The Modern Records Management Program: An Overview of Electronic Records Management Standards

by Jennifer Seymour

EDITOR'S SUMMARY

Standardization is fundamental for bringing a vast variety of electronic records under control. It enables capturing and preserving original records as well as evidence of any access or change to the records. Creating standards within and across organizations is an extreme challenge that must be met. The 2011 Presidential Memorandum on Managing Government Records and subsequent 2012 Directive established values and strategic direction for managing federal electronic records without creating strict standards. The Battelle Record Management Office relies on a Defense Department standard for enterprise content management systems to be secure, searchable and capable of preserving contextual relationships and on the Code of Federal Regulations regarding equivalence of electronic records and signatures to paper. The result demonstrates a record's integrity and authenticity and enables it to be discovered and accessed. Defined access permissions and an audit trail add further assurances. Interoperability through application program interface layers is another requirement, being addressed through advanced platform development, which may provide the solution for authenticity and contextual preservation.

KEYWORDS

standardization

records management

electronic documents

electronic document management systems

digital object preservation

document access

authenticity

interoperability

Jennifer Seymour is document control/records management coordinator in the Battelle Records Management Office. She can be reached at Seymour<at>battelle.org.

The vast array of electronic media, file formats and record types produced by the conduct of scientific research and the work of its supporting business units would be unmanageable without some form of standardization. The intricacies of the connections between the creator organizations would be incomprehensible, and the breathtaking volume of records would be unnavigable. Devising and implementing global standards across the parent organization, however, is nearly inconceivable. Every semantic business unit within each business line within the market verticals conducts business in the manner best suited to its success. As a result, even standardizing the records retention schedule can be a challenge. How many international non-profit 501c(3) charitable trusts dedicated to scientific discovery and its government and commercial applications across multiple market verticals have public-facing retention schedules to consult for comparison? In developing a comprehensive records management program, we aim to identify and create executable standards in each arena, building them into processes and procedures – frameworks designed to satisfy staff, clients and regulatory agencies – based on internal organizational and departmental expertise, exercising determined and judicious conversation and collaboration. Passion for artefactual preservation aside, our mission is to provide solutions to the needs of the business via the capture and preservation of the evidence of their operations.

On the Managing Government Records Directive

Effective records management programs balance efficiency, cost-effectiveness, transparency and risk. In November of 2011, President Obama signed the "Presidential Memorandum – Managing Government Records," officially beginning the federal government's design and implementation of

CONTENTS

< PREVIOUS PAGE

NEXT PAGE >

NEXT ARTICLE >

a modern records management program that forces the adaptation of government activities to the electronic environment. Self-assessment, the reduction of redundancy and knowledge management are all identified within the memorandum as the values of sound records management – not the preservation of the public record for its own sake or a democratic ideal. The directive puts efficiency first in the General Records Schedules as well, aiming to reduce schedules by aggregation of records series in a move that could result in more efficient dispositioning via reducing the burden on both users and records managers. Archivists will be familiar with Greene and Meissner's minimalistic "More Product, Less Process" approach, a modern compromise between adequacy and efficiency when faced with overwhelming volume. In records management, similar principles guide big bucket scheduling or scheduling records by aggregation. They're facets of the same attempt to preserve the mission of records and archives, carrying out retention activities efficiently without losing context and value in the face of skyrocketing volume.

The National Archives and Records Administration (NARA) and the Office of Management and Budget released the *Managing Government Records Directive* in 2012, demanding that all email be managed in an accessible electronic format by the close of 2016 – now an issue of national prominence under the focus of an election year's media spotlight – and all permanent electronic records be managed in an accessible electronic format by the close of 2019. A significant step toward standardizing the misunderstood and mistrusted format, the directive did not provide guidance or goals for other day-to-day records of government activity but is rather representative of the realization of a shift still in flux. Through providing strong strategic direction for email and permanent records, endorsing the transition from paper to electronic records management across the federal government, the agencies began to cement this change to information culture. But digitizing permanent paper collections, as is suggested, has been the default solution to preservation issues since the heyday of microfiche – so nothing much changes there. Born-digital records are different: as a society we are still wary of what is viewed as a manipulable, falsifiable medium. The trappings of context – primarily demonstrable via metadata – have become the lynchpin of authenticity in the electronic age. Standards dictate authenticity, not any inherent quality of the record.

The most stringent standards that we apply to our role in the business are the Department of Defense standard 5015.2 ("DoD 5015.2") Electronic Records Management Software Applications Design Criteria Standard and Title 21 Part 11 of the Code of Federal Regulations ("21CFR11"), establishing the United States Food and Drug Administration regulations on electronic records and signatures. While 21CFR11 is not a generalized electronic records standard, it can be discussed here as one of the most detailed standards in common practice that directly addresses and provides a standardized solution for the active management, in records management and its sister disciplines, of electronic records and their integrity, authenticity and reliability.

21CFR11

21CFR11 addresses the conditions under which "electronic records, electronic signatures, and handwritten signatures executed to electronic records" are to be considered "trustworthy, reliable, and generally equivalent to paper records and handwritten signatures executed on paper," applying to "records in electronic form that are created, modified, maintained, archived, retrieved, or transmitted, under any records requirements set forth in agency regulations… [or] submitted to the agency under requirements of the Federal Food, Drug, and Cosmetic Act and the Public Health Service Act" (21CFR11 Sec 11.1 Scope). The standard specifically notes that these are rules by which an electronic record can be considered acceptable in lieu of paper, implying that electronic records inherently have no authenticity or integrity. On the basis of that assumption, the standard goes on to describe the regulations and requirements for achieving those acceptable conditions. A facility must be able to demonstrate designed procedures and controls that ensure authenticity, integrity and confidentiality, requiring the validation of closed systems, the ability to produce exact copies for inspection, the administration of access controls, computer-generated and time-stamped audit trails and uniquely identifiable authority checks, among other controls. Records requiring signature validation must be verifiably unique, cannot be reused and where not based on biometrics must meet several further qualifications.

Some of these requirements have been decried as obstructions to the conduct of business, too burdensome to implement, but they serve as a lens through which we might analyze the significance of the commonplace assumption that electronic records have one sure quality in common: manipulability. It is easier to question and disprove a record's authenticity than it is to prove it, and if context is key to authenticity, then the record's physical integrity, its chain of custody and the security of the repository in which it is stored must be unquestionable. Determining how to prove that a SAS file, for example, remains unaltered during the archival process involves developing and validating an automated capture application that demonstrates the integrity of the data, represented by an MD5 hash check. Every time that an electronic file is opened or moved, there is a risk that the object could be altered in some way that challenges or endangers the integrity and validity of the data. Once generated from the original file, generating a new hash and checking it against the original proves that the file remains untampered; the validation of the process serves as documented proof that the archival mechanism does not permit alteration and therefore invalidation. The mechanism also maps structural and descriptive metadata, including the MD5 hash, from the electronic record to the enterprise content management system. The system captures provenance and chain of custody, then at ingest initiates an audit trail, a key feature of the DoD 5015.2 compliant application, to provide proof of security.

DoD 5015.2 and the Enterprise Content Management System

The cornerstone of a modern records management program is the enterprise content management system (ECM). The current software Battelle has implemented is DoD 5015.2 compliant – not always a required standard but endorsed by NARA and expected as a best practice for federal contractors. Most of the standard is below the notice of everyday activity, but key features are tied to security, searchability and the preservation of contextual relationships. The records must be made visible via the development and application of standardized and rich descriptive metadata; the application must accommodate dates and date logic – for example, the ability to search for a range of dates; it must support meta-tagging and organization-defined metadata; it must be capable of meeting particular security compliances, featuring strict access controls and audit capabilities. It is due to the flexibility of both description and searching within the application that we are able to accommodate so many varied needs in terms of access and description, and due to access controls and audit trails that we are able to depend upon the application to provide the requisite security. The combination of the application design standards of DoD 5015.2 and the technical demands of 21CFR11 ensure an electronic object's integrity and authenticity can be proven, and it can also be made accessible and discoverable by the guidance of principled metadata standards such as Dublin Core.

Though security begins with network authentication controls and other security measures, long before ever reaching the controls featured in application design, the ECM's ability to define access per records collection and, if necessary, per record, is what enables the records manager to guarantee authenticity and integrity. The system administrator identifies and assigns the appropriate permissions, defining who can take action with a record: accesses can be set such that although one can view the descriptive metadata record, they cannot view the electronic object attached to it, for example. The audit trail feature, which tracks those actions, enables the administrator to view a complete history of all actions performed on a record from the point of ingest, demonstrating the authenticity and integrity of the files in storage. The context-based metadata that create the audit trails are both structural and descriptive, with the potential to either manually enter that data or, ideally, automatically capture it at ingest. Some elements are captured by the ECM by default, such as Title (File Name), Date Created, and Date Registered, but not all, and so we develop what we need to maintain not only the integrity and authenticity of the electronic object but also the accessibility of its valuable content.

Developing standardized metadata for the entire organization might be

CONTENTS
< TOP OF ARTICLE
< PREVIOUS PAGE
NEXT PAGE >
NEXT ARTICLE >

out of the question, but developing standards based on provenance is not. Records collections are identified first by the creator organization, and collection management activities are tailored to the specific needs of each functional group, usually delineated by business unit. The Battelle ECM provides the ability to create bibliographic record templates called Record Types that use organization-defined and out-of-the-box metadata fields to describe the records to which they are attached. Keeping in mind the principles of Dublin Core, NISO, PREMIS and other descriptive metadata element standards, we have full autonomy to create our own data dictionaries and metadata elements. The fields can be string fields, text fields, dates or numbers, among other data types. Data types can be standardized by feature, such as a maximum number of characters, or restricted to organization-defined controlled vocabularies, which can be used like linked data. Any field created can also be queried, enabling multifaceted searching. A search using the Date Registered field, for example, can be filtered by Author and even Record Type itself. In concert with metadata searching, document content searching enables a text-based search of text-based file types, or files with an OCR layer, that are indexed for it by the ECM. Information can be identified as useful or relevant to an information request based on the actual content of the record, and not just an interpretation of the context.

On ESI and E-Discovery

Consider that Rule 34 of the Federal Rules of Civil Procedure inform us that one party may serve another with a request for "any designated documents or electronically stored information – including writings, drawings, graphs, charts, photographs, sound recordings, images and other data or data compilations – stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form" (Rule 34 a1A). Consider next that the "responding party" to which the rule refers is not the records management office but the organization as a whole, meaning that anything in the organization's possession, whether it has met retention but has not yet been destroyed, whether it is a copy or an original, whether it is the archived version or not, can be produced for review for litigation. The functions of an ECM that provide for multifaceted Boolean searching, multi-field filtering and document content queries to the benefit of enterprise-wide knowledge management can take the volume of e-discovery requests into the terabytes.

Volume is rapidly becoming a more difficult defense to an objectionable request. An ECM can reasonably perform a content-based query through tens of thousands of records in moments, identifying thousands of potentially responsive records based on a single search term or phrase. The operational risks are escalating dramatically. Redgrave, Peay and Bulander of Redgrave LLP presented an excellent narrative in the *Richmond Journal of Law and Technology* on the rapid transformation of e-discovery rules and precedents [1], reminding us that accessibility and discovery are not the end of identification, as attorney expertise will ultimately serve as the final filter before production. I will call particular attention to Case Assumption #3 and the myth that "Preservation of Electronically Stored Information Is Getting Easier with the Passage of Time." Courts are not yet setting official precedent for sanctioning parties that fail to adequately preserve and identify potentially responsive information under spoliation rules, but the risk is quite alarming, creating the potential for a save-everything failsafe. Does this conflict with the precedent being set by the federal government to reduce detail and simplify retention schedules? How minimalistic can schedules become before they can no longer be defended? Finding the balance between transparency and risk is avoidable by the establishment and execution of not only legal hold procedures, but well-designed, standards-conscious electronic records retention policies and procedures.

Conclusion: On Marist, Rockefeller, and the API layer

Multifaceted, interoperable, complex content management systems have moved information culture out of the sphere of document management and into that of information management, no longer as an added value but as a default expectation. The ECM's API layer is crucial to its success as an interoperable repository software. The ability to write code bridging the myriad business and collaboration applications in operation to the ECM, automating the intelligent capture of electronic records, lessens the burden on both the end user and the records management team. Often this relies on careful management of the records by their owners, or administrators of the systems in which the records reside. Users can work from SharePoint and their records can be captured with little to no overhead, capturing both provenance and original order; the code written for this task maps metadata from document libraries to the ECM, capturing key fields like Site Title, URL, Date Created, and Date Modified, as well as the organizational hierarchy of the folders themselves. The same tool can be tweaked to capture anything from a Windows Explorer style hierarchy. The interoperability provided by application program interface (API) layers enables the fluid exchange of records and information from origin to archive, and we're already seeing significant progress in this arena.

This year it was announced that Marist College and the Rockefeller Archive Center were taking this technology to the next phase [2], working together to develop a platform, with an API layer, that can support the complexity of managing electronic records originating from near-limitless varieties of creators and account for rapid changes to technological contexts. If they are successful, information professionals could be witnessing the technical solution to the problem of electronic records authenticity and contextual preservation. ■

Resources Mentioned in the Article

[1] Redgrave, J. M., Peay, K. H., & Bulander, M. K. E. (2014). Understanding and contextualizing precedents in e-discovery: The illusion of stare decisis and best practices to avoid reliance on outdated guidance. *Richmond Journal of Law and Technology, 20*(2). Retrieved from http://jolt.richmond.edu/v20i2/article8.pdf

[2] Marist, Rockefeller Archive Center partner on open-source technologies for digital archival processes. (August 23, 2016). Retrieved from www.marist.edu/publicaffairs/rockefellerarchive2016.html