



A DOCUMENT MANAGEMENT SYSTEM IS ONLY AS GOOD AS ITS INDEX **BY VIKING SOFTWARE SOLUTIONS**

In a document management system, an index functions as the means to retrieve documents. Without retrieval capabilities, the content would merely take up space. Given this vital role, companies should give indexing careful consideration when implementing a new system.

Index Validation

We are asked time and again about various aspects of document management:

- Web-based storage vs. in-house server
- Simple index vs. advanced taxonomy with metadata
- All-in-one system vs. interfaced open systems
- Searching time vs. indexing time

In all of these comparisons one thing emerges as the single most important factor; the validity of the index assigned to the document images. 85% correct isn't adequate, nor is 90%, or even 98%. An index must be 100% correct 100% of the time, or there will be lost images. These losses translate into to wasted time because you will not know that you cannot locate a document until after you have spent hours searching for it.

Think of the time it takes to locate a paper document that is missing. Searching for an image can be even more time-consuming. Therefore, indexing correctly is an essential part of document management.

[AIIM](#) president John Mancini's *Capture: On-Ramp for ECM*¹ describes the complete process of Electronic Content Management. In his diagram, at the intersection related to indexing, he says, "Indexing is NOT optional. No other way to find documents." Another essential part of the onramp, according to Mancini's diagram, is quality assurance. "Electronic images must be double-checked."

Indexes can consist of a few key fields per document or the full text. There are several different ways to capture the information that goes into the index. Many companies use automated character recognition on a number of levels from barcode to ICR. Others use key-from-image operators. Whether the initial process of gathering information is automated or manual, document management systems need a quality assurance process that guarantees accurate indexes. The bottom line for any type of data capture is that the system must have a way of validating the information through a second operator, a database lookup, or both, if you want a useful, accurate index.

Index Enhancements: Taxonomy and Metadata

For years, the records management industry has had a consensus on best practices for organizing filing systems. These same organizational structures are the road map to follow when establishing document index structures. This kind of in-depth organization is not always available during the scan stage of document management. Taxonomy is frequently added and enhanced during the quality assurance step. At that time, documents can be analyzed on a case-by-case basis and more data fields are added as necessary to define the document contents in greater depth.

¹ John Mancini, "Capture: On-Ramp for ECM," AIIM, November, 29, 2006,
<http://www.aiim.org/article-webinar.asp?ID=32261>

In his presentation at the 2005 [ARMA](#) conference, Lloyd Dugan suggested that 10 or fewer key data fields are generally adequate for records management needs. However, additional descriptive information can also be used as non-key data fields.² There should be a balance between the number of indexing parameters and facilitation of error-free entry of index data. To the question of how much time should you spend inputting/storing information, we reply, as long as it takes to be able to find every document image quickly without error. Once you lose even one image and spend the time trying to find it, you have lost any gains achieved by taking shortcuts up front.

Are there index data or metadata standards? The Dublin Core Metadata Initiative (www.dublincore.org) offers the following guidelines:³

- Title: The name given to the resource by the Creator or Publisher.
- Creator: The person(s) or organization(s) primarily responsible for the intellectual content of the resource.
- Subject: The topic of the resource or keywords or phrases that describe the subject or content of the resource.
- Description: A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources.
- Publisher: The entity responsible for making the resource available in its present form, such as a publisher, a university department, or a corporate entity.
- Contributors: Person(s) or organization(s) in addition to those specified in the Creator element who have made significant intellectual contributions to the resource but whose contribution is secondary to the individuals or entities specified in the Creator element.
- Date: The date the resource was made available in its present form.
- Resource Type: The category of the resource.
- Format: The data representation of the resource.
- Resource Identifier: String or number used to uniquely identify the resource.
- Source: The work, either print or electronic, from which this resource is derived, if applicable.
- Language: Language(s) of the intellectual content of the resource.
- Relation: Relationship to other resources.
- Coverage: The spatial locations and temporal durations characteristic of the resource.
- Rights: A link to a copyright notice or rights-management statement.

Software Capabilities

Many times, organizations find that a single product will not provide the depth needed for acceptable quality assurance of metadata. For example, a large energy company in the Southwest spent tens of thousands of dollars on their enterprise-level archival system only to find that they were spending measurable, unnecessary time searching for documents because indexes were incorrect or incomplete. The indexing capabilities of the system were inadequate. By adding specialized software for their onramp that enhances the index with taxonomy and metadata then verifies the information before it is stored, the company experienced immediate return on their investment. This quality assurance step has proven to be a time saver even after factoring in the time investment for validating the metadata. The search time savings outweighs the time spent verifying and adding additional information to make searching the archives easier.

² Lloyd Dugan, "Information Technology and Management: IT 101 for RIM Practitioners," presented at ARMA International 50th Annual Conference and Expo held in Chicago, Illinois September 20, 2005, 34.

³ Dugan 36.

Evaluating Your Requirements

Here are some key questions companies need to ask when considering document management indexing solutions:

- What problem are we trying to solve?
- Are our needs for retrieval growing?
- How does our plan for the next five years fit into our existing search/retrieval time allocations?
- What are the costs of adding a quality assurance step?
- What are the costs of NOT adding a quality assurance step?

About Viking Software Solutions

Viking Software Solutions has developed data entry-related software since 1980. Viking's philosophy of producing a quality product and providing incomparable technical support has generated a loyal client base, encompassing both government and business enterprises...large and small...local and worldwide.

Viking's mission statement is twofold:

- Provide best practice human interface tools for the most demanding data and document processes.
- Make those specific applications easy for end-users, non-programmers, and data entry personnel to set up and use

Viking Software Solutions can be found on the web at <http://www.vikingsoft.com>.