Long-Term Surveillance and Maintenance Records: Challenges Inherent to Managing Electronic Records – 15165

Jeanie Gueretta*, John Montgomery*, Ruth McKinney**, Cliff Anglim**, Peter Cameon**
*U.S. Department of Energy Office of Legacy Management, 99 Research Park Road Morgantown, WV 26505; jeanie.gueretta@lm.doe.gov, john.montgomery@lm.doe.gov
**Source One Management, Inc.

ABSTRACT

The U.S. Department of Energy (DOE) Office of Legacy Management (LM) is an integral part of DOE's strategy to ensure that legacy liabilities of former nuclear weapons production sites are properly managed following the completion of environmental cleanup activities. Records management is crucial to the protection of health, environmental, and legal interests of the Department and the public. LM is responsible for preserving long-term surveillance and maintenance (LTS&M) records in performance of its mission.

Electronic records (e-records) documenting LTS&M activities are becoming more prevalent. Managing e-records creates new challenges in maintaining their accessibility and readability for the duration of their use.

INTRODUCTION

LM is currently responsible for records from 89 decommissioned sites, with plans to manage records for 127 sites by the year 2020. Emphasis on maintaining information for current and future use by internal and external stakeholders is an important aspect to the LTS&M of former nuclear weapons production sites. Historical and current information is needed to perform surveillance and maintenance activities and will provide the framework to tell each site's story to future generations.

Technological advancement in the last few decades has created challenges for records management. In November 2011, the White House issued Presidential Memorandum – *Managing Government Records*, which sought to improve records management policies and procedures to provide accountability for agency actions and decisions [1]. In response, the Office of Management and Budget (OMB), in conjunction with the National Archives and Records Administration (NARA), issued M-12-18 – *Managing Government Records Directive*. The directive requires, to the extent possible, the elimination of paper records and use of electronic recordkeeping [2].

Although physical records management practices are applicable to e-records management, the inherent challenges must also be considered. NARA defines an e-record as "any information that

is recorded in a form that only a computer can process and that satisfies the definition of a Federal record per the *Federal Records Act*." [3] The term includes both record content and the associated metadata required to meet business needs. Conversely, most historical physical records do not require a machine to be read and the media (e.g., paper, photographic film, and microform) they are stored on are still readable today.

DISCUSSION

Challenges associated with the management of e-records include: capture methods, digitization of legacy physical records, and automation using content analytics and auto classification. Additionally, problems and solutions associated with maintaining the accessibility and readability of e-records, in a myriad of formats, for current and future LM decision making and stakeholder needs will need to be examined.

Capture Methods

Methods of capturing records for ongoing storage and retention vary among organizations and should be tailored to meet organizational needs. LM capture methods were selected to maintain compliance with NARA requirements and DOE guidance.

LM manages e-records (i.e., email, word processing documents, presentations, etc.) in a Department of Defense Standard (DoD STD) 5015.2-certified Records Management Application (RMA) [4]. A standard-certified RMA ensures the software meets the minimum NARA records management requirements. The RMA has the capability to store e-records, schedule records for retention using record retention schedules, and is a finding aid for physical record collections. The RMA search engine facilitates the search and retrieval of records.

Email capture and retention is a subject at the forefront due to its increasing use for formal correspondence and decision making. As an alternative option for the capture of permanent email records, NARA issued bulletin 2013-02, *Guidance on a New Approach to Managing Email Records*, introducing the Capstone approach [5]. The Capstone approach provides a technological solution for managing the high volumes of email traffic, while removing the onus of responsibility from the individual user.

The Capstone approach is a solution that allows Federal agencies to tie email retention to the organizational structure and the role of the individual sending or receiving the email. Retention is bound to the functions of key personnel and provides for automated processes relating to the submittal and capture of email records. Capstone enables organizations to leverage technology to cope with large volumes of email, while meeting Federal requirements for the capture of e-records. The Capstone approach requires coordinated efforts between information technology and records management personnel to ensure the capture method is both feasible and compatible with existing systems prior to implementation. An organization must identify key personnel for

capture of e-mail to ensure capture of important organizational permanent records.

Most LM e-records are submitted by end users for storage and retention via dedicated records coordinator email addresses. Items are submitted as an attachment to the submittal email and converted to portable document format (PDF), as applicable, for ongoing retention (applicable electronic file formats suitable for preservation are addressed in the Formats section). Records Management personnel monitor the email inboxes and process the record (i.e., collect required metadata for data entry and determine the applicable retention schedule) for storage in the RMA.

E-records contained in databases are captured, retained, searched, and retrieved using existing electronic information system (EIS) capabilities. An EIS is a system (spreadsheet, tracking database, etc.) designed to automate business functions, which may include official records. An EIS may or may not incorporate the requirements of an RMA, depending on its design characteristics. Each EIS should be assessed to determine record content and then scheduled and dispositioned as appropriate. LM recently completed a thorough review of all LM databases by analyzing the inputs and outputs of each system to determine records content, and applying a records retention schedule, if necessary.

Digitization of Legacy Physical Records

Similar to most organizations, LM has a combination of both physical records in a variety of media and e-records in various formats. E-records are convenient to access, require significantly less storage space, and may be readily shared at remote locations.

The process of digitizing records supports the LM goal to preserve, protect, and share records and information. When digitizing existing hard-copy record collections, LM must consider the cost versus benefits. Digitizing paper records can be accomplished in-house or by a vendor. However, when digitizing other analog formats (e.g., microfiche, microfilm, or analog video and audio), the cost of procuring specialized digitizing equipment, training or hiring professional digitizing personnel to produce archival quality end products, in many instances, is cost prohibitive. Using vendors specializing in the conversion of other analog formats is usually more cost effective, but can still be expensive.

LM currently digitizes paper records on an as needed basis.

Formats

Formats used for digital preservation is based on NARA regulations and guidance. LM currently converts digitized paper records to PDF using PDF/A-1b (ISO 19005-1:2005) conformance for preservation purposes. PDF/A is an archival format intended to be suitable for long-term preservation by embedding all fonts and disabling any security settings. Transfer documentation, acceptable image quality specifications, and restrictions on optical character recognition processes are required, as applicable.

In January 2014, NARA issued Bulletin 2014-04, *Revised Format Guidance for the Transfer of Permanent Electronic Records* specifying acceptable file formats for transferring permanent e-records to NARA. Additional guidance on metadata requirements for e-records will be issued separately [6].

A key change in NARA Bulletin 2014-04 is there is no longer specific PDF guidance for permanent e-records; PDF is now only listed for format categories where it is appropriate. LM's standard for archiving emails as records has been PDF/A-1b. In light of the new format guidance, LM is evaluating the implementation of the NARA preferred or acceptable archival formats.

Automation Using Content Analytics and Auto-Classification

As technology becomes more widespread, the volume of e-records produced will soon exceed allocated Records Management resources. Records Management in large organizations must test and deploy automated processes to augment its resources to categorize and schedule electronic files that are created or received that meet the definition of a Federal record (44 United States Code [U.S.C.] 3301).

Content analytics is the use of technology to assist in evaluating and interpreting large volumes of data. Software applications can assist with data lifecycle management through automated processes such as identification of duplicates, privacy and sensitive content, auto-classification and deletion of data, and holds management.

Auto-classification is intended to eliminate the need for end-user involvement in the identification and archiving of emails as records [7]. Auto-classification products rely on a combination of textual cues and organizational structure to classify emails, which can be incorporated into an organization's RMA for direct import. These tools match the growing volume of e-records and email, while automating the task of reviewing and capturing emails as records. The system is trained to operate at an acceptable level of granularity and can be paired with human oversight for verification and increased accuracy. Auto-classification products can yield acceptable to high levels of accuracy at reduced processing cost and with consistent results. Familiarity can also increase accuracy over time. The key to successful use of auto-classification is a thorough approach and an experienced implementation team.

Auto-classification technology is fairly new and experts disagree about the effort needed to adjust the software and overall accuracy. The software can analyze a large volume of data with 60-80 percent accuracy in a consistent manner, with less time and cost than manual processes. Most of the applications "learn" as they are exposed to more information, with accuracy typically improving over time. Additionally, the combination of human review with automated processes can vastly increase the reliability of these applications.

LM is not currently using content analytics or auto-classification software for records lifecycle management, but is reviewing the cost-benefit of implementation. Recently, LM deployed an automated application (electronic discovery or eDiscovery) for searching, locating, and securing potentially responsive electronic files for legal holds. LM is exploring other uses for the eDiscovery application.

SUMMARY

Today's organizations are faced with the challenge of managing e-records, which includes LM. To meet this challenge, LM will research options for improvements in ways to manage e-records. LM looks forward to working with NARA as they continue to promulgate requirements and solutions for e-records management for Federal agencies.

LM's records and information management program continues to enhance its capabilities to protect, preserve, and provide access to records and information systems. LM's information governance allows the program to effectively and efficiently identify and collect information from records sources and disseminate information to internal and external stakeholders.

Emphasis on maintaining information for current and future use is an important aspect to the LTS&M of former nuclear weapons production sites. Historical and current information is needed to perform surveillance and maintenance activities and provide the framework to tell each site's story to future generations.

REFERENCES

1. Presidential Memorandum - Managing Government Records (2011).
2. OMB/NARA, Memorandum M-12-18 – Managing Government Records Directive (2012)
3. 44 U.S.C. 29-35, Records Management, GPO (2002).
4. DoD 5015.02-STD, Electronic Records Management Software Applications Design Criteria Standard, DoD Chief Information Officer (2007).
5. NARA Bulletin 2013-02, Guidance on a New Approach to Managing Email Records (2013).
6. NARA Bulletin 2014-04, Revised Format Guidance for the Transfer of Permanent Electronic Records (2014).
7. NARA, Managing Government Records Directive: Automated Electronic Records Management Report/Plan (2014).