Tools of the E-Discovery Trade: Taxonomy Fundamentals for Managing Document Management Systems

by Sandra Nunn, MA, RHIA, CHP

Organizations rarely embark on the purchase of content management systems with a strategic content management initiative drafted, approved, and in place. Content management systems usually sneak in the back door as part of an emergent need to make specific domains of content more organized and accessible. If the organization is large enough, it may purchase multiple content management systems and develop them independently of each other, depending on the need for each of the disparate systems at the time.

Getting the Most out of ECM Tools

Failure to draft an enterprise content management (ECM) solution can cost an organization in wasted time, uncoordinated content sectors, redundant documents, unorganized responses to litigation requests, unskilled applications of legal holds, and misapplied efforts of skilled employees. The strategic potential that an ECM system brings to an organization is often lost on leadership until a module or two begins to produce a return on investment.

Once this occurs, the organization should reexamine these tools and strategically leverage them. This effort usually involves an assessment of which tools have been deployed, what content now resides within them, and who is responsible for the ownership of the content and its management.

If thinking strategically, the enterprise should begin to centralize the use of the content management tools, their management, and the standards used to structure the storage and retrieval of the documents and records archived within them.

The key to effectively retrieving diverse documents and records lies in the development of an organizational language or taxonomy. These tools permit documents, records, and other types of content to be “tagged” with metadata and indexed consistently for fast and efficient retrieval. This is particularly expedient when faced with litigation. Content that is systematically indexed against a standardized organizational taxonomy is relatively easy to retrieve, even when some of the older records may have been archived onto storage media.

Managing Document Management Systems

HIM professionals are most familiar with document management systems into which paper parts of medical records are scanned. These systems are content management systems whose sole content is usually clinical in nature, although there may be a business folder for billing information. The key to retrieval in such systems is usually the patient identifier. Taxonomy tools are not required because retrieval is usually based on one field (the patient identifier) for one type of content (the medical record).

Along with human resource and financial systems, often termed enterprise resource planning systems, these large clinical document management systems capture a large part of an organization’s unstructured information. They also capture most of the organization’s structured information, which then can be transmitted to data warehouse applications where decision support tools can be applied to massage and aggregate the structured information to make organizational forecasting and decision making easier and more accurate.

Structured and Unstructured Data

Although people use unstructured information every day, they rarely think of it in that manner. Whenever employees exchange e-mail, create spreadsheets and Word documents, or review images, they are engaged in the unstructured world of

information.

"Unstructured data consists of any data stored in an unstructured format at an atomic level. That is, in the unstructured content, there is no conceptual definition and no data type definition—in textual documents, a word is simply a word."[1] A clinical example is a written progress note created by a physician. An administrative example is a feedback report from the Joint Commission on hospital performance on a survey.

Work is also routinely carried out using structured data, "anything that has an enforced composition to the atomic data types. Structured data is managed by technology that allows for querying and reporting against predetermined data types and understood relationships."[2] When coders review the content of a medical record, they are analyzing the content and deriving codes that will become part of structured data (e.g., data that can be aggregated and from which decisions can be made).

Content management systems provide an electronic environment in which an organization can store clinical content, often scanned from unstructured paper documents, or an electronic library in which other administrative content can be archived and managed.

Data Categorization

Establishing content types to be archived into a content management system is one of the easier steps in managing enterprise content management systems. The content type (e.g., spreadsheet, e-mail, minutes, policies, protocols, guidelines, audits, etc.) occupies a single metadata field in a content management system. There may be subtypes such as "policy attachment," but a content type is assigned to *every* piece of content archived into a content management system.

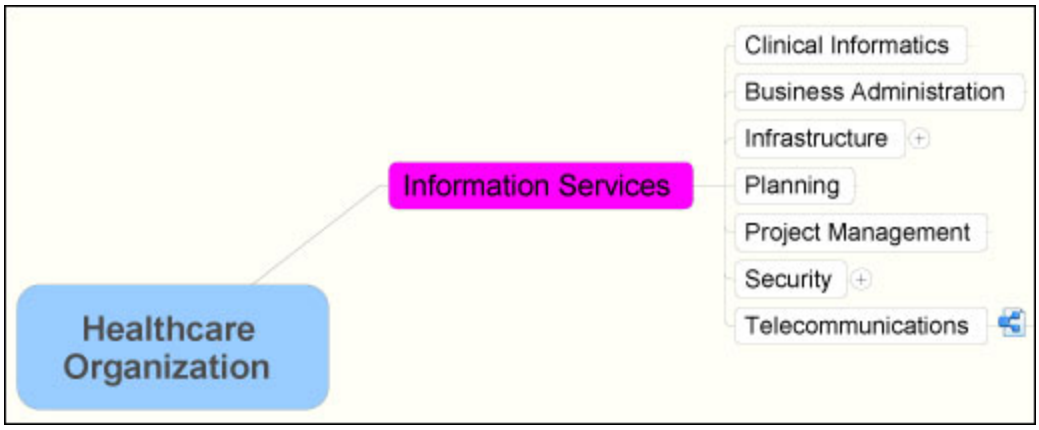
Part of the content management team's implementation work is to build a file plan that will describe all the metadata elements that will be used to tag each piece of content for later retrieval. Some metadata fields are automatically assigned, such as the content's submission date into the system, but others must be described in the data dictionary and defined by length and permitted characters or numbers. Examples of metadata fields include content author, content contributor, and content security level. Content containing protected health information would have a different security level than content meant to be available to all organizational employees.

Other metadata fields can help retrieve information for specific users, such as assigning a metadata field by Joint Commission chapter to allow retrieval of all content relating to compliance with that chapter of the Joint Commission standards.

Building the file plan requires a high-level view of the organization to determine how to categorize all unstructured organizational content. The unstructured content that will be archived into clinical systems is relatively straightforward and will be categorized into several source system buckets (e.g., clinical document management systems, imaging management systems, etc.). However, categorizing content for the remainder of the organization means understanding what the organization does beyond providing patient care.

Subtypes of Information Services Data

Data are often categorized by function rather than department name, which may change over time. In this simple illustration, information services is one functional category within a given healthcare organization, with multiple subcategories that comprise it. Further subcategories are possible within IS functions such as infrastructure and security.



Categorizing by Function

For most organizations, categorizing by the function the data perform allows it to catalog information without specific departmental or organizational names that may change over time. For example, most organizations carry out human resource management, which is a function. Within human resource management, the functions of employee relations, benefit management, compensation, and recruitment are often subfunctions of the master function. In the functional area of information services, subfunctions might include project management, information security, clinical informatics management, business information management, infrastructure, and telecommunications.

Content from these functional areas would be indexed by functional and subfunctional designations as well as by other metadata to allow for accurate and fast retrieval. The illustration "Subtypes of Information Services Data" (at left) displays a simple example of information services functional subtypes. The table on the next page shows how an information services policy related to the HIPAA security rule might be tagged for eventual retrieval.

As different domains of information develop content contributors, these staff will need more advanced knowledge management skills. Therefore, job descriptions and duties will need to be altered in each area. Human resources, information services, clinical services, and legal services will need skilled staff capable of interfacing a content management system and possessing the competencies necessary to manage the creation, archival, retrieval, metadata management, retention, and destruction of documents and records for each functional area.

HIM programs that offer training in content management system functionality will be embraced by organizations hungry for electronic content (documents and records) management.

The content management world is extending itself into the management of content types like e-mail and exploring how the law is evolving in response to records now described as electronically stored information in e-discovery terms. As HIM professionals reach beyond department walls to embrace electronic health records, a little more reach will find other records management domains within their grasp.

Sample HIPAA Security Policy Tags

Within the content management system, metadata elements are used to "tag" each piece of content for later retrieval. For example, an organization's security policy might be tagged with the metadata elements shown here.

Metadata Element	**Sample Data Element**
Author	John Doe

Effective Date	August 31, 2008
Contributor	John Doe's content contributor
Owner	Chief information officer
Date of Submission	September 2, 2008
Functional Category	Info Svcs-Security-HIPAA
Content Type	Policy
Regulatory Citation	Pertinent HIPAA citation
Reviewers	Security staff
Signer(s)	Security officer

Notes

1. Weglarz, Geoffrey. "Two Worlds of Data—Unstructured and Structured." *DM Review Magazine* (September 2004). Available online at www.dmreview.com/issues/20040901/1009161-1.html.
2. Ibid.

Sandra Nunn (snunn@phs.org) is enterprise records manager at Presbyterian Healthcare Services in Albuquerque, NM.

Article citation:
Nunn, Sandra L.. "Tools of the E-Discovery Trade: Taxonomy Fundamentals for Managing Document Management Systems" *Journal of AHIMA* 79, no.11 (November 2008): 54-55, 62-63.