Making Content Easy to Find

DC2010 – Pittsburgh, PA
Betsy Fanning
AIIM
Who is AIIM?

The leading industry association representing professionals working in Enterprise Content Management (ECM). We offer a Membership Value Program Focused On:

- Market Education
- Peer Networking
- Industry Advocacy
- Professional Development
About AIIM Standards

• ANSI Accredited
• ISO TC 171, Document Management Applications – Secretariat
• ISO TC 171, Document Management Applications, SC2, Application Issues – Secretariat
• U. S. TAG (Technical Advisory Group) to ISO TC 171 Administrator
• Industry Standards Developer – AIIM Recommended Practices (ARP)
• Open Source Standards for Document Management
• Liaison Relationships
What is ECM?

The tools and technologies used to:

- **Capture** — move content (in any form) into your repositories for reuse or retirement
- **Manage** — move it around the enterprise to drive key applications and processes
- **Store** — put it in a logical place for easy access
- **Preserve** — long-term archival and storage
- **Deliver** — get to the right audience on the right device

...documents and content related to organization *processes*.
What is content?

• Content comes in a variety of formats:
  • Unstructured content such as
    – Office files (e.g., word processing, e-mail)
    – Imaged documents
    – Media files
    – Complex documents (e.g., CAD files)
  • Structured content (often referred to as “data”) stored in database tables
    – Or increasingly, XML
  • Semi-structured content such as HTML
What is Expected?

- Information should be easy to discover or locate
- Information access is about helping users find documents that satisfy their information needs
- Remember, someone may be looking for something they’ve never seen or touched before
- Information should be easy to tag or assign the metadata
Organizational issues

Which of the following organizational issues have you experienced with your SharePoint implementation?

- No management plan as to which features we use and where
- Haven't defined how SharePoint sits with existing DM/ECM/RM systems
- Hard to reach agreement on standard classifications, etc.
- No defined ownership of team sites for management, end-of-life, etc.
- We have/we had insufficient expertise in ECM, DM, RM
- It is driven by the IT department with no real IM input
- Team sites are proliferating with little control
- Being implemented by individual departments with no coordination
- No real policy from Microsoft on dealing with emails in SharePoint
- None of these

40+% no planning or strategy
26% lack of information management expertise

N=362 SharePoint using or implementing
Know What You Have

- In order to improve information access, you need to know
  - How much content you have
  - What types of content you have, and its relative value
  - What content needs to be archived, retained, or deleted

- In order to undertake a successful ECM/WCM/RM/Search implementation or improvement effort you need to know:
  - What documents you possess
  - Who “owns” the content in order to determine proper security, roles and permissions
  - Who or what creates content in order to properly tag/index and otherwise contextualise and enrich content

- Ultimately, you need to create an overall Content Model
What is a Content Model?

- Components or “elements” that make up a body of content
  - The folder or “meta”-structure of a repository or enterprise information set
  - The document types
  - Associated metadata
  - Elements within a (structured) document

- A framework applied to content to create relevant information
  - Making those related pieces useful to the people who need it

This is how you need to see and think about content.
What is a Metadata Strategy?

• Identification and understanding of different metadata types and their purpose
• Synchronisation and adoption across a department, project, and ultimately the entire enterprise;
  – Agreement on terms, labels, and meanings
• Understanding of people, processes, and systems applying and interacting with metadata and vocabularies
• Understanding who owns various metadata and structures
• Planning for maintenance and changes

Source: Ed Stevenson, Really Strategies, Inc.
Benefits of Strategy

• Consistent use of metadata structures across the enterprise makes the metadata more powerful
  – Information and systems become more interoperable
• Lesser chance of ambiguous terms when metadata and its purposes are defined, helping to ensure quality in the metadata
• Understanding of how metadata changes can affect downstream processes
• Identification of gaps in what should have more metadata
• Communication of metadata information to others who may find uses for the content outside its original area
• Realistic appreciation for level of effort to “tag” or “index” content
• Establishment of someone or some group with centralised knowledge of the metadata processes

Source: Ed Stevenson, Really Strategies, Inc.
Governance

Which of the following governance policies do you have in place for SharePoint usage?

- Policy on roles, administrative rights and access
- Policy on who can set up a team site, and their responsibilities
- Approved site design templates
- Quotas (numbers/storage) by user, by sites, etc.
- Guidance on corporate classification and use of content types and columns
- Restrictions on stored content with regard to security – e.g., HR, Finance
- Acceptable use policy wrt other staff
- Guidance on use and longevity of team sites, blogs and projects
- End-of-life policy for sites and contents
- Policy on use of third party products and web parts
- Retention policies
- Policy on dealing with emails and email attachments
- Legal discovery procedures
- None of these

55+% trying to address team-site sprawl
22% guidance on classification and metadata
16% or less on retention, legal discovery - and emails!

N=391, Using or implementing, May 2010
Why?

• Digital content is expanding at almost unmanageable rates
  – New information worldwide has been increasing on average 30% a year (doubling every three years)*
  – Getting access to the right information is an increasingly acute challenge for enterprise employees and customers alike

• Better Information Organisation leads to better Access

*http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/
ECM Drivers

When you consider your document and records management projects and priorities, what is the most significant business driver for your organization? (Check only ONE)

- Improve efficiency: 45%
- Optimize business processes: 20%
- Compliance: 28%
- Mitigate risk: 15%
- Reduce costs: 10%
- Enable collaboration: 5%
- Improve customer service: 2%
- Faster turnaround/Improved response: 1%
- Competitive advantage: 0%

N=680 Non-trade,

Efficiency and business process: 45%
Compliance and risk: 28%
ECM Drivers

Thinking about the compliance benefits of ECM and Records Management, which of the following are the TWO most important compliance drivers in your organization?

1. Customer/supplier litigation
2. Financial reporting and audit

N=680 Non-trade,
Metadata and ECM

• Metadata often acts as a “great unifier” in the area of content technologies and enable them to work together

• Many content management systems depend on solid library and categorisation services order to add significant value
  – Essential for organising any large content corpus
  – Required for meaningful records management
  – Critical to effective findability

• How you choose to design the repository, and how the system you choose can use certain repositories and content structures, greatly influence the business value you can realise
ECM Drivers – content types

How are the following content types managed and archived in your organization?

- 40% with documents in ECM/DM/RM system (scanned and electronic)
- 15% storing emails in ECM/DM/RM, 29% in EMM system

- Scanned documents
- Electronic documents
- Faxes
- Emails
- Photo images
- Active web pages
- Archived web pages
- Audio recordings
- Video/CCTV recordings
- Internal blog posts
- Telephone recordings
- Instant messages
- External blog posts
- Twitter posts

In an ECM/RM system | In a stand-alone management system | In a well organized fileshare

N=604 Non-trade,
ECM Drivers - Electronic

How confident are you, that if challenged, your organization could demonstrate that your electronic information (excluding emails) is accurate, accessible, and trustworthy?

Electronic (not emails)

41% Slightly or not at all confident

N=607 Non-trade,
ECM Governance

Who is the highest person in your organization who has specific reporting authority, or management ownership, of document and records management?

- 28% have a CIO who really is a CIO
- Plus 11% with a CRO
- 39% have no board-level ownership

N=645 non-trade
How would you describe your use of SharePoint in the following ECM areas?

- Collaboration/workspaces/team sites
- Document management (check-in/check-out)
- File share replacement
- Web - internal/intranet/staff-facing sites
- Forums, Blogs & Wikis
- Portal/Company news-site
- Portal to multiple repositories
- Enterprise Search
- Business Process Management (Workflow)
- Staff profiles/directories
- Forms processing – internal, e.g. HR, expenses
- Records management
- Scanned image management
- Web - external/www
- Case Management
- Forms processing – external, via capture
- Legal discovery
- Email management

Most people manage documents in SharePoint.

Lots are using it as a portal to other systems.

RM low but set to rise.

Capture and emails v. low.

N=436 SharePoint using or planning May 2010.
ECM DC Use

- Many content technologies are now offering Dublin Core standard repositories and content formats out of the box
- SharePoint uses Content Types
  - Tied to business process or document type
  - Shared across site collections
- DC is used with file formats – PDF and PDF/A
So, what happens with no metadata?
For humans, adding metadata means work

- Taggers may not see the ultimate benefit of metadata themselves
  - Benefits tend to accrue to the enterprise and content consumers
- To be sure, clerical staff can be forced to index
  - In some imaging systems, it is a specialised skill
- In other cases: “Not my job”
- Sometimes humans provide incomplete or inaccurate metadata

So a question arises:

- Is there a way to get machines to add metadata for us?
Indexing a Scanned Image

- A person adds data to the record of a scanned document.
- That data is (typically) stored in a separate database, associated with the image file.
- Later, the scanned document can be retrieved by number, name, or date.
- The Notes field will likely need to be indexed by a text search engine to become searchable.
Capturing Metadata

![Booking Form 1 Properties window](image)

- **Name:** Client, Task completed, Department, Destination, Disposition, Division
- **Types:** Data
- **Value:**
  - Owner: AIIM
  - Purpose: Register for...
  - Language: English
  - Data created: 3/1/2008

*aiim*

Find, Control, and Optimize Your Information
PDF and Metadata

- General information about the document, i.e., title, author, creation and modification dates
- Used to help search for documents in external databases
- PDF metadata may be stored in document information dictionary or a metadata stream
PDF/A Metadata

- Requires the use of Extensible Metadata Platform (XMP)
  - Proprietary, but open format – soon to be ISO
  - Used for metadata creation, processing and interchange
  - Based on restricted form of Resource Description Framework (RDF) – W3C standard
  - Fosters re-use, re-purposing across domains
  - Enables metadata capture, preservation, and propagation across devices, applications, file formats
  - Not limited to a specific schema
Document Information Dictionary

- Title
- Author
- Subject
- Keywords
- Creator
- Producer
- CreationDate
- ModDate

Metadata entries are optional and are deleted if not provided.
## DocInfo – XMP Crosswalk

<table>
<thead>
<tr>
<th>Document information dictionary</th>
<th>XMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry</td>
<td>PDF type</td>
</tr>
<tr>
<td>Title</td>
<td>text string</td>
</tr>
<tr>
<td>Author</td>
<td>text string</td>
</tr>
<tr>
<td>Subject</td>
<td>text string</td>
</tr>
<tr>
<td>Keywords</td>
<td>text string</td>
</tr>
<tr>
<td>Creator</td>
<td>text string</td>
</tr>
<tr>
<td>Producer</td>
<td>text string</td>
</tr>
<tr>
<td>CreationDate</td>
<td>date</td>
</tr>
<tr>
<td>ModDate</td>
<td>date</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Property</th>
<th>XMP type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>dc:title</td>
<td>Lang Alt</td>
</tr>
<tr>
<td>Author</td>
<td>dc:creator</td>
<td>seq ProperName</td>
</tr>
<tr>
<td>Subject</td>
<td>dc:description[“x-default”]</td>
<td>bag Text</td>
</tr>
<tr>
<td>Keywords</td>
<td>pdf:Keywords</td>
<td>Text</td>
</tr>
<tr>
<td>Creator</td>
<td>xmp:CreatorTool</td>
<td>AgentName</td>
</tr>
<tr>
<td>Producer</td>
<td>pdf:Producer</td>
<td>AgentName</td>
</tr>
<tr>
<td>CreationDate</td>
<td>xmp:CreateDate</td>
<td>Date</td>
</tr>
<tr>
<td>ModDate</td>
<td>xmp:ModifyDate</td>
<td>Date</td>
</tr>
</tbody>
</table>
Metadata Nomenclature

The following are examples of metadata tags:

<dc:element>Content</dc:element>
<pdf:element>Content</pdf:element>
<xmp:element>Content</xmp:element>
Example of Description Metadata
Example of Advanced Metadata
Future

• IDC, Digital Universe Report
  – By 2020, 25 quintillion information containers
• How will we find the information we need when we need it?
Questions/Contact

• Betsy Fanning
• Ph: +1.301.755.2682
• Skype: betsy.fanning
• Email: bfanning@aiim.org
• Twitter: bfanning
• LinkedIn: www.linkedin.com/in/betsyfanning