

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?

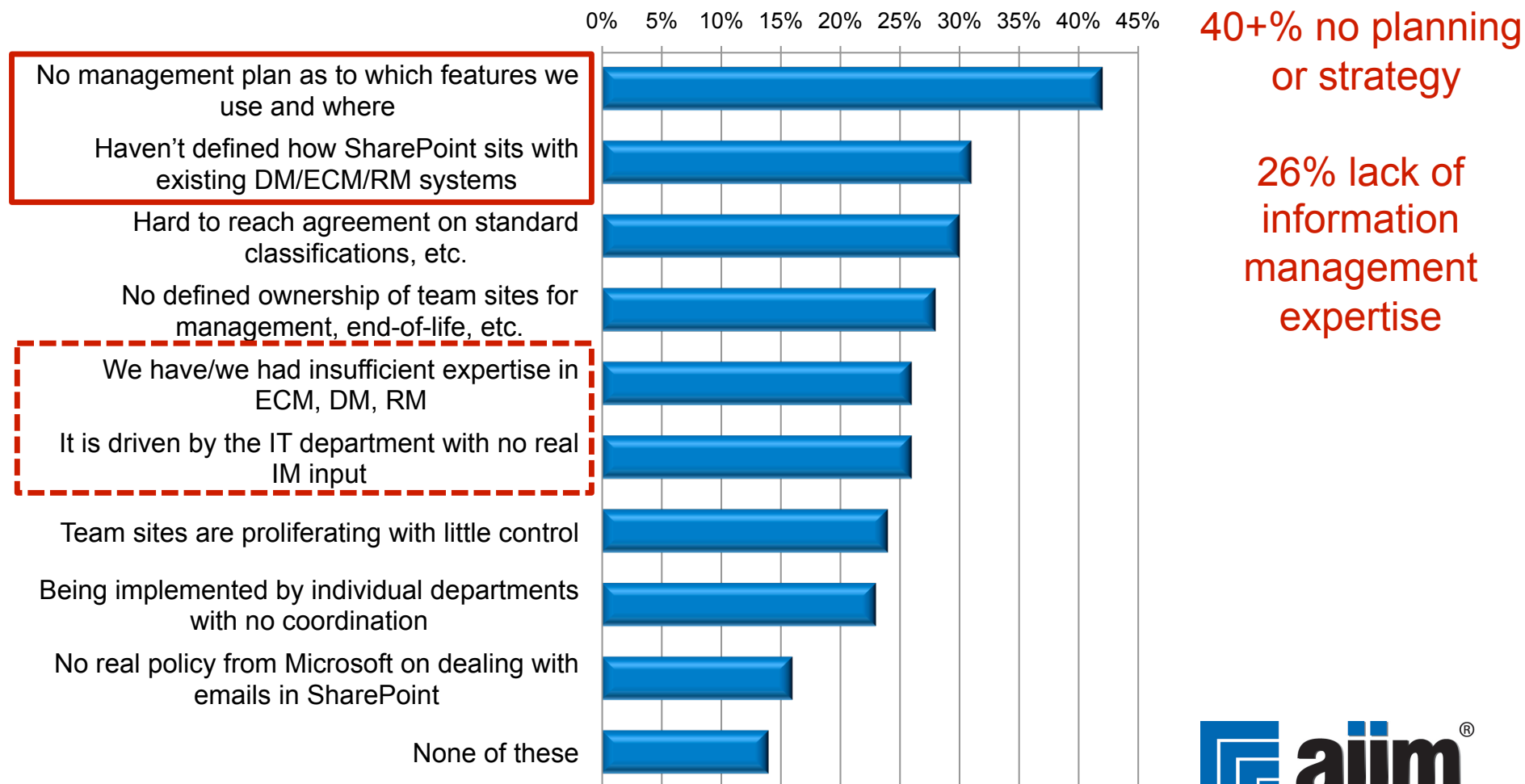
- Content comes in a variety of formats:
 - Unstructured content such as
 - Office files (e.g., word processing, e-mail)
 - Imaged documents
 - Media files
 - Complex documents (e.g., CAD files)
 - Structured content (often referred to as “data”) stored in database tables
 - Or increasingly, XML
 - Semi-structured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



40+% no planning or strategy

26% lack of information management expertise

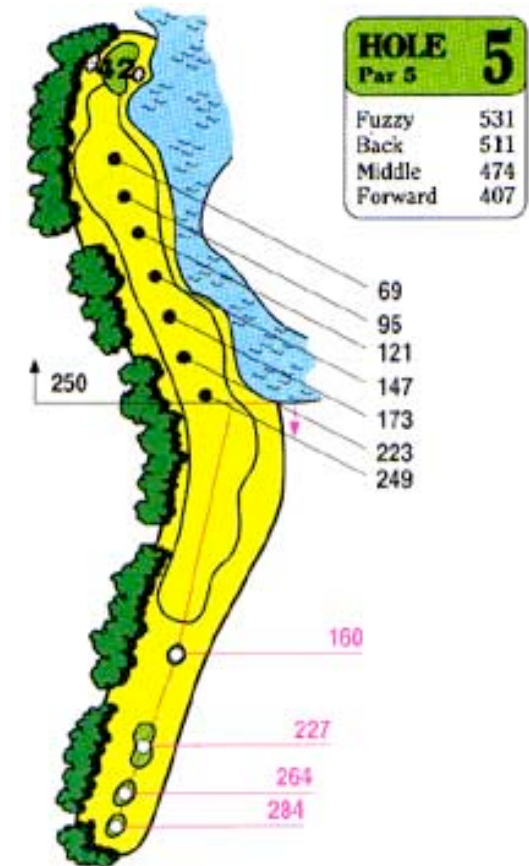
N=362 SharePoint using or implementing

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.



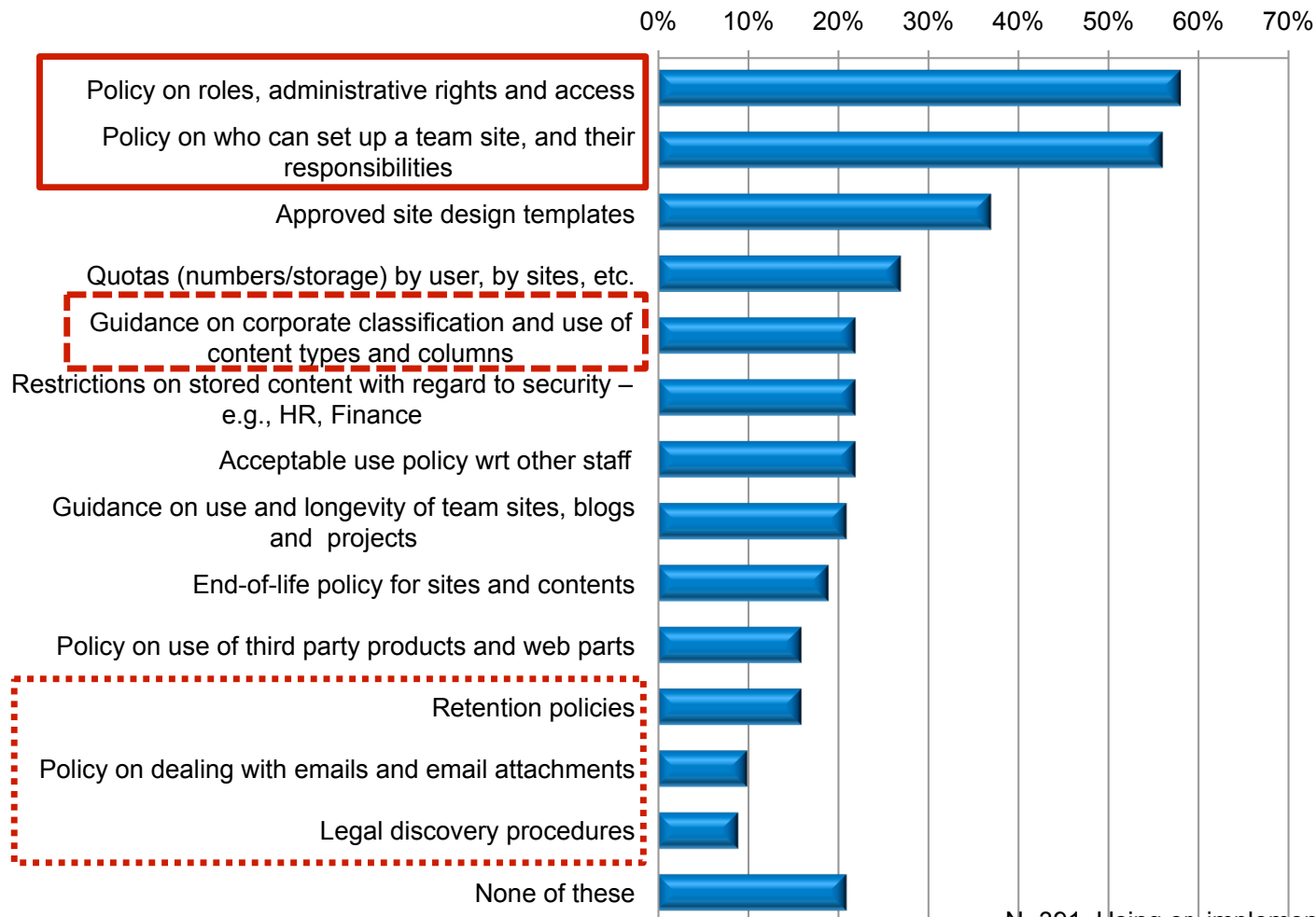
Benefits of Strategy

- 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?



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?

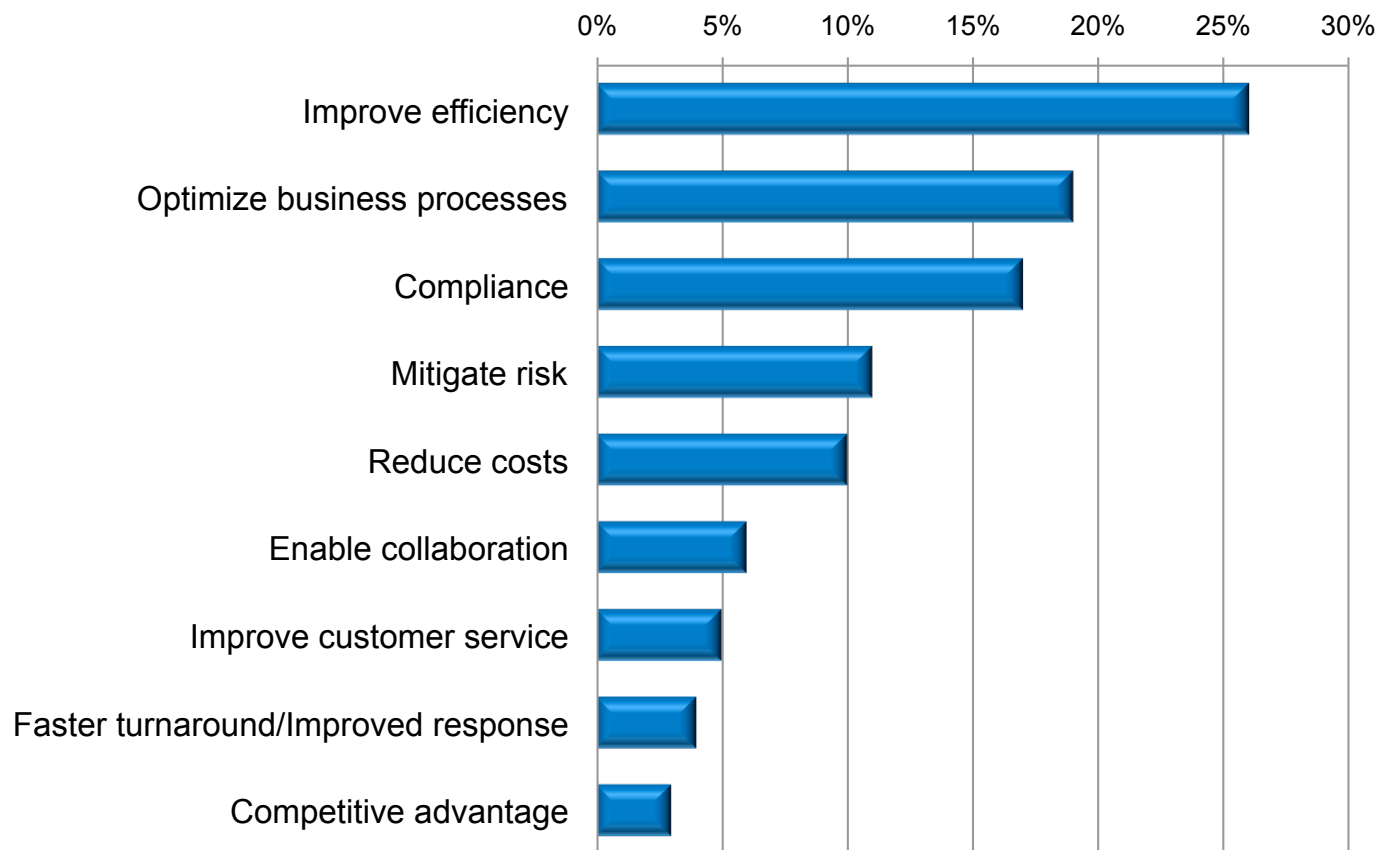
- Digital content is expanding at almost unmanageable rates
 - New information worldwide has been increasing on average 30% a year (doubling every three years)*
 - Getting **access** to the right information is an increasingly acute challenge for enterprise employees and customers alike
- Better Information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

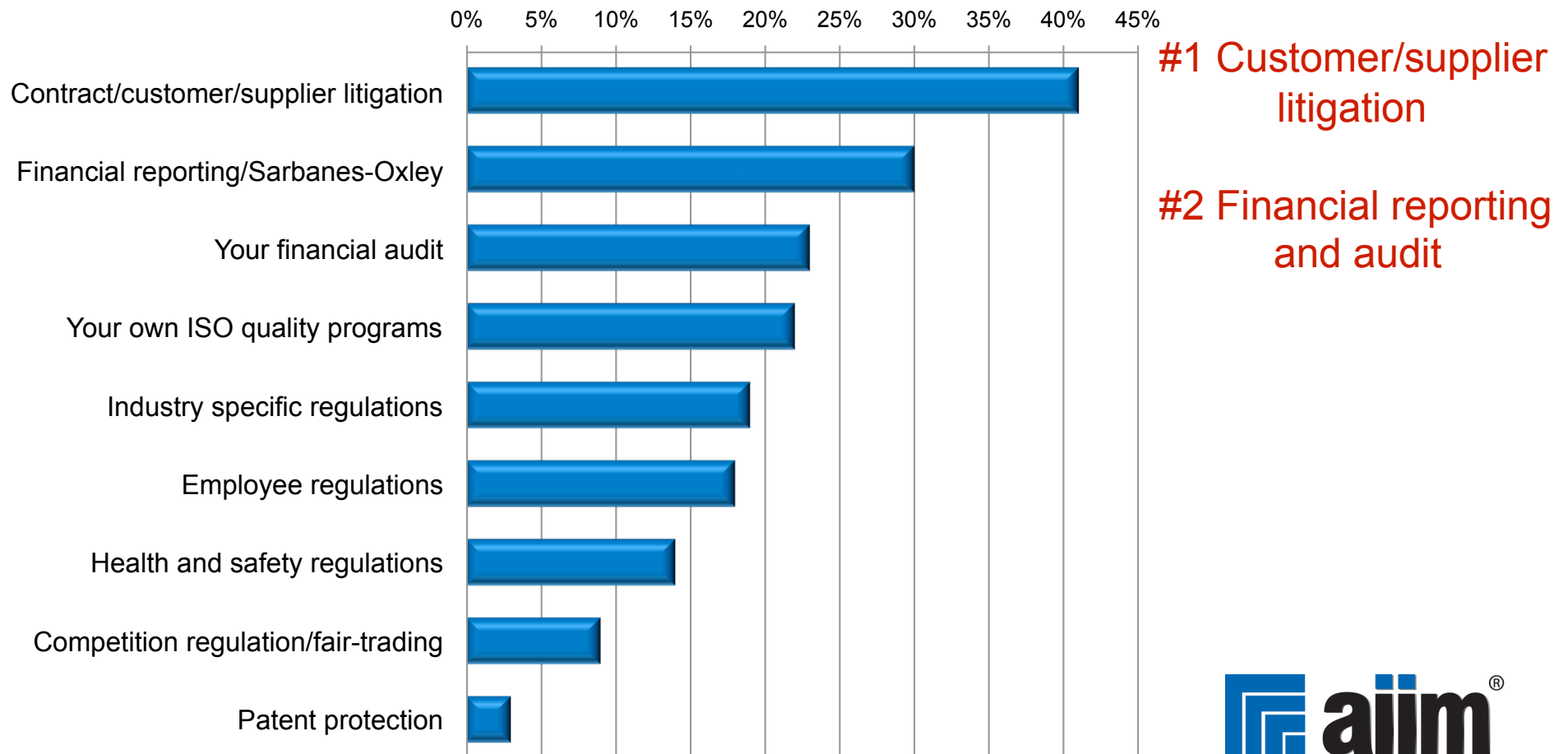


Efficiency and
business process:
45%

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?



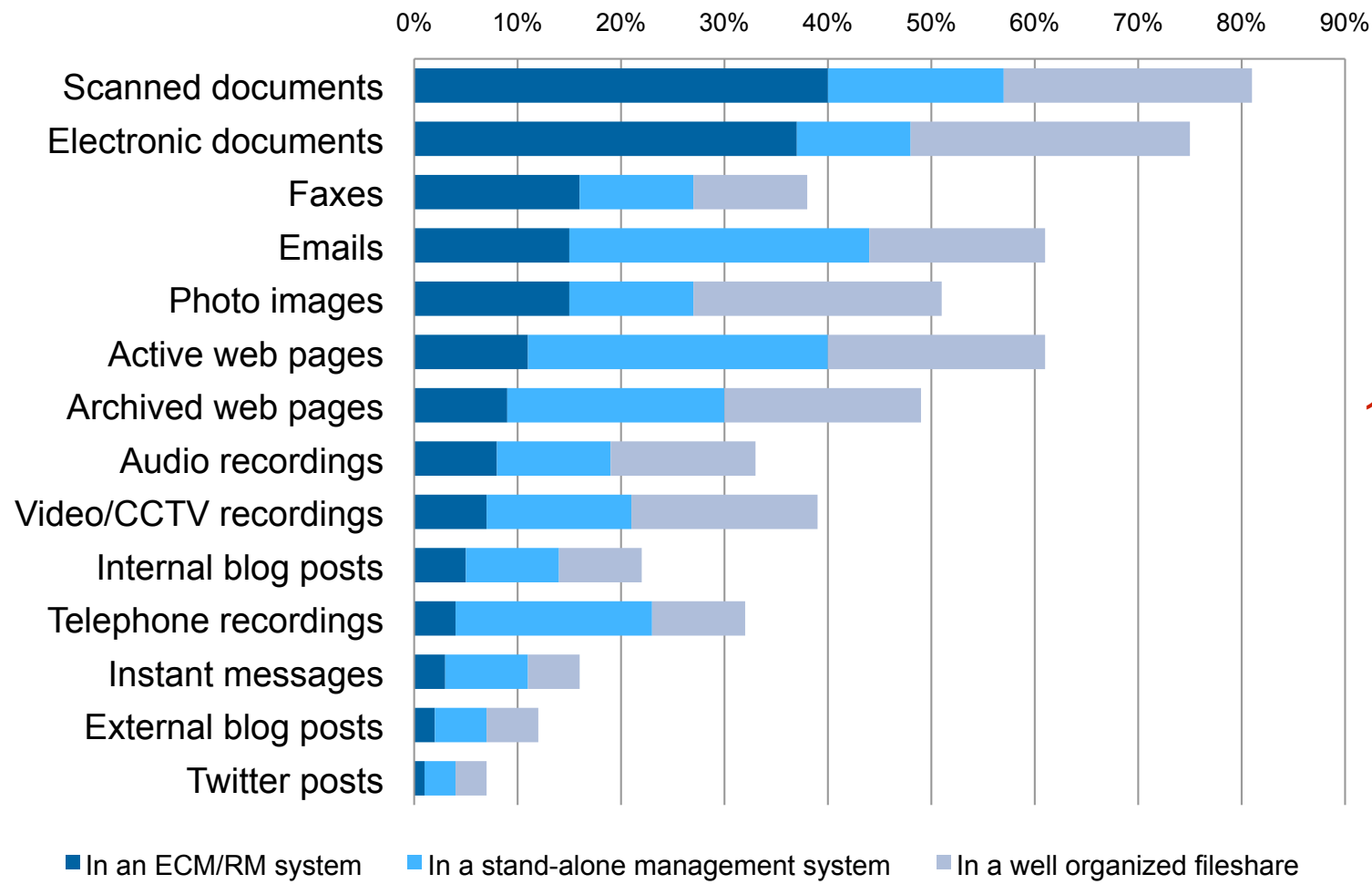
N=680 Non-trade,

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?

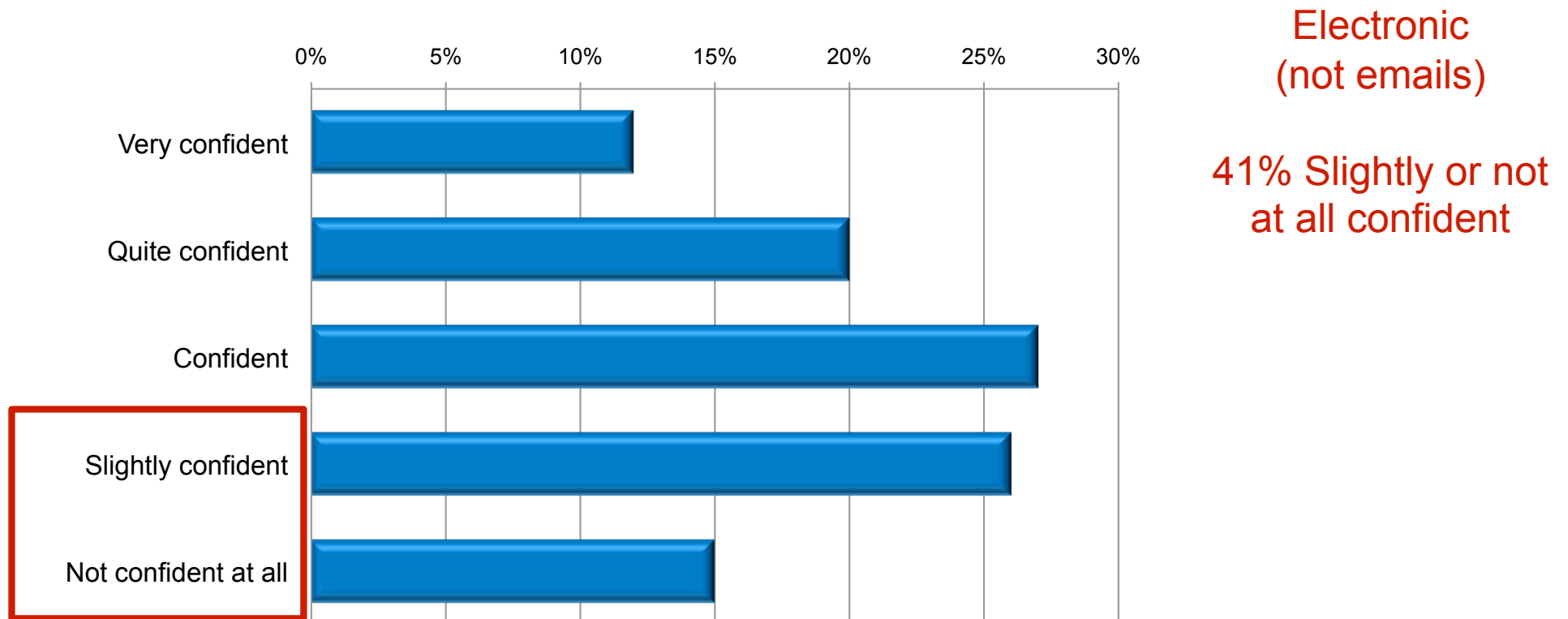


40% with documents in ECM/DM/RM system (scanned and electronic)

15% storing emails in ECM/DM/RM, 29% in EMM system

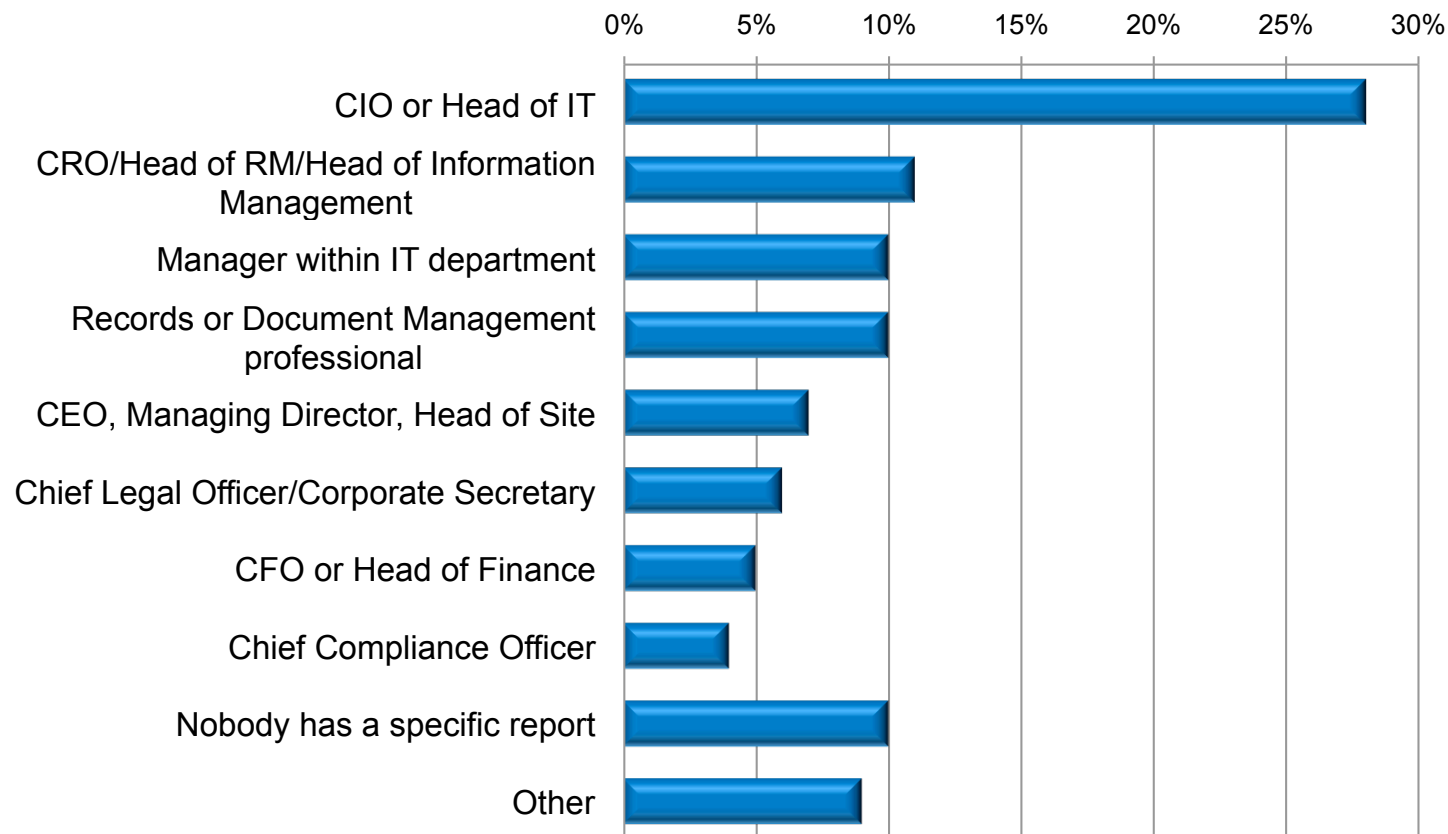
ECM Drivers - Electronic

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



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

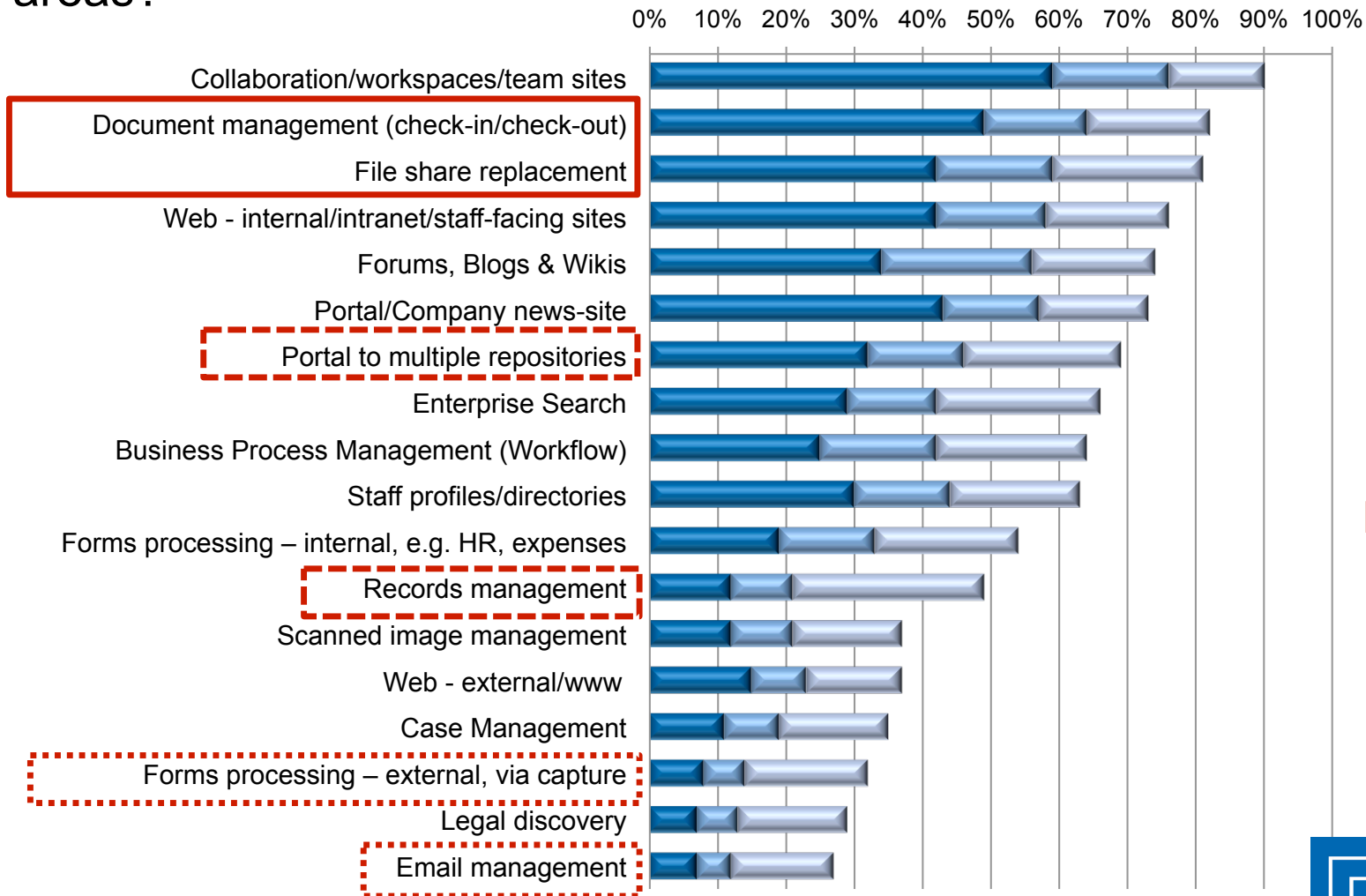
Plus 11% with a CRO

39% have no board-level ownership

N=645 non-trade

SharePoint - Use

How would you describe your use of SharePoint in the following ECM areas?



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

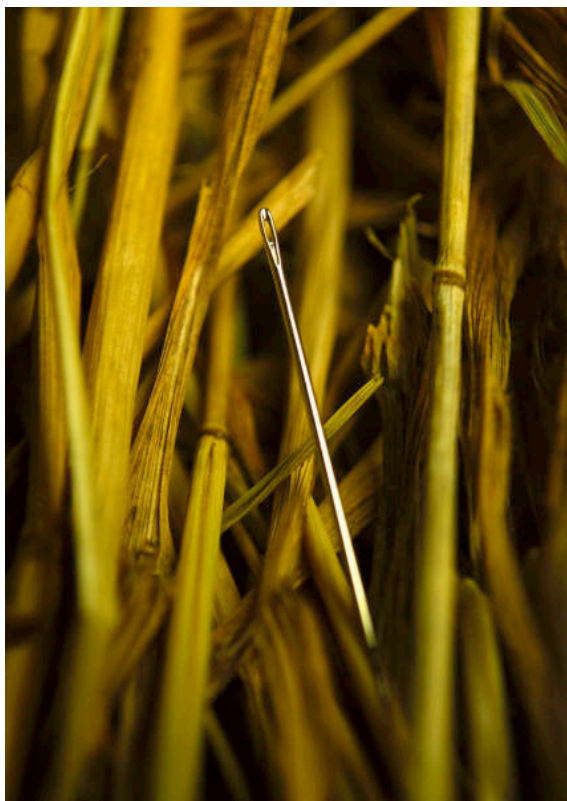
N=436 SharePoint using or planning
May 2010

■ Working live ■ Rolling-out ■ Planned in next 12-18 months

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?



Context to the Problem

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

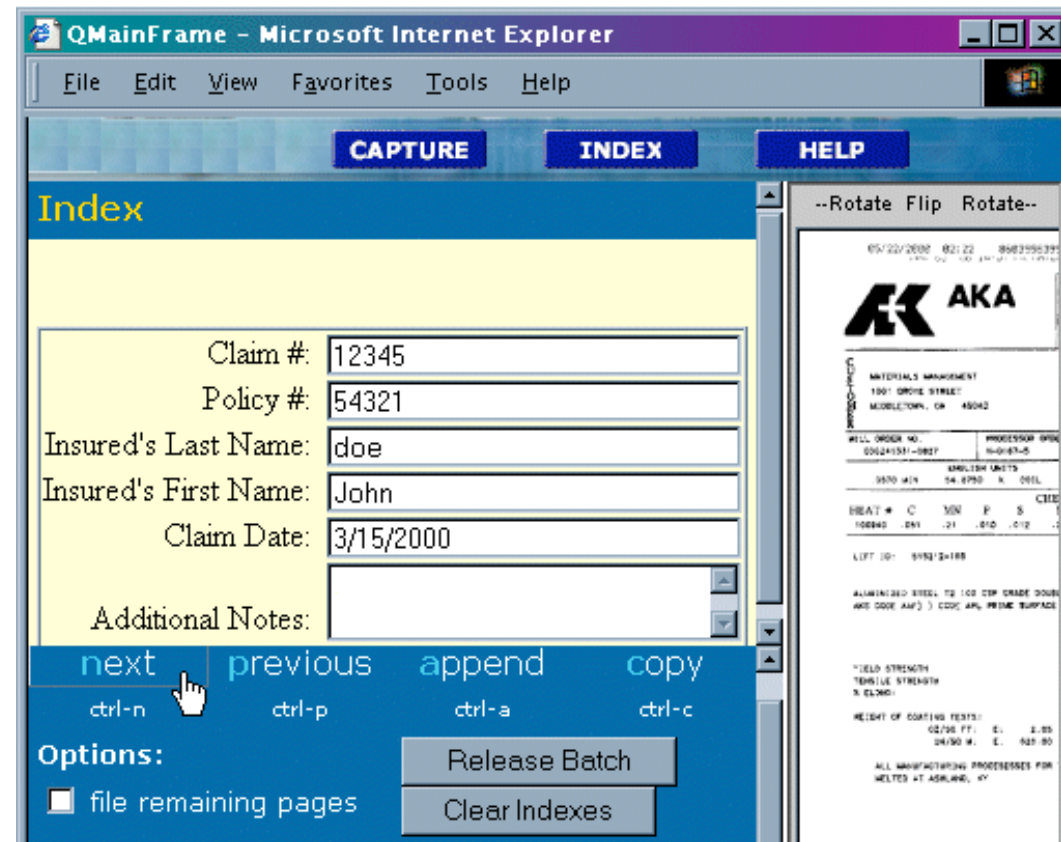
So a question arises:

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

Source: Taxonomy Strategies

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.



Source: QMainFrame

Capturing Metadata

The screenshot shows a dialog box titled "Booking Form 1 Properties" with a close button (X) in the top right corner. The dialog has five tabs: "General", "Summary", "Statistics", "Contents", and "Custom". The "Custom" tab is selected and active. It contains the following fields and controls:

- Name:** A text input field is empty. To its right are "Add" and "Delete" buttons.
- Type:** A dropdown menu is set to "Date".
- Value:** A text input field is empty. To its right is a checkbox labeled "Link to content", which is currently unchecked.
- Properties:** A table listing existing metadata.

Name	Value	Type
Owner	AIIM	Text
Purpose	Register fo...	Text
Language	English	Text
Date co...	3/1/2008	Date

At the bottom of the dialog are "OK" and "Cancel" buttons.

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

Document information dictionary		XMP	
Entry	PDF type	Property	XMP type
Title	text string	dc:title	Lang Alt
Author	text string	dc:creator	seq ProperName
Subject	text string	dc:description["x- default"]	bag Text
Keywords	text string	pdf:Keywords	Text
Creator	text string	xmp:CreatorTool	AgentName
Producer	text string	pdf:Producer	AgentName
CreationDate	date	xmp:CreateDate	Date
ModDate	date	xmp:ModifyDate	Date

Metadata Nomenclature

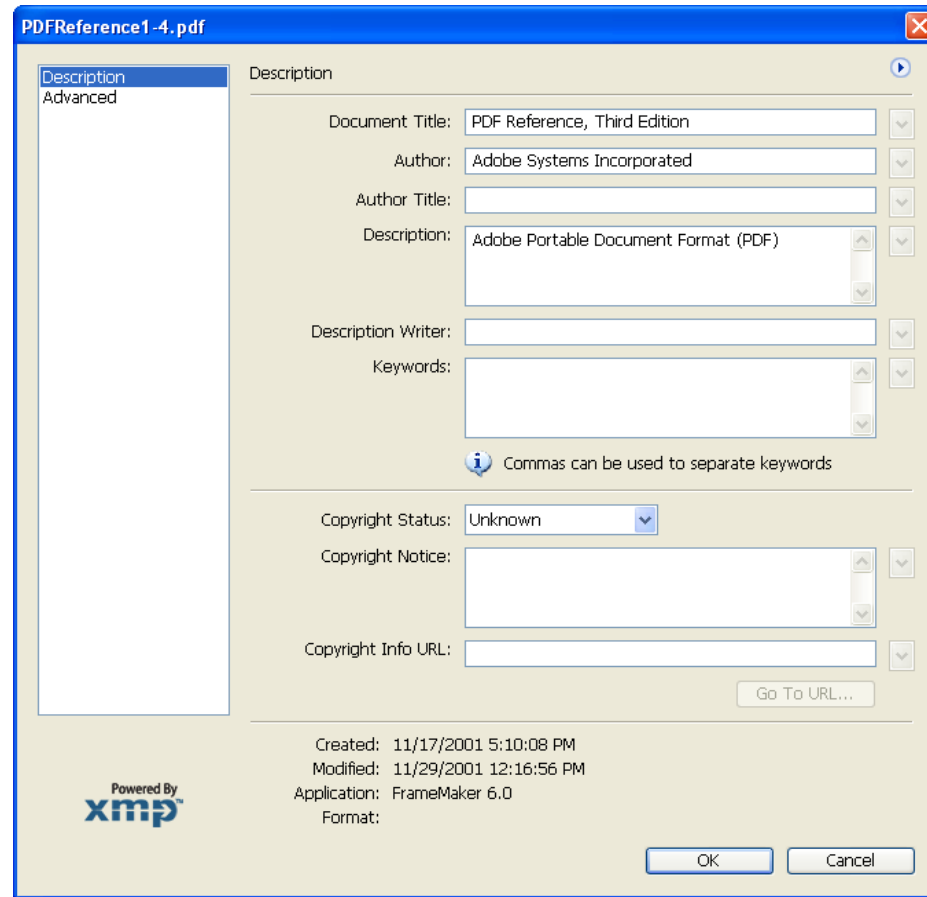
The following are examples of metadata tags:

```
<dc:element>Content</dc:element>
```

```
<pdf:element>Content</pdf:element>
```

```
<xmp:element>Content</xmp:element>
```

Example of Description Metadata



The screenshot shows a dialog box titled "PDFReference1-4.pdf" with a "Description" tab selected. The dialog contains several input fields for metadata:

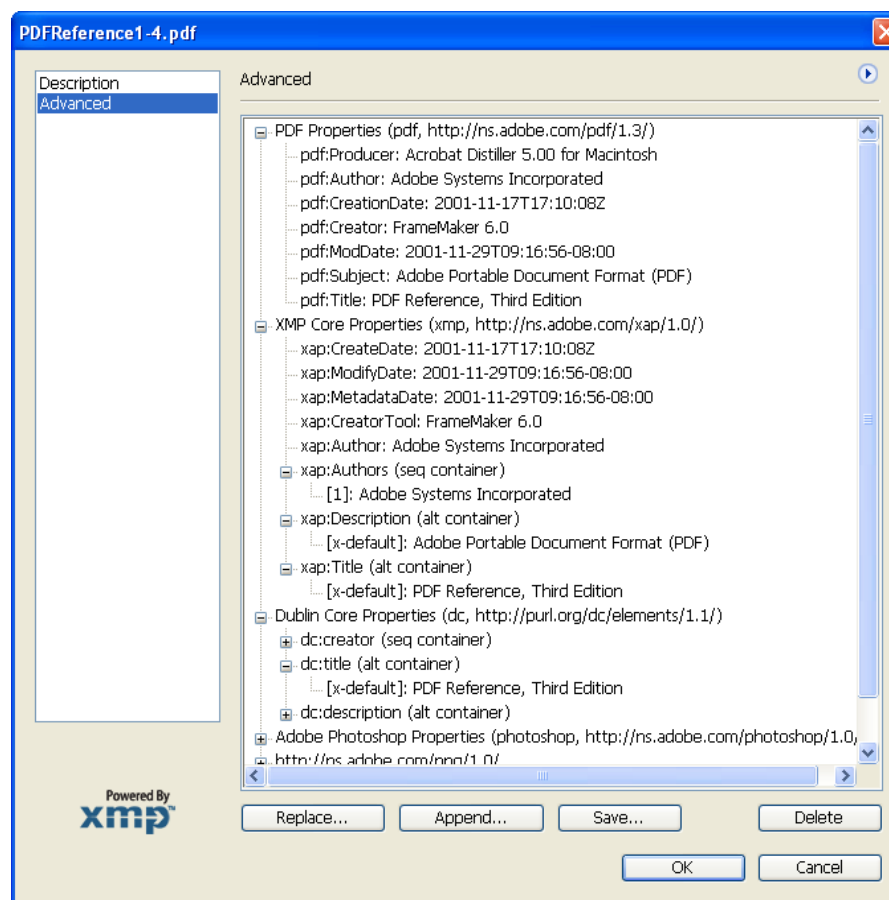
- Description:** A list on the left shows "Description" and "Advanced".
- Document Title:** PDF Reference, Third Edition
- Author:** Adobe Systems Incorporated
- Author Title:** (empty)
- Description:** Adobe Portable Document Format (PDF)
- Description Writer:** (empty)
- Keywords:** (empty)
- Copyright Status:** Unknown
- Copyright Notice:** (empty)
- Copyright Info URL:** (empty)

Additional information at the bottom of the dialog includes:

- Created: 11/17/2001 5:10:08 PM
- Modified: 11/29/2001 12:16:56 PM
- Application: FrameMaker 6.0
- Format: (empty)

The dialog is powered by xmp and includes "OK" and "Cancel" buttons.

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