Life Cycle Information Management: A CASE STUDY

KAREN A. SHAW AND GENE J. HICKOK

AT THE CORE

THIS ARTICLE EXAMINES:

- The components of life cycle information management systems
- The functional requirements for records management applications
- Tools for life cycle information management systems

The diversity of formats and media to be included in recordkeeping systems is growing dramatically. How can users and information managers capture these newer formats, such as video and audio, so that they can be effectively managed as records?

What are the fundamental challenges posed by non-traditional media and formats? To begin with, users must capture these products in such a way that they can manage them as records. Users must capture the appropriate metadata that will allow them to manage, use, protect, and disseminate these advanced record types. This technology requires communications bandwidth, digital storage systems, and techniques that dwarf the technology requirements of the past and merely scratch the surface when predicting where future requirements will lead. How do users incorporate compression techniques that reduce these requirements while maintaining the integrity of these record types?

These questions and many more are addressed by exploring the components required for an end-to-end life-cycle management system. The challenges in integrating these components and exploring the roadmap for constructing a system that is both useful and successful in the mission of records and information management is addressed in the following sections.

Editor's Note: *The companies and products named in this article are provided as examples by the authors. Their appearance here does not constitute endorsement by ARMA International.*

Center for Army Lessons Learned (CALL) System

The primary mission of the Center for Army Lessons Learned (CALL) system is to provide the right information to the right person at the right time in the right form, anywhere in the world. Although the CALL system is commonly referred to as the CALL DB (database), it is not truly a database. It is comprised of four principal systems – an Internet Gateway, the CALL Collection Observations Management System (CALLCOMS), the Virtual Research Library, and the Defense Information Technology Testbed (DITT), which is developing a life-cycle management system.

The CALL DB encompasses established processes and method-

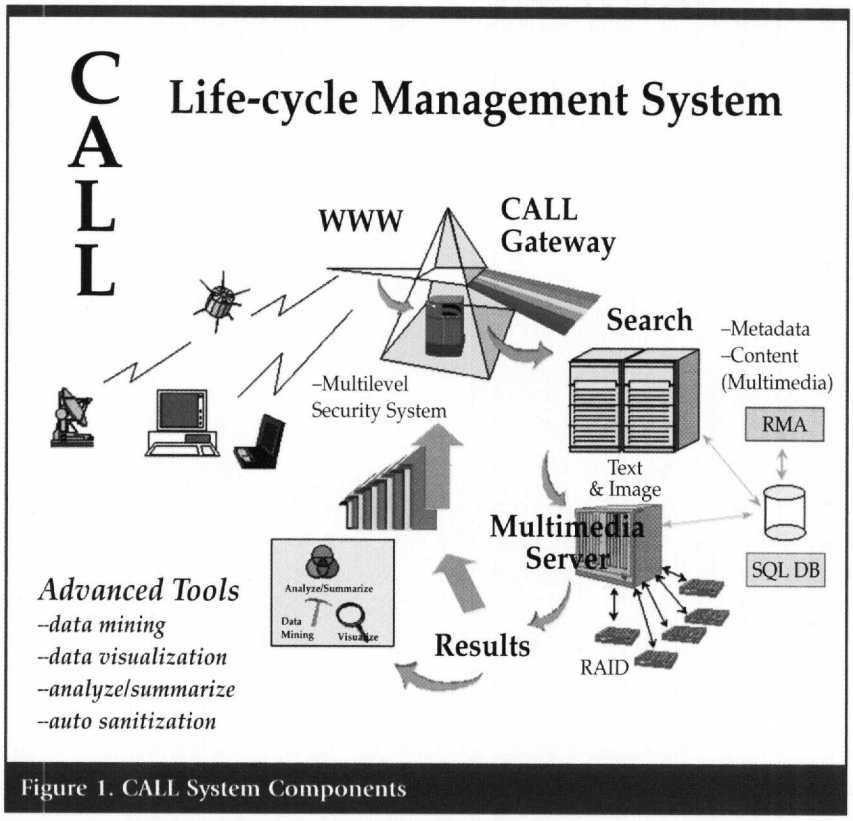

Figure 1. CALL System Components

ologies, search engines, advanced tools, a records management application, document management systems, communication lines, and numerous information systems, and it is available through the World Wide Web (WWW). The CALL implements the "thoughtware" required to sustain and evolve these robust information tools, systems, and their respective functionality as a coherent whole. To support the development, implementation, management, and use of the CALL DB, a team comprised of people from various information and military disciplines – military analysts, management analysts, library scientists, historians, archivists, archives specialists, records and information managers, lexicographers, search engineers, and information technology specialists – was assembled.

The life-cycle management system (LCMS) is an integral component of the Virtual Research Library and the CALL DB. Each component of the overall system was developed and implemented to perform a specific function as well as to contribute to the overall functionality and capability of the Virtual Research Library.

Resources (time, money, and personnel) required for one system are cross-utilized in others. This approach and philosophy has reduced overall investment costs, ensured open architecture, alleviated duplication of effort, and utilized the strengths of systems already developed for specialized purposes by incorporating them into a family of systems, thereby enabling the U.S. Army and other Department of Defense (DoD) agencies to do far more than their leaders had ever imagined. This approach applies the true definition of a knowledge management system: many systems working together to provide information at the level desired or needed, whether it is raw data, primary or secondary documents and records, or analyzed information.

Components

The components of a life-cycle management system are illustrated in Figure 1.

Capture

One of many challenges facing records and information management (RIM) professionals today is to capture records and information effectively. CALL subject matter experts (SMEs) have been struggling with this dilemma for some time and have reached the conclusion that RIM managers must have the capability to capture or acquire electronic records at their creation or approval points. Components of a LCMS should have the capability to identify records at their creation point and to establish and capture the audit trail and status of records as they are routed through an organization for review, revision, approval, and final distribution. The system would then be able to capture the official record once the division or person who has been delegated the authority to "sign" and release records has approved and saved the document/record as "final." This process may sound simple but, in fact, it requires careful planning, integration, and implementation of numerous systems and cross-utilization of information and RIM personnel. Records and information managers can insert themselves as SMEs and assist with or physically configure the functional side of a records management application (RMA) and workflow system. These systems utilize existing local area networks and require functional and technical configurations of workflow and document management systems that will enhance users'/action officers' capabilities and provide them with helpful tools at their desktops. The desktop tools and functional configurations contribute to the identification

of a record through the capture of metadata. These tools establish the integrity of the record through audit logs and "official" signatures, enabling all or any part of the process to be captured and stored in the RMA software, its relational database and repository.

CALL-DITT continues to review and work on the development of software applications that are integrated at the personal computer level. These tools must be simple to use, provide file management capabilities, and contribute to the user's/action officer's daily business process of creating, routing, and identifying records. If action officers can use the same software to organize and locate records in their personal file space, then RIM managers have contributed to the overall process of identifying important records so that they can be acquired for management, legal, and long-term preservation.

CALL-DITT has concentrated on implementing an RMA software against the Virtual Research Library multimedia holdings. The components of this effort are discussed next.

Collection, Identification, and Upload

One of the most basic challenges in acquiring records for long-term preservation and use is being able to identify them. In November 1997, the DoD finalized the standard DoD-STD 5015.2, *Functional Requirements for a Records Management Application.* It also provided, for the first time, a reference tool and guideline for all DoD agencies. Most importantly, it provided a minimum set of metadata required to identify and manage information as a record. It also defined system requirements for performing the management function using a software application. This standard is worth looking at, digesting, and applying to an organization's records system, whether analog or digital.

Although the DoD standard includes technical requirements and capabilities for a RMA, it also includes the minimum requirements for identifying a record's metadata, so that the record can be managed effectively. CALL-DITT began reviewing and implementing the minimum metadata requirements from this standard as it was being developed. By actually trying to implement the emerging standard, numerous portions of the standard were identified as either unrealistic or too vague, enabling it to be revised prior to staffing and finalization. This practice of "spiral development" – the actual testing of a functional requirement prior to implementation

or technical testing – has enabled the DoD to save precious resources and provide a realistic policy and standard that is implementable.

Conversion

The conversion component of any system typically deals with migration of legacy data that must be maintained in accordance with disposition requirements. These records and/or data types are usually a secondary capture process of legacy data. This component may require both software and hardware implementation, and it must maintain the integrity of the record throughout this process. Two primary processes accomplish this step.

1. **Analog to digital** – In this component the conversion takes the form of translating one media type into another. This translation could be accomplished by scanning paper into a digital format or converting analog video (i.e., VHS) into digital video format. In the example of paper to digital conversion shown in Figure 2, the correct image format should be chosen such that the digital version is an exact duplicate of the original and that lossless compression techniques are used to prevent information loss during the use of compression algorithms. Other formats, such as a distribution format like portable document format (PDF), can be linked to the lossless format to fulfill search, retrieval, and distribution requirements.
2. **Software** – Software conversion is the more difficult process because the differences are very subtle over time. The conversion moves through incremental changes that may go unnoticed until major changes are required. Two issues related to software conversion are version control and operating system changes.

   • Version Control – An example of version control would be converting Microsoft® Word version 2.0 to version 5.0. This conversion would be done at a system level, and it would assist in keeping records as close as possible to the current technology with minimal impact on financial investment. Typically, the newest version of a commercial software can read only the last or the two previous versions of a software release.

   • **Operating System Changes** – Changes in operating systems (OS) can be quite crippling if not monitored closely. This possibility is one of the biggest fears for agencies such as the U.S. National Archives and Records Administration, whose mission is to preserve the record of our nation's history. Systems are developed around OS, and when the OS changes, the older systems are no longer supported. This lack of support results in costly changes that can take months, if not years, to correct.

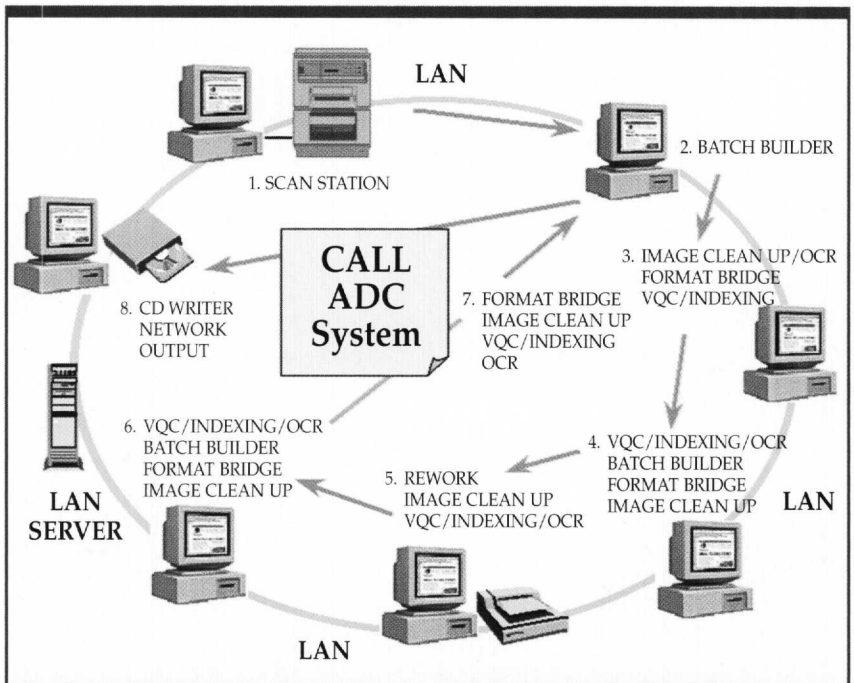

Figure 2. Paper to Digital Conversion

Records Management Application Software

RMA software can mean different things to different people. Define the capabilities you expect your RMA to perform prior to purchasing RMA software and avoid being disappointed in its capabilities. You may be surprised that your expectations require a document management *and* a workflow system in addition to the RMA.

Although document management and workflow software have been around for a while in the United States, RMA software has not. In fact, two of the first RMAs to be certified as meeting the DoD standard were developed outside the United States. Such development is not unusual when you consider that in countries outside the United States, records management is performed as an integral part of

their archives and archival function. Canadian-developed ForeMost®, Version 6.3, by Provenance Systems, Inc., and Australian-developed TRIM®, Version 4.2, by TOWER Software Corporation were the first to have been approved by the Joint Interoperability Test Command (JITC) at Fort Huachuca, Arizona, as meeting the minimum requirements set forth in the DoD standard. Several other software applications and integrated solutions have also been approved. The mission of the JITC is to support warfighters in their efforts to manage information on and off the battlefield. The purpose of the RMA certification-testing program is to perform technical certification testing of RMA products for compliance with DoD-STD 5015.2. The JITC will provide a test report containing the results of successful RMA certification testing to each participating vendor and to government program managers engaged in the procurement of RMA products. The JITC will continue to evaluate and test RMA software, posting summary results for public access and utilization. For the DoD, the JITC has been delegated the authority to maintain the DoD-STD 5015.2, and they are facilitating a working group for the first enhancement to the DoD standard concentrating on minimum system and functional requirements for security classification and declassification. This follow-on effort includes identification of standardized metadata requirements as well.

CALL-DITT has chosen to utilize Tower's TRIM as their RMA first implementation. This decision was based on the workflow, document management, and RMA capabilities TRIM offered. When a RMA is packaged with a workflow and/or document management system, implementation and integration costs are less, as are the requirements to purchase and implement other compatible systems that will provide the metadata capture, audit trail, and routing/staffing requirements needed to support a life-cycle management system. Companies will also have a proven, seamless product that provides capabilities and tools at the desktop level.

The purpose of the records management applications (RMA) certification-testing program is to perform technical certification testing of RMA products for compliance with DoD-STD 5015.2.

Through interaction with CALL-DITT and other U.S. companies, Tower's TRIM desktop tools and functionality are becoming more user-friendly. The customers of RMA software packages, such as TRIM, are not just records and information managers, but action officers, department heads, and employees throughout an organization.

TRIM has the capability to set up metadata input templates using various structures such as a file plan, file series, or file retention. Most legacy holdings do not come with their respective metadata. Therefore, CALL has chosen to implement TRIM by functional area, which enables minimum metadata input responsibilities to be decentralized, i.e., performed by the various information disciplines in-house. Decentralization enables a records management specialist (RMS) to input metadata that applies to all records as a "batch" and route the "batch" to archive technicians who apply long-term preservation, search, and retrieval metadata at the document level. When metadata has been placed against individual records, these records are sent to a team chief, who is usually a historian or an archivist. This individual has the capability to view individual records with their associated metadata or groups of records to determine the best archival structure and arrangement. This value-added structure provides users with an additional option to search for, or "browse," through records within the life-cycle management system using metadata, document content, or both.

Because CALL had high-production input requirements that TRIM was not originally designed to handle, the CALL-DITT team developed a graphic user interface (GUI) that enabled rapid metadata input and metadata carry over from previous metadata input screens' completed records to the next record to be processed. This template and added capability has streamlined the metadata input and upload process to the long-term repository. A simple-to-use GUI can also be provided to remote customers who wish to utilize the technology and systems within the CALL LCMS. By providing them with this GUI, remote customers are able to utilize their own personnel resources to identify their records and, consequently, provide CALL with their records already arranged and identified. This simple GUI has created a win-win situation for CALL and its remote customers. The ultimate benefit is passed on to the end users.

Digital Storage Systems

Digital storage systems required to maintain records may be as simple as a laptop computer or as elaborate as a video server that can stream video to 5,000 simultaneous users. These systems should support open architectures and provide timely and consistent performance in delivering records to the appropriate customers.

When choosing a media type for storage, the longevity of the media must support the disposition of the records types that will be stored on it. The fact that your records have been stored and are physically in your custody does not mean you can recover the records. The storage system and media chosen are affected by mechanical and environmental elements that can dramatically improve or degrade your ability to maintain your records; so choose wisely based on all your needs.

Another fact to consider is how the media type may affect your ability to maintain dissimilar disposition schedules. If you have chosen CD-ROM as your media type, you should store only records with similar dispositions because you cannot selectively remove individual records from this media type.

Advanced Tools

Search engines, data mining and visualization, abstract, summary, user-defined graphical user interfaces, and system and information security capabilities are advanced tools that CALL, under the DITT program, is working on.

Search Engines

Search engines are extremely important to internal and external customers of your records. Search and retrieval is where all the rewards of metadata and document and records management become apparent to users. Most customers do not care how records or information get into the repository; they only want to access, search, and retrieve them easily. However, a repository is only as good as the information it contains and the search engine(s) used. Ensuring data integrity, system and information security, and files identification are key functional components that make a search engine perform its primary function of providing customers with the right information in a timely manner.

In 1994, CALL documented their functional requirements and, after doing much homework, chose Excalibur's EFS search engine for the Virtual Research Library. The search engine software was acquired first but had limited capabilities for metadata capture, input, and functional management. Because the strength of EFS was search and retrieval, the team chose to concentrate its efforts on document/record organization and arrangement with minimum metadata input to get records online and accessible. That was the right decision at the time, but the target system still required capture, input, and RMA components. With migration from Excalibur's EFS to their newer product, RetrievalWare, these components will be implemented.

Like EFS, RetrievalWare provides users with Boolean, content, or pattern recognition search capabilities. Because its strength is to search and retrieve, CALL wants external users (read-only) to continue accessing the search engine GUI first, offering the metadata and document content as an option for search and retrieval. Although CALL thought this to be an everyday implementation method, it is not. This method of accessing records via the search engine side instead of going through the RMA creates challenges of its own. Remember that the primary function of a records management application is to manage records, not to search and retrieve.

More Search Engines

The CALL-DITT target system provides users with numerous ways and capabilities to get at the data or information needed. For example, the search engines CALL utilizes are all state of the art. There are more than 40 search engines available at CALL's Web site at *http://call.army.mil* (accessed 2 October 2000).

For records in HTML format, CALL utilizes a visualization search engine – SemioMap® – that displays results automatically organized into related records groups. The team is exploring the capability of applying this type of technology to converted records and pure electronic records in formats other than HTML.

Data Mining and Visualization

Data mining and visualization packages are critical tools that will assist users of information and knowledge sets in pairing down the massive amounts of data that correspond to their query. Customers may use standard Boolean commands to enter queries into the systems, and networks and visualization tools will show the links between data and concepts that may have gone unnoticed during the use of conventional query tools. These advanced tools will allow researchers to see on which areas and/or documents of a database they must focus their efforts.

These tools may exist as two- or three-dimensional representations and use advanced algorithms to evaluate word co-occurrence and concepts as well as pattern recognition to represent tens of thousands of records types in a single image. The example provided in Figure 3 is from SemioMap® Text Mining search engine, using the search clue of "information warfare." This advanced tool can be found on the CALL home page. Under the "Search Engines" button, choose "SemioMap Visual Search."

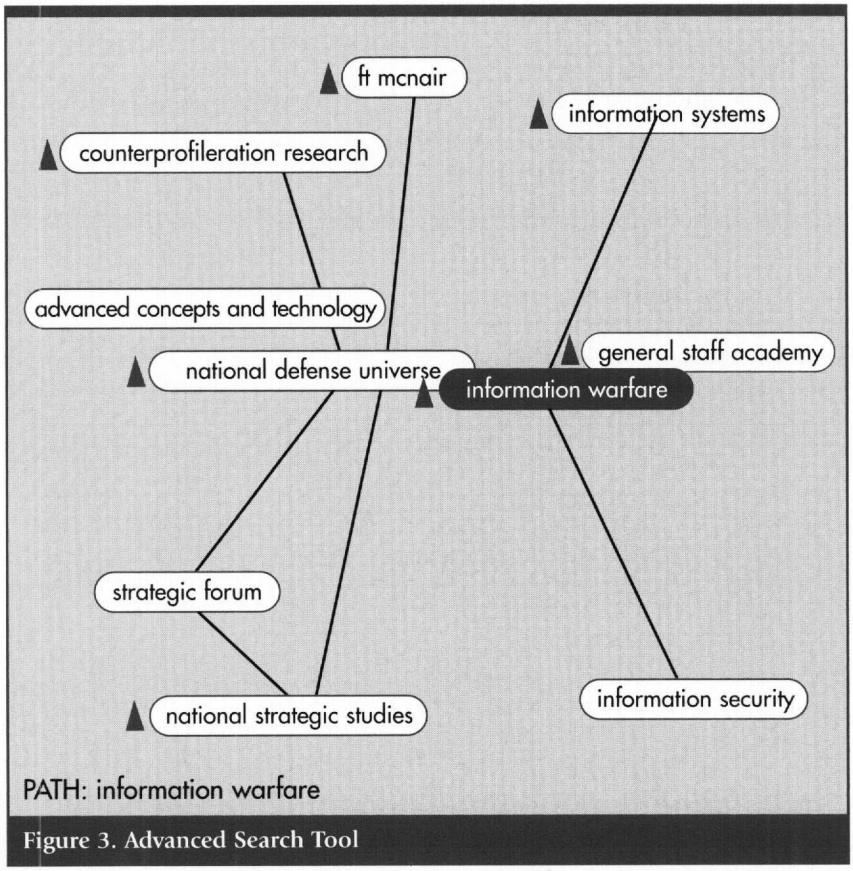

Figure 3. Advanced Search Tool

Abstract, Summary

Those who have used the Internet to locate information have discovered that too much information is as bad as not enough. An automatic abstract and summary tool, however, will enable users to scan a document summary quickly to determine whether the record presented to them is what they are looking for. Some software applications provide this capability by presenting users with extracts from the first paragraph of a document. This advanced tool would actually look for and determine the important parts of a document and summarize them. Another opportunity presents itself with this type of tool – utilizing the tool internally to create an abstract and summary and populate a metadata field with the results.

Graphic User Interfaces (GUIs)

A GUI is a good way to customize software to meet the needs of different customers. GUIs should be transparent to the customer and seamless interfaces with the systems they support. Companies may have to consider developing GUIs for their users to assist with and contribute to their responsibilities for data input or electronic filing. This component of a life-cycle management system should not be overlooked because it is the first thing users may see each time they log onto their computers every day.

Security

Security is a crucial component used to demonstrate the integrity of an end-to-end records management system. It is the component that provides access to the appropriate people and protects the system and records within it.

Access

Access to each system must be controlled. This control is typically accomplished by using passwords or other more sophisticated means. Some control methods use fingerprints, voice, retinal scans, and facial features to authenticate access; however, the level of protection used must be weighed against the loss or corruption of the records sets.

Multilevel Security

Multilevel Security (MLS) is a tool that is used in national security situations, and it is being investigated for use in the private sector. The incorporation of a MLS system would remove the need to have completely separate systems for each level of restriction. Instead, a single access point would be used for all levels, and the data would be returned as requested, returned with modification, or simply returned as a null set in which the user would think that simply no matches were made to the query sent.

Redaction and Sanitization

Redaction and sanitization tools in collaborative use with a MLS system would perform the required functions of protecting privacy information by replacing real names and places with fictitious ones (sanitization) or simply covering up the protected information with an opaque marking that cannot be removed (redaction).

Communications

The communications component of this system refers to the landline and satellite infrastructure to support the access and dissemination of knowledge and records from anywhere on the globe. This structure represents a means to transmit and receive massive amounts of data. If you consider that the amount of data held on a conventional floppy disk is 1.2MB, you have an idea of how much information is stored on that disk. Now, relate this concept to high-

definition television, which needs 1.2GB (1,000 floppy disks) per second to be transmitted and stored through a relatively narrow pipe. The following example gives a feel for how large this problem is: if hundreds of thousands of people request a two-hour movie simultaneously, the equivalent information stored on 7.2 million floppy disks would have to fly flawlessly through the air to the exact places that it is needed.

This same daunting task is facing landline communications. Researchers must continue to develop new ways to get more information inside the same size pipe without causing any loss of resolution or content of these priceless knowledge sets. Numerous projects are under way to investigate ways of splitting light into its various color components and sending different information on each color band (wave division multiplexing) or of compressing digital information with advanced techniques that compress at 500:1 and only transmit information that has changed in a video scene or document.

Web-based Technologies

The World Wide Web has changed the way almost every company and organization provides and presents information and products to its customers. The emergence of Web technologies has led the development of software applications. Most organizations, including the DoD and the federal government, continue to have requirements for Web-compatible software and applications. The DoD is a heavy user of the Internet, both as a customer and a consumer. However, the DoD does not just consume. Vice President Al Gore has supported an initiative to invest government funding to improve Internet connectivity and capabilities. The DoD is collaborating with major commercial communication companies to ensure future communication requirements are being identified through operational prototyping and spiral development projects such as those implemented under the DITT.

Identification, Longevity, Preservation

Web servers are not records management systems or long-term repositories, but they were designed to provide convenient access to information. Normally, when records are readied for the Web, they are converted from their native format into a format called "hypertext markup language" (HTML). Metatags are identified along with keywords that are hyperlinked, thereby enabling Web browsers to search for and find information. To the authors' knowledge, no standardization exists for identifying records and their respective metatags and hyperlinks. Additionally, unless a company has instituted a policy to maintain an electronic record in its native format for migration and functional management purposes, only the HTML format or a nonmigratable format such as PDF is retained. The HTML or PDF document is then placed on a public or open system that was never intended to ensure data integrity or provide long-term maintenance of that record.

Web servers remain a way to access and disseminate information and records. CALL-DITT is exploring and testing ways to enable Web team personnel to create a document for distribution with its associated metatags for the Web while maintaining and sending the original electronic record with its associated metadata/tags to a long-term repository for preservation, management, long-term reference, and utilization.

Challenges in Implementation and Integration

Two important challenges to implementation and integration of the CALL-DITT system are technical and political.

Technical

The technical challenges that exist in implementing a fully integrated records management system are in two key areas: functional performance and standardization. Functionality of the system should be driven by its intended use and its customers and not by production statistics. Many technical solutions can perform the same function. Functionality also prevents "big brother" from dictating which vendor supplies your system, as long as the functions are performed.

Difficulties arise when one system must interface with systems at other locations that may or may not have compatible databases. This nonstandardization means that data structures and formats may not be compatible. The result of this incompatibility will manifest itself when you are not able to see and/or use the information and knowledge in systems dissimilar from yours. If functional processes and standardized metadata are used to describe records, these challenges are alleviated.

These same concepts address the issue of front-end collection of records and supporting metadata, and thereby prevent the costly capture and conversion of records late in their life cycle. Once records have reached the point where the processes for handling them are primarily manual, they provide limited value-added usefulness.

One of the remaining challenges is in the area of remote use and control of a records and information management system. Records have

limited usefulness if they are not shared, so your systems should be targeted toward the lowest echelon of use. That user may be someone in the office across the hall or in a foxhole on the other side of the world. In either instance, the end user should be able to put records into the system and search for records already resident within the system.

Political

The primary political issues surround the "not-invented-here" syndrome and around the old axiom that "knowledge is power." The not-invented-here issues address the misperception that "I" am the only one who has a vested interest in these records, and therefore only "I" will protect and provide access with the integrity required. If users can agree on the functional processes and the required metadata, protection and access become a moot point because the end state would be the same in any agency and/or situation.

The second political issue is one of power. One who holds the knowledge holds the power and, more than likely, the budget to go along with it. These issues are addressed in the reasons for promoting distributed archives. If all knowledge existed in a single location, users could not build enough communications infrastructure to support the system. A risk of terroristic threats or acts of war that could cripple the nation by simply destroying a single site would be added as well.

Roadmap to Construction of a Successful Records Management System

Development of a life-cycle management system should be accomplished in small steps. Spiral development is the best approach to constructing a successful system.

Small Steps With End-State Vision

Concept plans must be developed; however, when you actually begin to implement the concept, start small. Determine a process, function, or piece of the overall system that will provide a foundation for other systems or processes to build on or feed into. Employ a team concept. Life-cycle management systems developed in a vacuum or those with only one information discipline focus will not flourish in today's information age.

Spiral Development

Spiral development techniques either in-house or contracted out, enable companies and organizations to design a little, test, then refine their requirements. This type of system development has become more successful. Spiral development and implementation enables companies and organizations to have a continuous line of communication with the developers and users and actually assist with the development and testing phases. This process of "develop a little, test a little" enables functional users to clarify or adjust their requirements to enhance the capabilities or functionality of a product or system while it is in the development stage.

Small steps, common sense, teamwork, a lot of luck, and good planning will enable your organization to develop and implement a system that will manage multimedia records and provide a value-added business asset to your company or organization.

ABOUT THE AUTHORS: ***Karen A. Shaw*** *is senior information and records manager and project officer for the Center for Army Lessons Learned, Defense Information Technology Testbed at Fort Leavenworth, Kansas. She has 14 years information management experience and specializes in enterprise electronic records management, information and knowledge lifecycle management, and metadata development and standardization. Shaw is a member of ARMA International and has served as the Region IV vice president on its Board of Directors and as the president of her local chapter. She is also a member of the Federal Information & Records Managers Council and the Department of Defense, Enterprise Document and Records Management Integrated Product Team. She received achievement awards for her influence on effecting changes to the Department of Army's information strategy and is recognized in the Department of Defense as a records management expert and visionary leader as project officer for the DITT. Shaw may be reached at* shawk@leavenworth.army.mil.

Gene J. Hickok *is deputy director and senior market development specialist for National Media Laboratory, Defense Information Technology Testbed, in St. Paul, Minnesota. He has 12 years information management experience and specializes in technology's role in records management, information and records lifecycle management, and metadata development and utilization in records management. He is a member of the Department of Defense, Enterprise Document and Records Management Integrated Product Team and of the Intelligence Community Working Group on the Use of Multi-level Security Systems in Redaction and Declassification. He has received two Circle of Technical Excellence awards for supporting the government in the recovery and long-term preservation of information critical to national defense and for defining government strategy for managing all DoD information as records. He may be reached at* gjhickok@mmm.com.

BIBLIOGRAPHY

[1] DoD-STD 5015.2, Assistant Secretary of Defense for Command, Control, Communications, and Intelligence (ASD, C3I); Design Criteria Standard for Electronic Records Management Software Applications, November 1997.