

DATA BASE MANAGEMENT

STRATEGIES FOR AN ARCHIVES MANAGEMENT PROGRAM

Michael J.D. Sutton

INSIDE

Key Business Drivers; Archives Management Problem Statement; Information Preservation Strategies;
Archival Management Process; Management Strategies

INTRODUCTION

Few professionals have navigated through few archives management (AM) projects, teams, successes, and failures. Why? Because there have been few successes to point to, and because there has been no poignant business reason to proceed with such initiatives (at least until recently). Nonetheless, many corporations and public sector organizations are now experiencing significant pain and financial liability from the loss of their corporate memory. The lack of archives management strategies is holding back “well-intentioned” but untrained project directors and managers from coming to grips with the issues, concerns, problems, and obstacles associated with an archive repository (AR). The overall strategy outlined here could well be an important reference document for one’s next initiative.

FOUNDATIONAL CONCEPTS

Why does one need a strategy? One needs a strategy because the AM initiative one has been asked to embark upon has no anchors to hold anyone to the *present*. Archives management is about time, and the effect time has

PAYOFF IDEA

The proper planning, design, construction, and use of an archives repository for an enterprise can successfully position a firm to avoid “corporate amnesia” — the loss of corporate memory. Since archives management initiatives are still new, very few skilled and experienced project directors and managers have any material or business rationale to draw upon in constructing a long-term archives repository. Many project managers will waste a considerable amount of time just “cutting their teeth” on their first or second unsuccessful archive management initiative. By acquiring access to the principles, concepts, and strategies highlighted in this article, the project manager will be assured of a higher probability of success for architecting a long-term corporate memory for the enterprise. A succinct but relevant bibliography is provided to start the reader in the right direction.

on artifacts created in the *present*. Archives management is, paradoxically, about the *future* (which most people cannot see unless they profess to be prophets), but only when that *future* has become the *past*. AM does not incorporate in its conceptual model a classical linear timeline; it presents a reverse timeline where the *past* only becomes valuable again when it is many years old.

Suppose one is asked to plan an archives repository in such a way that information objects created in the *present* can be “read” (or at least understood) in the *future*, within their original, rich context. However, someone (or some automated system) in the *future* must, at its *present* moment, contend with information objects from the *past*. The information objects at that point in the *future* never seem to have enough context or content to permit their proper understanding.

Archives management, especially as it is defined in digital terms, is a very young discipline. The key business drivers and benefits are not yet well-defined or clearly mapped. The project director has no means of testing whether information objects created as an output to an AM initiative will be usable in the future, other than through simulated testing. This testing cannot take into account software product or system evolution over a five- to ten-year period, or even worse, over a 25- to 100-year period. The best a project director can do is make an educated guess, and hope that when five or more years have passed, he may have been a prophet in his own land.

Although there is little to go on at this time to justify an AM initiative, there is anecdotal material that points to business drivers and benefits. There are **key business drivers** that can justify an AM initiative, including:

- an antiquated or traditional records management system that cannot cope with the emerging challenges and requirements of a digital workspace
- increasing online space requirements for current operational databases, data marts, and data warehouses
- short- and long-term technological obsolescence
- an enterprise’s drive to harmonize all systems onto a heterogeneous environment
- an inability to find, locate, catalog, and use current or historical digital information objects
- new uses for “old” data from sales, human resources, marketing, facilities, engineering, and financial databases

There are numerous potential benefits in implementing an archives management program that may apply to an organization, including:

- decreased corporate exposure during audits or legal inquiries
- increased compliance with various levels of laws and regulation (international, federal, state/province, county/region, and city/municipal)

-
- improved records management and control
 - appropriate retention and timely destruction of business records
 - streamlined and rationalized processes where business owners do not have to be concerned with archiving information objects from their current systems

Take the Year 2000 issues and problems that most organizations are currently contending with, for example. These were a significant archives management problem. As automated systems grew in the 1960s and 1970s, no one could predict how memory costs would fall or how software would handle date-related information. In fact, most people who worked on such systems never expected COBOL, PL/I, RPG, or even Assembler compilers to last more than five years. Alas, one is now in the *future*, and trying to cope with the professional inability to plan for legacy applications that were just too poorly documented to re-write, and too integral to the business to throw away.

An AM initiative differs substantially from a contemporary information system project. The AM initiative can be described by a number of characteristics:

- generally five to seven years in length
- an average team size of five members for an enterprise of 1000 client users
- historical focus on record and file structures that are not currently in vogue
- primarily concerned with the disposition of the digital records of the enterprise after they are at least seven years old

Contrast this with the characteristics of a contemporary information system project:

- generally less than two years in duration (and preferably less than one year)
- an average team size of seven to ten members for an enterprise of 1000 client users
- *avant garde* focus on record and file structures that are just currently in use
- primarily concerned with the creation and use of digital records of the enterprise created in the *present*, less than two years from their date of creation

There is a significant difference in the business goals and objectives of the two projects. In addition, the owners and technology supporters of the current system projects have been allocated no time or budget to worry about or plan for the eventual retirement or disposition of the sys-

tems, data, or documents. These *present-focused* business and technology leaders have incentives, i.e., bonuses, dividends, or stock options, based on what they help the company accomplish *today* or *tomorrow* — not five to seven years from now.

The tools used within an AM project must contend with both old and new technologies, software, media devices, methods, file structures, record formats, etc. The tools of an AM initiative are Janus-like in their employment. (Janus was the classical Roman god who had two faces: one looked into the past while the second looked into the future. Janus was identified with doors, gates, and beginnings and endings). The AM project team must contend with having one foot in the *past* and one in the *future*, while the *present* passes them by.

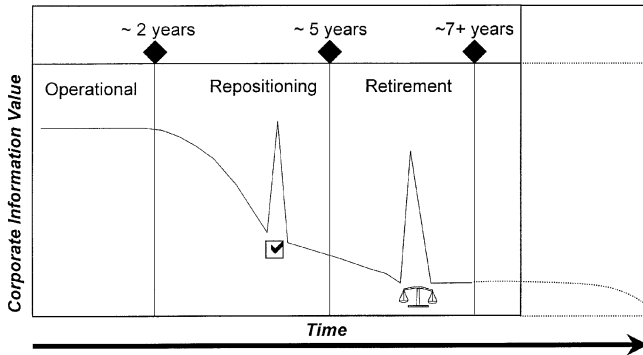
Information management professionals — at the behest of private and public sector leaders — have worked for over 40 years to automate (digitize) less than 20 percent of the enterprise information assets and intellectual capital. Regretfully, in doing so, they have forced most enterprises to be short-term in their vision of the value of information objects and systems, to rely too heavily on the digital nature of the information. Thus, many organizations almost ignore the digital as well as the hardcopy storage and long-term retrieval value of information. These enterprises may not have “lost their minds,” but they have lost their corporate memory. When an audit or legal discovery process tries to contend with digital information that is five to ten years old (or even three years old), there are significant problems in reconstructing these information objects and their context. And for many enterprises today, this “corporate amnesia” results in hefty legal costs, fines, and lost revenue.

Archives Management Problem Statement

Most organizations that wish to survive in today’s aggressive business climate must continually improve, evolve, and reposition their market, products, and services, and aggressively maintain their profitability. The enterprise must create and achieve short-term business goals, objectives, and strategies. Many businesses have invested heavily in automated legacy systems. These systems may be nearing or have already passed their retirement date. The valued corporate memories in these systems may be incorporated into new, emerging corporate systems; or, alternatively, the digital information objects may be stored offline to comply with legal and regulatory requirements. The worst case is that these out-of-date systems, data, and document objects are ignored until it is too late to determine how they could be migrated to physical media or logical format to be useful.

All enterprises are facing a challenge to preserve digital systems, data, and documents over short-, medium-, and long-term periods of time. Some companies have lost significant income from audits because substantiating data and documents (stored digitally) were lost years earlier;

EXHIBIT 1 — Information Preservation Stages



but not discovered until it was too late. The business owners of current operational information systems are generally preoccupied with operational challenges, issues, and concerns surrounding applications performance and availability. Managers are asked to contend with short-term business objectives and present problems, not with hypothetical problems that could emerge in the distant future.

The oversimplified diagram shown in [Exhibit 1](#) might help to illustrate the preservation challenge that one faces. Current information systems create or acquire information objects that have immediate corporate value. This is the operational stage of a corporate information value life cycle. The operational stage may encompass one to two fiscal years. The requirement during this stage for instant retrieval dictates that the information objects are stored online. This stage might also encompass data marts and data warehouses that require relatively quick access and retrieval response times.

The corporate value of the information will decrease over time because it simply is no longer current; and current information is very important when generating profits or increased stock price value. The information objects migrate to the repositioning stage of the corporate information value life cycle. The repositioning stage may incorporate three to five fiscal years of information objects. The use of near-line or offline storage media (optical disks, CD-ROMs, QIC tapes, etc.) generally applies at this stage, and may continue for up to five or six years. Instantaneous retrieval is not as critical at this juncture in the life cycle. Nonetheless, during this period of time, external audits by regulatory or tax agencies may take place; this creates a peak in the perceived corporate value of the information where its value may be as much or more than its original corporate value.

Generally, information objects during this period of time exhibit certain characteristics of aging, which include:

- the loss of the original information and its context
- the original media may be unreadable or devices to read it are no longer manufactured
- the original file and record structure may be unknown, foreign, or undocumented
- any significant business rules stored in the software programs are unreadable, unusable, or unknowable
- any individuals who may have worked on the original system or database have moved on to other opportunities

In short, corporate amnesia exists.

Finally, the corporate information moves into the retirement stage of the corporate information value life cycle. Information objects can remain dormant here for six to 20 years, depending on legal, historical, and archival requirements. The information objects are loaded onto offline media such as magnetic tapes, DAT (digital archive tapes), or possibly CD-ROMs. Historians, social scientists, or lawyers are the expected users for these seldom-accessed information objects. Nonetheless, the information objects must retain enough content, logical structure, and context to be more useful than just a historical artifact.

A business framework or archival facility has rarely existed in present-day enterprises to accommodate the corporate information life cycle as described above. Thus, a digital archives repository rarely exists or is available in today's public and private sector institutions to use as a model. A digital archives repository is required to contextually preserve the database objects, document objects, business rules, system documentation, and other descriptive data (meta-data). Such a repository must accommodate information objects over a medium- to long-term period of time (i.e., from five to 20 years) in a software- and hardware platform-neutral format (e.g., that proposed in ISO Reference Model of the Consultative Committee for Space Data Systems White Paper, *Open Archival Information System (OAIS)*). The archives management strategy might help to position this commitment to a repository.

FOUNDATIONAL DEFINITIONS

Before describing an archives management strategy (AMS), some basic vocabulary intrinsic to archives management is required. An archives management program consists of specific information objects, facilities, processes, and preservation domains. One can start with the information objects.

At the fore of a good AMS is the *business record* — a by-product of a business transaction preserved for future use as evidence of transact-

ing business. A business record must maintain a number of preservation characteristics. Charles Dollar, in *Ensuring Access over Time to Authentic Electronic Records: Strategy, Alternatives, and Best Practices*, proposes a number of preservation characteristics that help preserve the legal integrity of a business record while stored and archived; these characteristics include:

- *authentic*: the measure of the reliability of a record (i.e., its ability to remain unaltered, unchanged, and uncorrupted)
- *encapsulated*: the measure of the self-referential linkage of logical components in a record
- *identifiable*: the measure of the specification of unique identification boundaries in a record
- *intelligible*: the measure of the integrity of the bit stream represented in a record
- *readable*: the measure of the integrity of the bit stream device processing of a record
- *reconstructable*: the measure of the quality of rebuilding the same structure and intellectual content of a record
- *retrievable*: the measure of the capability to locate objects and parts of a record
- *understandable*: the measure of the quality of the context of creation of a record

Business records are aggregated into files. An *operational file* is an information object that contains information of immediate, instantaneous interest to a reader. Digital copies of operational files are normally created in the course of data management procedures, and are referred to as *backup files*. A *repositioned file* is a specially formatted copy of an original operational file that can be retrieved through near-line storage media instead of online storage media. For longer-term storage, repositioned files are migrated to retired files.

A *retired file* contains information objects that would be the foundation for reconstructing an authentic instance of a file in its original format and context. A retired file can encompass system, data, and document objects. The information content can be conveyed to the user or another computer as audio data, bit map data, data fragments and databases, spatial (geographical), spreadsheets, text, vector data, and video data.

The files are managed by different facilities (as illustrated in [Exhibit 2](#)). An *operational facility* is accountable for managing the current information assets of an enterprise. The information assets are digital and are stored online for immediate access. A *repositioning facility* manages the nearly current information assets (i.e., information that may be between two and five years old). The information objects are stored on near-line media for near-instant access. A *retirement facility* manages the dormant

EXHIBIT 2 — Time-Based Migration of Information Between Current and Successive Stages

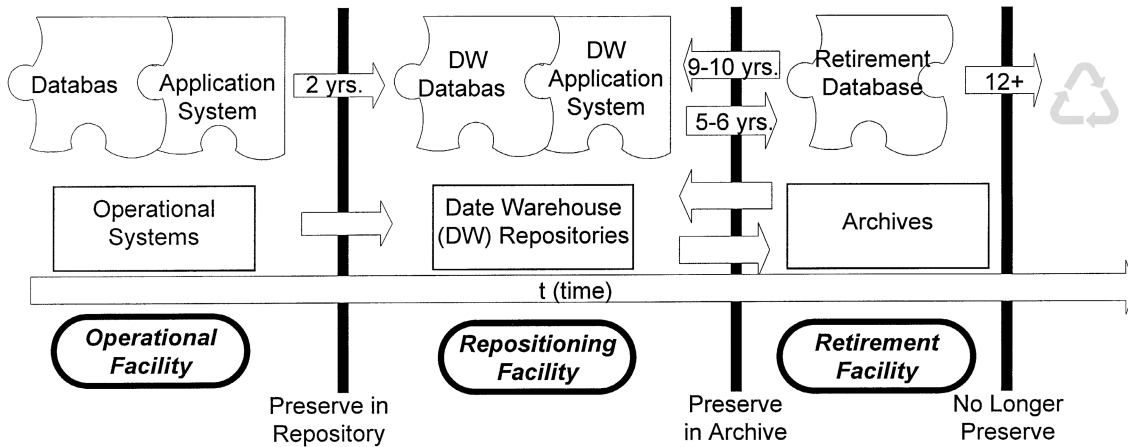
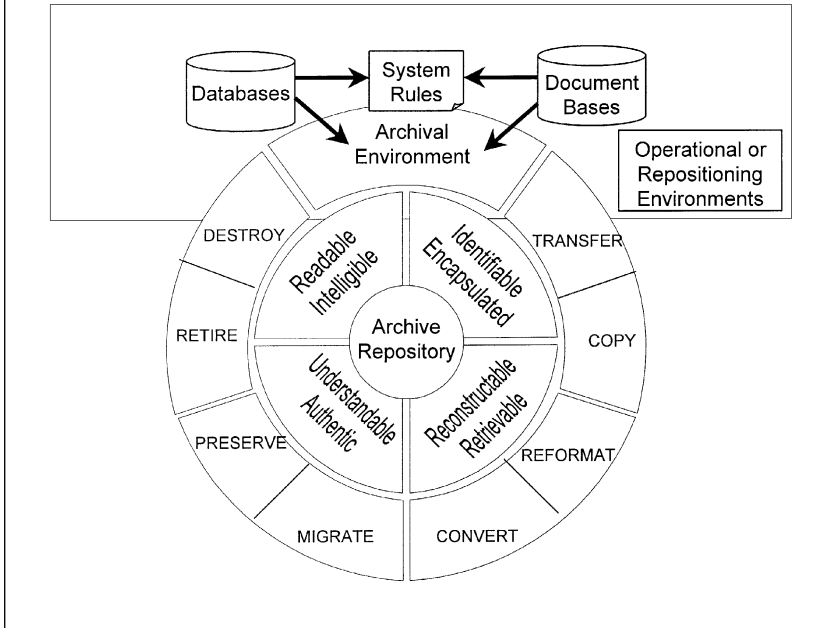


EXHIBIT 3 — Archival Management Processes



information assets (i.e., information that may be between six and 25 or more years old). The retirement facility can store the information objects in a hardware- and software-neutral format to diminish the problems of technological obsolescence. All the facilities execute specific processes upon the information objects under their control (see [Exhibit 3](#)).

A number of specific processes can be executed on the operational and repositioning environments to move them into an archival environment, including:

1. *convert*: importing or exporting of records from one software-dependent environment to another while ensuring the preservation of structure, content, and context
2. *copy*: the creation of a digital binary twin of the original file
3. *destroy*: physically disposing of the media and information objects so as to make them totally unreadable (i.e., degaussing, crushing, melting, etc.)
4. *migrate*: moving authentic electronic records from legacy information systems in online systems to another storage media, such as near-line or offline storage, while preserving the logical view of the original records

-
5. *reformat*: transferring records from one medium to another without alteration of the bit stream of a record (e.g., with no change in appearance, content, or logical structure)
 6. *retire*: moving information objects to very dormant offline storage to protect and preserve records from corruption and make them relatively inaccessible
 7. *transfer*: repositioning electronic records from online or near-line storage to offline storage for infrequent retrieval

Finally, there is *metadata*. This is specific descriptive data about particular data objects which increases the precision in recalling information objects from a search of a repository of data, document, or systems. *Metadata* can describe an information object with fields such as author, business unit of the author, creation date, modification date, security classification, or subject. This overview of the emerging vocabulary of AM will serve as a foundation for subsequent discussion.

STRATEGIES FOR AN ARCHIVES MANAGEMENT INITIATIVE

The following set of strategies and guiding principles are proposed for constructing a framework to design and successfully engineer an archives management program.

Strategy 1: Develop a Repositioning Facility

The repositioning facility will depend on available budget as well as corporate technical resourcing, knowledge, and support. In addition, the enterprise will require skills in data mart and data warehouse design, and an informed and experienced digital records management group. The repositioning facility will ensure the availability of a corporate memory for between two and five years. The operational facility feeds the repositioning facility after two or three fiscal years have elapsed for the operational data.

Strategy 2: Develop a Retirement Facility

The retirement facility must construct a long-term program, which most enterprises are not willing to fund. This retirement facility will need long-term financing so that it is not cut from the budget during lean times. This is a corporate commitment to preserve the memory of the institution. This business unit will be responsible for the archival activities associated with appraisal, collection, migration, protection, reformation, retention, and, finally, the destruction of digital information assets. The retirement facility will warrant that valuable corporate memory is available and accessible. A corporate records management program may already handle many of these functions for hardcopy records. But beware, most records management staff are ill-equipped to cope with managing

digital records, especially over such a long term; a new business unit may need to be defined in the organization.

Strategy 3: Employ Guiding Principles in Architecture and Design

The following guiding principles are proposed to jump-start the design and deployment of an AMS.

Principle 1: Manage long-term information assets in the same way one manages short-term information assets. Most organizations ignore the problems and challenges of storing information assets over the long term. This is metaphorically similar to the problems that industrial and manufacturing sectors are experiencing with pollution. If ignored in planning and design, it will cost 100 times as much to reconstruct information assets, or pay the fines and costs of mismanaged archival data.

Principle 2: Business owners of the information assets are concerned with the present — not the future. Business owners should not be made responsible for ensuring medium-term repositioning and long-term preservation of their information objects. They hardly have enough time to keep their current, operational systems performing within acceptable operational limits and backed up for disaster recovery. If the enterprise forces the operational facility to worry about repositioning and retiring, then these activities will never be done. (How many people have omitted backing up their hard disks when they first started on a PC? — my point exactly.) A separate set of facilities (with their own goals and incentives) must be brought into play to relieve the stress and pressure exhibited in the operational facility. This can even be outsourced to companies that are starting to sell services for digital records and archives management facilities.

Principle 3: Aggressively pursue the preservation of critical information assets. A digital retirement facility must be mandated to proactively collect and acquire the critical information objects that need to be preserved. Otherwise, corporate amnesia is guaranteed. Corporate amnesia is expected to increase as companies expend incredible budgets on the Y2K problem instead of on more significant areas such as enterprise document management systems. Neglect and incompetence invalidate any warranty for corporate survival.

Principle 4: Destroy archived information assets in a timely manner. Maintaining access and availability of all digital objects over a long-term period is both impractical and expensive. Assets must be categorized according to their records retention schedule to facilitate their timely destruction. Periodically evaluate the retention periods of archival

information assets and destroy the records according to their legal requirements before they can be used against the institution in legal proceedings. The only thing worse than corporate amnesia is “photographic recall” — very risky in any court proceeding. Nonetheless, there may be corporate records of a historical nature that should be preserved longer than their normal destruction date. Make sure that only the historical records are preserved.

Principle 5: Assess compliance with all legal and statutory rules and regulations. Some business unit must be made responsible for maintaining a comprehensive checklist of relevant laws, statutes, and regulations. Often, the knowledge about retention is scattered among different business units; no one is really accountable for knowing or enforcing the legal and statutory requirements. This must change, and accountability must be assigned. There are too many financial liabilities that can cost an institution if these requirements are ignored.

These guiding principles should help in creating a firm foundation for AM activities.

RECAP

This article has presented a vocabulary, broad strategies, and guiding principles for embarking on an archives management program. The strategies revolve around the creation and staffing of a repositing facility and a retirement facility. The existence of these two facilities will relieve the operational facility of the stress and pressure to try and find a short-term solution for a long-term problem. The solution is not really within their mandate or skill set to achieve. Separate business units must be “ano-int-ed” with these tasks.

Beware of the lack of funding and commitment when starting an AM initiative. Because of the short-term thinking prevalent in most organizations, these would be the types of facilities that would be cut from the organization when times get tough. If the enterprise sees corporate value and the substantial benefits of an AM program, then it must be committed to keep that program intact. An archives repository with gaps is a corporate memory with missing fragments — it may take a great deal of effort to understand, and there will be gaps that could leave out important data, legally required data. Can the enterprise afford that?

Michael J.D. Sutton, Adm. A., CMC, ISP, MIT, directs the Business Process and Document Management Services Group of a leading, international systems consulting firm, {EMail: michael.j.d.sutton@usa.net}. He has participated as a member of numerous industry specification groups and international ISO committees and is author of the book *Document Management for the Enterprise: Principles, Techniques, and Applications* (ISBN 0-471-14719-2). His e-mail address is michael.j.d.sutton@usa.net.

Bibliography

1. Barry, R., *Best Practices for Document Management in an Emerging Digital Environment*, <http://www.rbarry.com/UKRMS1/UKRMS1.html>.
2. Barry, R., Electronic Records And Objects Circa 2001, *University College London, School of Library, Archive and Information Studies Conference Proceedings — 75th Anniversary Celebration*, 1994 <http://www.rbarry.com/CON-PAP2/CON-PAP2.html>.
3. Barry, R., Electronic Document and Records Management Systems: Towards a Methodology for Requirements Definition, *Information Management & Technology*, 27, 6, 1994.
4. Beagrie, N. and Greenstein, D., *A Strategic Policy Framework for Creating and Preserving Digital Collections* (Arts and Humanities Data Service: London, 7/14/98).
5. Bennett, J., *A Framework for Data Types and Formats, and Issues Affecting the Long Term Preservation of Digital Material* (British Library Research and Innovation Centre: West Yorkshire, U.K., 1997).
6. Commission on Preservation and Access, Research Libraries Group, *Preserving Digital Information: Report of the Task Force on Archiving Digital Information* (Yale University, May 1, 1996).
7. Consultative Committee for Space Data Systems, *Open Archival Information System (OAIS) White Book*, CCSDS 650.0-W-4.0, 08/17/1998.
8. Dollar, C., *Ensuring Access over Time to Authentic Electronic Records: Strategy, Alternatives, and Best Practices* (Manuscript: Vancouver, BC: Nov., 1998) [available soon as a digital book from www.cohasset.com].
9. Dollar, C. and Williams, R., New Strategy for Migrating Long-Term Electronic Records: Meeting Operational Needs with Less Risk and at Lower Cost, *Cohasset Associates, Inc. Managing Electronic Records Conference Proceedings*, 1995.
10. Getty Conservation Institute and Getty Information Institute, *Time & Bits, Managing Digital Continuity*.
11. Haynes, D., et al., *Responsibility for Digital Archiving and Long Term Access to Digital Data* (British Library Research and Innovation Centre: West Yorkshire, U.K., 1997).
12. Hedstrom, M., From Practice to Theory: Applying Archival Theory to Electronic Records Management, *Cohasset Associates, Inc. Managing Electronic Records Conference Proceedings*, 1996.
13. Jones, M., *Bibliography of Materials Relating to the Preservation of New Technology and Preservation Using New Technology* (National Library of Australia: Canberra ACT, 1994), <http://www.nla.gov.au/1/pres/pubs/bibmj.html>.
14. Kahle, B., Preserving the Internet, *Scientific American*, March 1997, <http://www.sciam.com/0397issue/0397kahle.html>.
15. Kelly, K., et al., *Models for Action: Practical Approaches to Electronic Records Management and Preservation*, Center for Technology in Government: Albany, NY, 1998.
16. Kowlowitz, A. and Kelly, K., *Functional Requirements to Ensure the Creation, Maintenance, and Preservation of Electronic Records*, Center for Technology in Government: Albany, NY, 1998.
17. Long Now Foundation, Clock/Library, <http://www.longnow.org/library/library.html>.
18. MacCarn, D., *Toward a Universal Data Format for the Preservation of Media*, WGBH Educational Foundation: Boston, undated.
19. McGovern, T. and Samuels, H., Our Institutional Memory at Risk: Collaborators to the Rescue, *Cause/Effect*, 20(3), 19–21 and 49–50, 1997.
20. Rasmussen, D., *Digitization — The Issues, Projects and Technology: A Selective Bibliography*, National Library of Canada: Ottawa, 1995.
21. Ross, S., Acting to Avoid Loss, but When Disaster Strikes — Relying on Digital Archeology, *Cohasset Associates, Inc. Managing Electronic Records Conference Proceedings*, 1998.
22. Rothenberg, J., Ensuring the Longevity of Digital Information, *Scientific American*, 272(1), 24–29.
23. Rothenberg, J., Metadata to Support Data Quality and Longevity, *IEEE Metadata Computer Conference Proceedings*, 1996, http://computer.org/conferen/meta96/rothenberg_paper/iee.data-quality.html.
24. Shepard, T., Universal Preservation Format Update, *D-Lib Magazine*, Nov. 1997, <http://www.dlib.org/dlib/november97/11clips.html>.
25. Skupsky, D., Legal Requirements for Optical Disk Records and Electronic Imaging Systems, *AIIM Show and Conference Proceedings*, March 31–April 3, 1996.
26. Sutton, M., *Document Management for the Enterprise: Principles, Techniques, and Applications*, John Wiley & Sons: NY, 1996.
27. Williams, R., Managing the New Corporate Memory, *International Council on Archives — Section of International Organizations Workshop*, June, 1998.
28. Williams, R., The Seven Key Legal Issues for Managing Electronic Records, *Cohasset Associates, Inc. Managing Electronic Records Conference Proceedings*, 1995.
