

Why METADATA Matters

Records managers must be involved in the development and design of metadata structures to ensure that digital records are captured, maintained, retained, preserved, or destroyed in accordance with their organization's recordkeeping requirements

Pat Franks, Ph.D., CRM, and Nancy Kunde, CRM, CA

At the Core

This article

- ▶ Introduces basic metadata concepts
- ▶ Explains metadata's role in RIM
- ▶ Provides metadata schema and standards sources

As the volume of information born or stored digitally increases, concern also increases over the ability to ensure that records will continue to be accessible throughout their life cycle. To accomplish this end, information about the content, context, and structure of records must be recorded and managed. Many experts believe metadata is the key to meeting this challenge.

Metadata's Relevance for Records Managers

Records management is a professional discipline concerned with systematic planning, control, direction, organization, training, promotion, and other managerial activities related to the creation, maintenance, use, and disposition of records in properly documenting an organization's policies and transactions.

In carrying out their responsibilities, records managers have always been required to deal with metadata. Recording how, when, where, and why a given paper record was created is essentially capturing metadata. However, the growth of digital

records necessitates a different approach. Records managers must now understand the role metadata plays in the management of electronic records.

Information and records originate from a variety of sources and are created in a variety of formats. Records managers must be involved in the development and design of metadata structures to ensure that digital records are captured, maintained, retained, preserved, or destroyed in accordance with the organization's record-keeping requirements.

While managing paper records has been the undisputable domain of records managers, managing electronic records

requires teamwork. Today, it is common to see representatives of information technology, business units, records management, and legal services or compliance coming together to develop policies and procedures to address electronic records and information management issues. As part of the team, records managers are in a position to bring their expertise in categorizing, classifying, and indexing documents to the development of metadata.

Metadata Defined

ISO15489-1 Information and Documentation - Records Management, the first international records management standard, defines metadata as "data describing context, content, and structure of records and their management through time." In the contemporary information management field, however, metadata must also be defined and understood in terms of the function it performs, so one appreciates not only what it is, but how it operates. *The Minnesota Electronic Records Management Guidelines* provides one of the best descriptions and rationale for metadata:

Metadata allows users to locate and evaluate data without each person having to discover it anew with every use. Its basic elements are a structured format and a controlled vocabulary, which together allow for a precise and comprehensible description of content, location, and value.

The definition of metadata provided by the National Information Standards Organization is also helpful. It states:

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.

One way to think of metadata is in terms of provenance, which is the origin or source of something. For example, when buying a painting, it would be important for the buyer to be assured of the authenticity and past ownership of the artwork. The buyer would want to be able

The Foundations of Metadata

Categorization is the process of associating a document with one or more subject categories. *Cataloging* and *classification*, two terms that come from the library field, tend to be used interchangeably with categorization. Records managers sometimes refer to a records classification scheme as a file plan or record plan. Metadata associating information to categories helps provide context.

Indexing involves analyzing the document to determine terms that can be used to describe the document. Indexing terms are then used as pointers to a possible solution to a search. The more specific the index term, such as a unique identification number, the better the chances are of locating and retrieving the document quickly and easily. For paper-based records, indexing is a manual process. However, the volume of electronic records demands an automatic process.

In the United States, DoD 5015.2 *Design Criteria Standard for Electronic Records Management Software Applications* specifies baseline functional requirements that electronic records management software must meet. Included among the requirements are those for indexing; specifically, the records management application must uniformly create and maintain indexes for record identifier, date of record, classification code, addressee(s), creating organization, media type, and subject. It is also recommended that the record format, for example Microsoft Word, version 6, be recorded as well.

Because indexing documents with defined terms improves retrieval rates and reduces search time, it is important to develop a *controlled language* (or controlled vocabulary). A **controlled language** is used to standardize terminology and includes tools such as business classification schemes, thesauri, and taxonomies. Controlled language is most often used in classification and indexing, but it can also be applied to such activities as retention, disposal, and content management. Controlled language can be the source for allowed terms for particular metadata elements. For example, a metadata element may be "country" and the language used to describe it could be selected from a list of countries; for example, United States. In this case, the controlled language ensures anyone selecting the United States uses those two words instead of U.S., US, or US of A.

to verify that they are getting what they paid for. The proof behind the artwork is the metadata – verification of who painted it, when it was painted, the trail of ownership, and other characteristics about the painting or artwork. Essentially, the records manager needs to be able to do the same thing for records. Records need to be uniquely identified and authenticated over time and space. Metadata that associates records to categories or classifications helps provide context. (See sidebar: “The Foundations of Metadata.”)

In addition to its use for locating and retrieving records and information consistently and accurately, metadata has a direct relationship to records management functions. It facilitates data sharing, ensures the authenticity of the content, describes the content, and facilitates implementation of retention and disposition. For electronic records management, metadata is critical to ensuring meaning, manageability, and preservation of electronic records.

Types of Metadata

Though the types of metadata vary from one environment to another, metadata generally falls into three types, or categories.

1. **Descriptive metadata** describes a resource for purposes such as discovery and identification. Metadata in this area can include such elements as author, title, and abstract. Elements that one would find in a typical bibliographic record are examples of metadata.
2. **Structural metadata** indicates how compound objects are put together. They identify data format, media format or the type of data representation and file types, hardware and software needed to render the data, and the compression method and encryption algorithms used, if any.
3. **Administrative metadata** provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. Very often, there are

subsets in this category. The two most typically included are rights management and preservation metadata.

How Metadata Works

Metadata enables the creation, registration, classification, access, preservation, and disposition of records through time and across domains. It can be used to identify, authenticate, and contextualize records, the people, processes, and systems that create, manage, maintain, and use them, and the policies that govern them.

Metadata can be used to describe an object so that it can be located when needed. It can help organize electronic records, facilitate interoperability across systems, provide digital identification, and support both archiving and preservation.

At the point of records creation or capture, metadata should include information that relates the date and time of creation or receipt, the agents involved in creation, the record structure, the record form, and the relationship between the data or format elements that comprise the records. (See Figure 1: “Metadata Capture and Use.”)

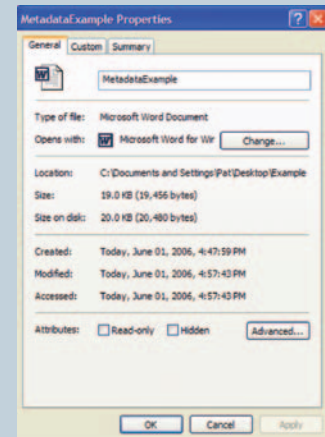
Additional metadata can be added to the file throughout the life cycle of the record by the originating system, an electronic document management system, an enterprise content management system, or an electronic records management system.

After records capture, metadata continues to accrue to define changes in the logical and physical structure of records, define the changes in technical attributes, describe new context in which records are used, and document new relationships with other records or aggregations.

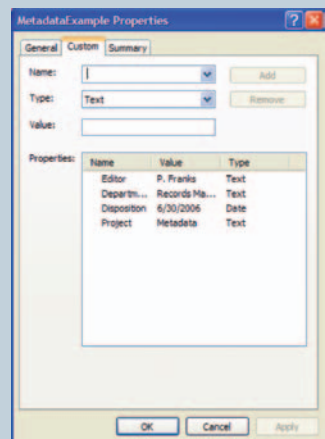
Metadata has been referred to as the information discovery tool. As the information infrastructure of many organizations has become complex, silos of information – that is, isolated information resources with no connectivity to other resources in the organization – are commonplace. Metadata forms bridges between the silos and permits employees to quickly access consistent data on a variety of organizational functions and topics.

Figure 1: Metadata Capture and Use

Any saved Microsoft Word document has a properties box that reveals the file name, the type of document, and the program needed to open it. The location information in this example adds the name of the folder, i.e., Pat\Desktop\Example. The date and time created are recorded, as are the dates and times the file was accessed and modified. This information is appended by the system.



Additional information can be added manually in the same property box. For example, the user can click on the custom tab, select from a drop down list of fields (metadata elements) and data types, then add the value desired. The four metadata elements in the second column were inserted manually.



Both the automatic and manually captured fields are metadata that is transmitted along with the document as into an electronic document management system or an electronic records management system, either of which can extract the metadata to automatically populate database fields.

How to Develop Metadata Sets or Schema

Formalizing metadata involves the development of a schema – sets of metadata designed for a specific purpose – that defines the types of data that are expected to be included in an appli-

cation. It is not uncommon for schemas to include the descriptive metadata, structural metadata, and administrative metadata types discussed above.

Developing metadata depends on the types of metadata needed and the tools available to do the job. It is impor-

tant for the records manager to be cognizant of the business processes involved and the need to partner with information technology specialists and others in the organization in determining the types of metadata needed. Although one member of the team may have more input in certain instances, deciding how to handle metadata creation and management takes a cooperative effort.

For records professionals seeking to understand metadata within the realm of given information systems in their organizations, obtaining answers to the following questions will likely lead to metadata.

- Who owns or is responsible for the particular data?
- Who or what is responsible for the origin of the data?
- Who has rights to access the data?
- Who owns or developed the data definitions? When were they last updated?
- When was the data created or entered into the system?
- What quality control was performed on the data?

Several approaches are possible. Organizations can use an existing schema or expand upon an existing schema by adding metadata elements. They may also develop their own schema or use one designed specifically for their industry.

Established Metadata Schemas

Many schema sets have been formalized and are recognized as official standards. A number of industries have developed industry-specific metadata element sets as well. (See Sidebar: “Metadata Standards and Guidelines.”)

Professional associations recognize the need for additional metadata elements to address specific technical requirements. *ANSI/AIIM/ARMA TR48-2004 Framework for Integration of Electronic Document Management Systems and Electronic Records Management Systems* describes metadata needed for the integration of

Metadata Standards and Guidelines

Below is a sampling of the wide variety of metadata standards and guidelines. The Metamap project at the University of Montreal is a good resource to guide the reader through the world of metadata standards and practices. Access it at <http://mapageweb.umontreal.ca/turner/meta/english/index.html>.

- Dublin Core Metadata Initiative (DCMI)
<http://dublincore.org>
- Australian Government Locator Service (AGLS) Metadata Element Set
www.naa.gov.au/recordkeeping/gov_online/agls/agls_all_jurisdictions.html
- VERS Metadata Scheme
PROS 99/007 (Version 2) Specification 2 Public Record Office Standard.
Public Records Office Victoria
www.prov.vic.gov.au
- OAIS (Open Archival Information System)
<http://nssdc.gsfc.nasa.gov/nost/isoas/>
- Publishing Requirements for Industry Standards (PRISM)
www.prismstandard.org/news/2000/0209.asp
- Content Standard for Geospatial Metadata and Spatial Data Transfer Standards
www.fgdc.gov/standards/standards_publications/
- Metadata Encoding and Transmission Standard (METS)
www.loc.gov/standards/mets
- Preservation Metadata: Implementation Strategies (PREMIS)
www.oclc.org/research/projects/pmwg
- Extensible Markup Language (XML)
www.w3.org/XML/
- ISO 15489, *Information and documentation: Records management*, Parts 1 and 2
www.iso.org or www.arma.org/bookstore
- ISO 23081-1:2006, *Information and documentation – Records management processes – Metadata for records—Part 1: Principles*
www.iso.org or www.arma.org/bookstore

electronic document management systems and electronic records management systems.

ARMA International is developing several standards or guidelines that incorporate or address metadata, including ones on these topics:

- Program planning, requirements, and procedures for the conversion process, which will identify metadata necessary for the conversion of digital records so as to preserve the integrity of such records as evidence of business transactions
- Records management issues in collaborative environments, which will address myriad recordkeeping issues associated with collaborative work tools
- Issues and considerations for managing electronic messages as records, which will provide guidance for those who deal with practical managerial issues involving the informational content of e-mail and other electronic messages

Customized Metadata Schemas to Meet Business Needs

Organizations can also develop their own metadata schema to aid sharing information resources across the enterprise. The successful development of metadata schema involves the devel-

Figure 2: Metadata Schema Example

Name (Metadata Element)	Unabbreviated Name	Comment	Sequence Number	Data Type	Length	Scale	Nulls	Default Type
DMV_PRDT_ID	DMV PRODUCT ID	The specific product key assigned by various DMV bureaus which is used to uniquely identify the product issued.	1	DECIMAL	11	0	N	N
PRDT_STUS_CD	PRODUCT STATUS CODE	A code representing the status of a product. (Logical foreign key).	2	CHAR	3		Y	Y
APLC_CD	APPLICATION CODE	The type of application made. Examples are: Original issuance, renewal, re-issuance, etc. (logical foreign key).	3	CHAR	3		Y	Y
BUSN_TXN_TMS T	BUSINESS TRANSACTION TIMESTAMP	The date and time assigned to a business transaction by the controlling application and reflected on all tables created or updated by that transaction.	4	CHAR	8		N	N
DL_CLS_FL	DRIVER LICENSE CLASS FLAG	Indicates if a license is an old system license type or contains multiple operating classes on a single document.	5	CHAR	1		N	N
DL_BTCH	DRIVER LICENSE BATCH	The batch number used to file a driver license application.	6	INTEGER			N	N
DL_PREV_BTCH	DRIVER LICENSE PREVIOUS BATCH	The previous batch number used to file a driver license application.	7	INTEGER			Y	Y
DL_ISS_DT	DRIVER LICENSE ISSUE DATE	The date a product was issued.	8	DATE			N	N

opment of a metadata application including:

- Defined and effective data best practices (or internal standards) that provide definitions of the characteris-

tics of the data elements going into the application

- A set of metadata business rules that ensures accuracy and consistency when the metadata schema is applied

- A reliable data creation process including the identification of those who create and maintain metadata, the tools that will be used, resources, procedures, and quality control to produce the desired outcome

Figure 2 shows the Wisconsin Department of Transportation's well-defined metadata structure to support the driver's license process.

Why Metadata Matters

Organizations in public and private sectors understand that metadata is central to managing business-critical information. Metadata aids organizations in re-using and re-purposing data and thereby avoids the expensive process of re-collecting it. It facilitates the decision-making process by providing for fast and efficient retrieval of information. Metadata serves as the backbone of electronic records management systems by providing consistent identification of records, preserving their authenticity, and implementing retention and disposition requirements. Finally, as organizations begin to address long-term digital storage and preservation, metadata plays a vital role in consistently identifying key information necessary for conducting the conversion/migration processes, such as hardware and software used to create the digital information object.

Records managers must be knowledgeable about metadata functionality and how it can be used to facilitate their work. This knowledge base is an important asset that can be brought to the new collaborative structure necessary for contemporary electronic records management. Records managers must be able to communicate with information technology staff and vendors when creating electronic records management systems and with business units and legal services or compliance when developing policies and procedures.

As the *Minnesota Electronic Records Management Guideline* states, "Whatever

you want to do with the information (e.g., protect its confidentiality, present it as evidence, provide citizens access to it, broadcast it, share it, preserve it, destroy it) will be feasible only if you

and your partners can understand and rely upon the metadata describing it. Using metadata effectively means understanding and applying the standards appropriate to your needs." ■

Pat Franks, Ph.D., CRM, and Nancy Kunde, CRM, CA, are members of ARMA International's Standards Development Committee (SDC) and have been actively involved in various SDC projects, including one on metadata. They will be developing a primer on metadata for records managers. Franks teaches online for the State University New York Learning Network. She can be contacted at pfranks@stny.rr.com. Kunde is records officer for the University of Wisconsin-Madison. She can be contacted at nkunde@library.wisc.edu.

For More Information

"A Case Study in Metadata Harvesting: the NSDL." William Y. Arms, Naomi Dushay, Dave Fuller and Carl Lagoze. *Library Hi Tech*, Vol. 21, No. 2, 2003, pp 228-237.

"An Operational Metadata Framework for Searching, Indexing, and Retrieving Distributed Geographic Information Services on the Internet." Ming-Hsiang Tsou. *GI Science*, 2002, pp. 313-332.

Digital Libraries: Metadata Resources. This is an online resource on metadata developed by the International Federation of Library Associations and Institutions. Available at www.ifla.org/II/metadata.htm (accessed 24 July 2006).

FirstGov.Gov. This site is the official U.S. government portal that provides a gateway to government information. A search of this site will result in a listing of numerous guidelines and best practices related to metadata in government. Available at <http://firstgovsearch.gov/search?v%3Aproject=firstgov&query=metadata+guidance> (accessed 24 July 2006).

"Preservation Metadata." Michael Day. Prepublication draft of chapter published in G. E. Gorman and Daniel G. Dorner (eds.), *Metadata applications and management, International Yearbook of Library and Information Management, 2003-2004*. London: Facet Publishing, 2004, pp. 253-273. Available at www.ukoln.ac.uk/metadata/publications/iylim-2003/ (accessed 24 July 2006).

References

International Organization for Standardization. *ISO15489-1 Information and Documentation - Records Management - Part I: General*. International Organization for Standardization: Geneva, Switzerland, 2001.

Minnesota State Archives. *Minnesota Electronic Records Management Guideline*. Available at www.mnhs.org/preserve/records/electronicrecords/ermetadata.html (accessed 24 July 2006).

National Information Standards Organization. *Understanding Metadata*. Bethesda, MD: National Information Standards Organization, 2004.

State Library of Tasmania. *Understanding Metadata*. Available at www.servicetasmania.tas.gov.au/papers/understanding_metadata.htm (accessed 24 July 2006).