	Occurrences of conceptually related terms and phrases
Statistical pattern recognition/ Clustering/relevancy	Statistical occurrences of words in title, abstract, and/ or body of text
Semantic network/Entailment mesh	Conceptually linked information based on logically derived relationships

The Benefits of New Models of Information Processing

From a user's perspective, the transition from file systems to integrating databases within the operating system will allow the construction of truly integrated search-and-retrieval tools. The huge and expanding market for so-called document management systems, which essentially must reimplement their own operating systems to accomplish anything significant, is one sign that information professionals are searching for retrieval and access tools vastly better than simple file management or full-text search engines.

Unlike the regular and predictably structured inputs of traditional data processing, information-processing must deal with uncertainty, ambiguity, and incomplete information. Indeed, in the classic information-processing domains, the questions are rarely asked in a form dealt with by the tabular report generators of traditional data processing. Rather, questions take other forms: "What other information relates to this bit of data?" "How does this new bit of information change the probability of a given hypothesis?" "How reliable is this information?" "What other information, collected by others for very different purposes, can be brought to bear on this problem?" "For this hypothesis, what is the weakest piece of data in the evidentiary chain?" Beneath each of these questions are other more fundamental questions of epistemology and taxonomy which are directly relevant to the archivist: given the variety of uses to which a document, image, or artifact may be put in the future, how should it be classified and categorized to ensure maximum usability?

It is this fundamental problem confronting analysts and researchers—that documents may contain information whose relevance was not foreseen by their creators—that suggests why the unification of search-and-retrieval paradigms is so very important. Often, the most interesting research and analysis problems force us to access information that we didn't know that we would need for this purpose when we originally reviewed, classified, and stored it.

With the new underlying technologies, particularly object-oriented databases as replacements for today's file systems, a researcher or analyst will for the first time be able to ask "What does the computer network know about my problem?" Then modern software technology provide the answer by making use of the variety of search strategies appropriate to the variety of information stored in the underlying information object database. By their very design, object databases know the retrieval and display methods appropriate to each class of objects they contain. Thus, a new generation of searching and finding aids will be developed to allow all information to be accessed from a user's perspective and in a consistent way, while defining search strategies appropriate to different kinds of information. And this is the way it should be.

Unlike the structure, certainty, and regularity of data processing, information processing forces us to deal with unstructured information, information having differing degrees of probability, often deriving from ad hoc information acquisition procedures, and having varying degrees of ambiguous interpretation. Triangulation, cross-referencing, value-of-information analysis, probabilistic analysis, chains of inference, information audit trails—this is the stuff of true information processing, applied to rich collections of information using flexible and multifaceted information retrieval and filtering tools. At base, our collections must be stored using appropriate technology if we are to gain the benefits of the emerging revolution in information processing.

Conclusion: Rising Expectations for Networked Information Access

This paper has suggested that the fundamental model of information processing will change from a traditional applications-oriented architecture to a document- and database-centered architecture. Today's desktop computing applications will become object oriented, and the user interface will move away from its current application-centered focus to a focus on projects and documents. Bits of information will be uncoupled from the applications that create them and will, as component data, be reusable across a wide spectrum of applications.

To summarize, object-oriented applications will create a world in which software tools collaborate (see figure 6). This collaboration will be structured via workflow engines to automate mundane tasks; object-oriented applications will behave as tools that are brought to manipulate information using notebook and document-oriented user interfaces. Among the popular new user in-

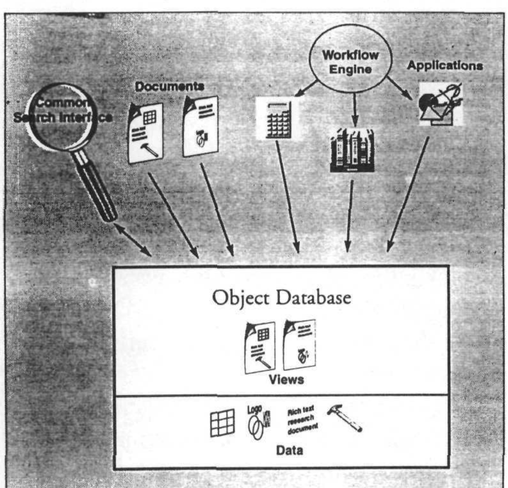


Figure 6: Emerging Information Architecture

terfaces for computer systems will be the notebook interface, in which the user thumbs through a notebook of related information and brings relevant editing tools to different kinds of information stored in the notebook. These living computer documents will be virtual entities composed of views of underlying information; information will be stored in networked databases, not in today's file systems; and, finally, these databases themselves will permit us to unify our many separate modes of finding and accessing information.

Networks are expanding in reach and power, both nationally and internationally. Innovative publishers (and this is becoming less and less of an oxymoron) are seeing themselves in the role of information providers, not as printers. Publishing, communications, entertainment, and computing, as has been observed many times, are converging. Dazzling as it is today, the range of on-line information services currently available to subscribers in North America and Europe is only a glimmer of what will be available by the end of the decade. From esoteric, specialized, one-of-a-kind custom searches of private databases, to the most familiar information—such as newspapers and magazines on line—sophisticated users and information professionals will expect to find what they need, and to find it on line, using newer storage, searching, and user-interface and information access paradigms. The point here is *not* to talk about new information services but about raised user expectations.

The desktop computer revolution forever changed users' expectations about what computing could offer and how friendly it could be, and this happened to the detriment of traditional mainframe operations lacking the interest or, in some cases, the awareness to compete as service providers capable of meeting these radically changed expectations. In 1993, the mainframe computer market shrank for the first time. The issue to ponder is to what extent the archival community will embrace enthusiastically the new information technology and provide on-line access to collections in line with users' rising expectations about how one accesses information of any kind.

As work practices in other fields come to rely increasingly on access to electronic information, traditional modes of accessing information, such as records in hard-copy form, will become less valued. This will happen because they will be less usable. Paul Strassman, former chief information officer for the U.S. Department of Defense, recently put it this way:

> I consider business files, in facsimile or printout form, the graveyard of organizational knowledge. Filing cabinets represent to me a mechanism for the destruction of reusable information assets. Perhaps as much as one half of the administrative costs of our private and public organizations are consumed in replicating, reentering, copying, retransmitting, rephrasing, restating, recoding and reinventing information that already resides in somebody's files unless it has already been discarded as garbage.[4]

I began this paper in my role as a computing professional. Let me end it by returning to my own roots as a historian of Renaissance Italy and an active user of archives. I returned a few weeks ago from my first visit to the newly built state archives in Florence, Italy. Having used the old archives for nearly twenty years, I went with a certain anticipation. How would the new, modern archives improve the quality

[4] Paul A. Strassman, *The Business Value of Computers* (New Canaan, Conn.: Information Economics Press, 190), 363.

of information access for perhaps the best collection of pre-modern historical documents in Western Europe?

The Archivio di Stato is housed in its own newly constructed building, an archival research center located far from the more historic Uffizi Gallery that served as its home since its eighteenth-century foundation. What wonders would the new archives, built some twelve years after the personal-computing revolution had taken hold, offer for the information professional?

On first glance, things looked quite promising. The reading room, for example, is now equipped with facilities allowing users to bring in their own notebook computers. But first glances were misleading. An examination of the fundamentals showed that nothing had changed. Indeed, the newest inventories produced by the archives for one of its most important collections, its notarial records, were entirely handwritten—not even typewritten, which at least could have offered the possibility of eventual scanning, character recognition, and conversion to a machine-readable and searchable format. What is the new Archivio di Stato? A fabulous Renaissance treasure trove, housed in a twenty-first century repository, using Medieval scribal methods of document management.

As the Florentine archives illustrate, technology is not self-implementing. To make effective use of the new models of information processing that are now emerging, archives not only must embrace new technology enthusiastically, as many have done, but must participate in fundamental redefinition of roles and expectations. First, users will soon assume that access to knowledge equals electronic information access. Very soon, users will assume that all well-managed information repositories, archives included, are networked information services, because electronic networks will become one of the dominant modes of accessing virtually all other kinds of information during this decade. And the role of the archivist will continue to ensure that information so accessed is appropriately contextualized in terms of provenance, history, and organizational integrity—hopefully by creating on-line documentation about core collections.

Second, users will assume a far greater degree of integration across information sources, given the new storage and searching possibilities offered by object-oriented systems and databases and by related technologies. Users will be less concerned about the traditional division of information among the various document- and collection-oriented specialties than about finding all relevant information across a wide range of documents, media, and information sources. Yes, the organization of archives into distinct types of information—that is, into collections—will continue to matter, but so will flexible access to information conceived by tomorrow's archival user in terms that do not always correspond to traditional collection-oriented distinctions. Again, it will be the responsibility of the archivist to ensure that, while flexible access to information across collections becomes easier, all documents so accessed are appropriately contextualized and documented so that flexible access does not mean a loss of integrity or documentary context. Indeed, I would argue that on-line database technology offers better and more usable contextualization tools than existing paper technology.

It is my hope that archivists will take the lead in determining how technology can significantly improve the range and depth of access to their holdings. To do so will require a fundamental rethinking of the role of archivists as information-engineering experts, and an affirmation that archives not only will continue to be places to visit but also will become networked services to access intelligently, flexibly, and appropriately from afar.

Technology Trends

Commentary

LUCIANA DURANTI

About the author: Luciana Duranti is associate professor in the Master of Archival Studies Program at the School of Library, Archival and Information Studies, University of British Columbia. Prior to this she was a professor in the Special School of Archivists and Librarians at the University of Rome, served as state archivist in the State Archives of Rome, and was project archivist for the National Research Council of Italy. Professor Duranti has earned a Doctorate in Arts and a Master of Archival Studies from the University of Rome.

RONALD WEISSMAN'S PAPER has presented a panoramic view of technological developments in the next decade which appears to be very close to a researcher's heaven. From my point of view, the developments described, if realized and implemented without appropriate administrative, procedural, and standards-related controls, would be much closer to everybody's hell, researchers included.

This fundamental difference between our positions is probably due to the fact that they are based on different assumptions. Weissman's assumptions, as revealed in his paper, are that:

1. archival material is "meaning-rich information."
2. archivists have "the fundamental task" of managing such information and documenting the history of records creation.
3. archival institutions are "information repositories" that have been for a long time "places to visit" in order to find information, and should become "networked services to access . . . from afar."
4. the relevant users of the new technologies will be researchers with the most varied purposes.
5. the main consequences of such uses will be knowledge related.

My assumptions, when thinking about technological developments, are somewhat different:

1. Archival material is the byproduct of administrative actions and transactions and is created as a means to practical purposes. Because of the reasons for its creation, archival material is impartial evidence of actions and transactions and provides a reliable account of them. By doing so, archival material also provides accountability and "enables individuals and organizations to maintain legitimate relationships of delegation and

to uphold the rights and obligations that flow from those relationships."[1]

2. Because accountability through archival material implies accountability for archival material, archivists are assigned the fundamental responsibility of ensuring its integrity, authenticity, and availability. They must also preserve its meaning in context—its evidential significance—a responsibility that Hilary Jenkinson called "the moral defense of archives."[2]
3. Archival institutions preserve "perpetual memory"[3] of societal deeds for the practical and moral benefit of the people.
4. The relevant users of the new technologies will be the records creators, mainly in large organizations.
5. The main consequences of such use will be related to administration and accountability and, in more specific terms, to the probative value and effectiveness of the documentary product of those technologies.

On the basis of these assumptions, the new information-processing model which will emerge in a decade, as described by Weissman, appears to be very similar to that of the old paper world: a model that allows for such freedom that, if the process itself is not structured and standardized, and if controls are not established to guarantee that procedures are followed (as was done in the paper world), all records creation will reach a state of anarchy. Administration and business transactions will be made more and more difficult, accountability nonexistent, and scholarly research unreliably speculative. Undoubtedly, as Weissman contends, the process by which official documents are created would be documented, but in the absence of a routine creation process for each category of documents, we would need to document the creation process for each single official document. More important, would the documented process be a reliable and trustworthy record of what happened? If yes, how so? If not, what is the benefit of having it?

If the probative nature of the evidence produced by using the new technologies is not ensured by a series of controls built into both the administrative and the information systems, law and scholarship have no use for it. Because law and scholarship respectively rule our society and explain its conduct by trying to establish the truth, they both need accurate and authentic records. In consideration of the fact that records are inherently unreliable—as they can only tell that which someone else told them—it is essential that they are provided with a circumstantial guarantee of trustworthiness; that is, that the circumstances of their creation and preservation are known and reliable.[4] Thus, both courts and researchers, like administrators and managers, need to be able to rely on an organization's procedures of creation, receipt and control of its records, and on the proper custody of them throughout their life continuum.

Examining the provenance of the records

[1] Jane Parkinson, "The Concept of Accountability and its Role in Archival Theory," Master of Archival Studies Thesis, University of British Columbia, 1992.

[2] Hilary Jenkinson, *A Manual of Archive Administration*, new and rev. ed. (London: Lund, Humphries & Co., 1937), 83 ff.

[3] That of "perpetual memory" is a juridical concept which has governed all archival endeavours from the Roman period until the end of the eighteen century, and according to which the documents preserved in archives are authentic and eternal evidence of past actions. See Elio Lodolini, Lineamenti di storia dell'archivistica italiana. Dalle origini alla metà del secolo XX (Rome: La Nuova Italia Scientifica, 1991), 53. The damage in which archives have incurred because of two centuries of partial neglect of such a concept has made archival theorists reconsider its centrality in archival science.

[4] See USC APP RULE 803 Rules of Evidence for U.S. Courts, Article 8; USC 1732, Judiciary and Judicial Procedure 115- Evidence, Documentary (about reproductions); and Canada, "Canada Evidence Act," *Revised Statutes of Canada*, 1985, c. C-5, s. 30.

makes it possible to ascertain the nature and characteristics of those procedures. But Weissman's description of "object-oriented applications" and "document-centered computing" does not address any capability related to the capture and preservation of provenancial information—that is, information on the records creator and on its structure, functions, activities, and transactions, which together constitute the administrative context of the records. Without such capability, all the archival documents created and used within the new technological environment would be useless for any administrative, legal, or cultural purpose.

It is therefore the responsibility of archivists, as part of the moral defense of archives, to intervene in the design of the new information architecture and require that the ability to capture provenance be built into it. Moreover, archivists must ensure that the new information architecture has built into it the ability to distinguish between the organic and unique evidence of action and the myriad of other types of information in the same database, so that the integrity and authenticity of archival material can be guaranteed by protecting it from all the manipulation to which the other types of information in the system are subject.

One type of manipulation with which archivists feel rather uneasy is outright destruction. There is a tendency among researchers to believe that when technology allows for limitless storage, archivists will preserve every document created, of whatever nature. Admittedly, nobody likes the idea of arbitrary destruction of information, particularly records, and for this reason it is absolutely essential that any new information system has the ability to retain automatically and without possibility of change the evidence of affairs.[5] But archival principles and concepts have taught archivists that preserving all the information generated goes against the interests of present and future users. Archivists want to be able to select for continued retention a compact, unencumbered, meaningful, complete, effective, and reliable record of societal deeds and to preserve it within its administrative and documentary context.

Weissman expects that tomorrow's debates about retention will "be as much about what classes of documents to compress at very high rates as about what to preserve or discard." This statement conjures up a situation very similar to that of the medieval period, when there were three classes of documents: (1) the *useless documents*, which were to be routinely discarded; (2) the *thesaurus* (i.e., the precious documents), which was to be kept in the treasury of the authority; and (3) the *sediment* (i.e., the documents accumulating on the benches of the offices), which was to be left undisturbed and in the end disappeared for apparently natural causes but really for lack of appropriate care. This ability of the new technology to compress documents at very high rates provides archivists with an excuse to relax their appraisal standards, with the illusion that they are not making any irrevocable decision, and ultimately with a means of avoiding their responsibilities. In fact, it might produce the largest sediment class ever. Given the unlimited multiplication of information which will inevitably derive from the use of the new technology, the principles and methods governing appraisal do undeniably need to become more strict and rigorous. Actually, the entire approach to the management of

[5] That the retention be "without possibility of change" is essential, because "records have evidential value precisely because they have an element of stability." Frank Upward, "Challenges to Traditional Archival Theory," in *Keeping Data: Papers from a Workshop on Appraising Computer-Based Records*, edited by Barbara Reed and David Roberts (Sydney: The Australian Council of Archives and the Australian Society of Archivists Incorporated, 1991), 106.

archival documents must become more rigorous, not only to protect the rights and interests of the people but also, more specifically, to provide the best service to archival researchers in both current and future archives.

At this point, it seems appropriate to move from the use of the new information technology for the creation, maintenance, and preservation of archival documents in electronic form to its employment to support archival functions within archival repositories, which is really the main focus of Weissman's paper. Weissman expresses his hope that the information architecture of the new technology will "significantly improve the range and depth of access" to archival holdings. However, he rightly remarks that technology is not self-implementing, and that archivists must embrace it consciously and enthusiastically. I believe it is equally necessary for archivists, before trying to implement any electronic information access system, to determine and consistently apply the principles, methods, and standards routed in archival science which must guide their efforts to make archival material accessible to all kinds of users. This implies a need, much stronger than before the advent of new technologies, for a rigorous and standardized education in the concepts of the science and discipline of archives.

This also implies that archivists must take a much firmer stance on what their role is, and make users understand that by *not* taking up the role of "information-engineering experts," they are acting in the best interest of the users themselves. Archivists share with users a sensitivity to the need for "flexible access to information," as Weissman puts it, but they also understand that the primary need of any user is access to reliable evidence, and no information constitutes evidence if it is deprived of structure. In fact, records both have structures and are created within structures; they have value as an information resource because of those structures. As Terry Eastwood writes,

> the assumption behind much recent thinking is that 'the new technologies' will negate the need of any comprehensive account of records structure. Everything becomes the query and its satisfaction. This reduces archival documents to a question of their content. Their meaning in structure, their place in a complex of documents recedes, their attachment to specific objects and procedures may be lost, and with it the concern to give the world knowledge of how records came into being.[6]

Archivists have an obligation to communicate archival material in its documentary and contextual integrity, in order to preserve the impartiality of the records (i.e., their characteristic of not purposely serving any particular use other than and after that for which they were created). As Jane Turner notices, "the Jenkinsonian notion of the moral defence of archives reflects the primacy of the probative nature of the records and directly links the notion of impartiality to use."[7] The usefulness of records is directly dependent on the preservation of their archival nature and of their structure through the application of the principle of provenance, which guarantees the respect of the external and internal integrity of each archival group. When describing archival material to make it accessible to researchers, Michael Cook wrote, the archivist provides "a statement on meaning and authenticity . . . ensures that the evidential

[6] Terry Eastwood, "Provenance, Structure, and Content in Archival Information Retrieval," unpublished paper presented at the Seventeenth Annual Conference of the Association of Canadian Archivists, Montreal, 12–15 September 1992.

[7] Jane Turner, "A Study of the Theory of Appraisal for Selection," Master of Archival Studies Thesis, University of British Columbia, 1992, p. 119.

meaning of the archives be understood for ever afterwards,'' and ''protect(s) and demonstrate[s] the significance of the information content.''[8]

Of course, there is nothing wrong with using new powerful technologies to manipulate the information contained in archival descriptions or in the text of archival material for exploitation purposes, as long as provenance is understood at all times. However, the image of archives as ''networked services to access . . . from afar'' is very disturbing. Archivists must make a serious effort to explain to technology experts and researchers that archives are not like any other information service. Their role in every society, but especially in a democratic society, is so intertwined with the legal, political, social, and economic system, with the relationships between the government and the governed, that their documentary content cannot just be left to disseminate freely in electronic ubiquity without proper controls and guarantees. Nor can it associate with information from non-archival sources without proper distinctions being made. An analysis of the consequences of such actions would frighten researchers even more than administrators.

On the contrary, a special effort should be made to establish and maintain access to archives through archival networks so that information about archival material preserved in many different archival repositories can be easily available to researchers living in geographically distant places. With respect to electronic records in particular, networks should be created involving both archival institutions and records creators to allow for direct access to records from far away. This would, of course, require the establishment of proper procedures not only to ensure the preservation of the integrity and authenticity of the records and their administrative and documentary context but also to guarantee confidentiality and privacy.

The creation of this kind of network would facilitate another development seems inevitable: the end of centralized archival repositories. In a recent study, Charles Dollar suggested that archives will become ''repositories of last resort'' for archival material at risk.[9] In such a scenario, the archivist's role of mediator between records and users (both records creators and researchers) would be enhanced, but for this to happen the existence of agreed-upon standards will be essential, and a solid education in the science and discipline of archives will be vital.

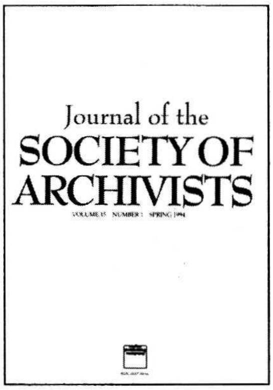
In conclusion, I believe that, confronted with the developments described by Weissman, archivists must resort to the knowledge embedded in their science and discipline, and use it (1) to participate in a knowledgeable and effective way in the design of archival technology and (2) to guide organizations in establishing the procedures and controls necessary to use new information technologies without compromising the effectiveness and trustworthiness of their records. After all is said and done, the first duty of the archivist is to the records' integrity and impartiality.

[8] Michael Cook, *The Management of Information from Archives* (Brookfield, Vt.: Gower, 1986), 81.

[9] Charles Dollar, *Archival Theory and Information Technologies* (Macerata, Italy: Università degli Studi di Macerata, 1992), 54.

CARFAX PUBLISHING COMPANY

Journal of the Society of Archivists



EDITOR
Chris Webb
Archivist, Borthwick Institute of Historical Research, University of York, UK

DEPUTY EDITOR AND REVIEWS EDITOR
Philippa Bassett
Archivist, Birmingham Central Library, UK

Journal of the Society of Archivists deals with all matters of interest to archivists, conservators, records managers and those involved in any of the historical disciplines, whether on materials, conservation, storage archival procedures or the study and interpretation of records.

Published in association with the Society of Archivists, this journal deals with the very latest developments in this field, including the archival challenges and opportunities of new media and modern information technology. It also relays invaluable information about training opportunities and developments in administration techniques for archivists, conservators and records managers.

Volume 15 - 1994 (2 issues). ISSN 0037-9816.

Subscriptions
Institutional rate: EC £38.00; Outside EC £44.00/US$76.00, post free.
Personal rate: EC £19.00; Outside EC £22.00/US$38.00, post free.

ORDER FORM

Please invoice me at the ☐ institutional ☐ personal rate

☐ Please send me an inspection copy of ***Journal of the Society of Archivists***

Name ____________________

Address ____________________

____

____

CARFAX PUBLISHING COMPANY
PO Box 25, Abingdon, Oxfordshire OX14 3UE, UK
PO Box 2025, Dunnellon, Florida 34430-2025, USA
UK Tel: +44 (0) 235 521154. Fax: +44 (0) 235 553559.

Technology Trends

Commentary

JOHN McDONALD

About the author: *While with the National Archives, John McDonald has served in a number of capacities related to the archival management of electronic records. During this time he also served on the Treasury Board Task Force responsible for the implementation of the Access to Information and Privacy legislation. Recently he assumed responsibility for a program that assists institutions in the management of their records, regardless of their physical form. A particular area of interest is the management of information in office systems. He received a B.A. in history and geography and an M.A. in geography from Carleton University, Ottawa.*

IN THE FIRST PARAGRAPH of her commentary, Luciana Duranti cautioned us about what the future could hold as a result of the introduction of the new information technologies described by Ronald Weissman. According to Luciana, unless controls are in place, the emergence of these technologies could lead to a researcher's hell rather than a heaven. For my part, since we appear to be neither in hell nor in heaven, I can only assume that we are in some form of limbo. That is, there has yet to be a judgment made as to whether our use of new technologies is so atrocious that we should be condemned to a world without context or meaning—a world inhabited by data rather than records. Nor, however, have we been allowed to pass through the pearly gates to a heavenly world of information in context; of records that are available, understandable, and usable; and of people (or perhaps angels) who are enlightened and enriched as a result of the sources of knowledge available to them.

There are three points I would like to make in my commentary. First, it is not just archivists who are concerned about the preservation or survivability of records, or what I like to refer to as *corporate memory*. There is a growing concern among organizations that they are in danger of losing the recorded memory they need to support their business and hold themselves accountable.

My second point is that the new technologies that Weissman has described have the potential to serve as the technical solutions that will facilitate the preservation of memory, regardless of whether that memory exists at the level of the individual, the work group, the organization, or the society as a whole.

Third, there is much that we as archivists can bring to the developers and implementors of these technologies, but we can accomplish only this by being clear about our requirements and our goals.

Sensitivity to the preservation of context and evidence—the "record"—is on the increase, but to understand this sensitivity, one must understand the rapid evolution that has taken place in the design and use of the computer. Many of the technologies that Weissman has described were designed by enterprising individuals who wanted to develop a *personal* computer. Originally, these computers were designed as *personal* support utilities to help individuals by automating tasks associated with, for example, the preparation of letters, memos, and reports or with the design of small databases. Meanwhile, on the other end of the spectrum, organizations have for over thirty years benefited from the *corporate* support utilities associated with mainframe computers that support the automation of major business functions, such as payroll, tax collection, and licensing.

During the past few years, however, we have been witnessing the convergence of the personal and the corporate worlds as the automated tasks of the individual user have been linked together into automated processes that, as applications, directly support the business functions and activities of organizations. In the stand-alone world of the personal computer, what we store in our directories has generally been the results of automated tasks that led mainly to the production of hard-copy documents, the transmission of which was governed by whatever corporate procedures and rules happened to be in place for paper records. But Weissman has described a different landscape, one where the flow of information through most modern organizations will be increasingly electronic. As these electronic highways are built, users are discovering that there seem to be few, if any, rules of the road. In fact it is difficult to find anyone who is prepared to develop the rules of the road and even harder to find people who could enforce them.

Given this landscape, records creators, and those who are distanced from the records creation process but who are concerned about the integrity of the records in an organizational context, are raising questions regarding their status and management. What is a record? How should accountability be assigned for its definition and care? How can authenticity for electronically transmitted information be established? When should something be filed and why? What criteria should be used in making the decision? Where should a record be stored? How should it be protected? Who should look after it? What tools and techniques should be used to manage it? These and other questions are being raised because users are concerned about being able to get the information that they need to do their work, about ensuring that their actions are accounted for in case they are challenged, and about being able to trace the evolution of decisions in order to help substantiate a case or support the development of new arguments and proposals. Essentially, users have an intuitive sense (often based in the culture of the organization) that evidence of certain transaction must be kept; they are just not sure when, or what, or—in precise terms—why.

The potential to address these concerns exists in many of the technologies that Weissman has described. For instance, he discussed the new automated routing systems and workflow control systems that are appearing on the market. I believe organizations will be able to incorporate rules and procedures into the design of these systems so that the record is identified, captured, and preserved. But this will only occur when the organization begins designing applications that permit documents to be transmitted electronically through various review and approval levels. Unlike the ad hoc transactions associated with our current e-mail systems, the automation of corporate workflows, such as the development of policies and the preparation of responses to executive correspondence, will generate almost automatically the need to capture evidence of key transactions. The problem is that organizations lack the people with the skills and knowledge required to determine how this should be done. Archivists, given their perspective, are in the best position to fill this important gap. As Weissman points out, "a future opportunity for the archivist will not only be to store and document the official record of organizations; but, given the versioning tools soon to be at our disposal, to document the process by which records are created." It remains for the archivist, however, to spell out the requirements.

Another example of an area in which the tools are already available is the area of storage. I believe the keep-everything issue will be a nonissue for most organizations. The private sector is famous for disposing of records that it no longer needs in the course of its business so that it cannot be brought into court on the evidence contained in its own records. And public sector organizations rid themselves of personal information to comply with the retention and disposal requirements imposed by privacy legislation.

On the other hand, within the context of the retention and disposition standards that institutions ought to have established, there are all sorts of technological ways of storing electronic records, as Weissman has illustrated. I just wish he had emphasized the fact that the potential exists for these tools and techniques to be made available to help organizations ensure that they can comply with their own retention and disposition standards. Before these can be put into place, we need criteria to help us decide what, as a record, should be kept. Should we keep the pool of objects and count on technology to help us recreate the views that will provide the evidence of the transactions we are concerned about? Or should we keep all of the views? And how much contextual information should we keep, particularly when it is scattered around a host of, as Weissman calls them, "integrated object-oriented applications"? And we haven't even mentioned the degree to which we need to keep the original user interfaces and functionality associated with the record, regardless of how it is defined. How can we deal with continuously updated object-oriented systems? Again, the tools are either available or are emerging; what we need are the requirements.

In the area of retrieval, much has been said about the valuable role that full-text retrieval systems can play and the extent to which certain kinds of classification schemes continue to be necessary. I am a big fan of full-text retrieval tools, particularly those that support sophisticated search analysis capabilities. My concern, however, is with the contextual information that is available to ensure that the records, once retrieved (and regardless of the way in which they were retrieved), are available, understandable, and usable. I believe that the tools and techniques that have been described can provide us with the opportunity to capture more contextual information than we could ever dream of capturing in the paper world. Again, however, we need to define our requirements. What does a policy officer need to help him or her develop policies? Surely they would not retrieve earlier memos, correspondence, and reports without understanding the context in which they were prepared and used. Records creators and users are just as concerned about context as are good researchers. In many organizations, for instance, the development of information locator systems, corporatewide thesauri, authority and vocabulary control tools, and other related instruments is becoming a priority, both as a means of facilitating access and retrieval and as a means of ensuring that, once retrieved, the record is understood in the context of its original creation and use.

Descriptive standards are as important to organizations as they are to achivists, and we need to draw more heavily on the experience of records creators as they grapple with the development of descriptive tools.

What archivists can bring to this complex world is context. But we need to express our requirements—both corporate and archival—if we are to accomplish the objective of putting into place the policies, standards, and practices that will enable organizations to have records that are in context and that are available, usable, and understandable. Where do we begin? Although it is not my intention to present a long list of requirements, I have given some thought to some of the goals that should be addressed. The following are intended to serve as catalysts for further discussion:

- Records should be created and maintained in a manner that ensures their quality, relevancy, integrity, accessibility, and currency.
- Records should be described and organized in a manner that preserves knowledge implicit at the time of collection, facilitates linkages among related records, and allows for efficient retrieval and dissemination within the context of the administrative procedures that governed their creation and use.
- Records should be protected from unauthorized access, disclosure, and alteration and from deterioration and loss.
- Records of archival value should be identified and stored in a manner that ensures their preservation.
- Records not of archival value should be discarded or otherwise disposed of when no longer required, in accordance with approved disposition authorities.

Sometimes it feels as though no one is listening—as though the technology is outstripping our capability to manage it, especially when it is so transformative. We feel frustrated by our lack of knowledge and expertise and by the inexperience of professional disciplines that should be helping us. But organizations are beginning to respond. As an example, the information systems community is shifting its role from one of control to one of service. The more innovative members of this community are becoming the standards setters for their organizations and, within the context of corporate policies and standards, are becoming facilitators to help users meet their requirements (including those that address the management of corporate memory). As a result, the climate in many organizations is ripe for the establishment of cooperative ventures that can include archivists.

The future is not bleak. What Ronald Weissman has described appears to complicate our lives and threaten our ability to preserve our recorded memory. Nevertheless, the technological trends he has outlined hold the building blocks of the solutions that we will need to develop in the future. It remains to us as archivists, operating in cooperation with others who share similar concerns, to express the requirements that will make these solutions a reality and turn our limbo into a users' heaven.