2005

A Strategy Development Process for Enterprise Content Management

Ramon O'Callaghan
University of Tilburg, r.ocallaghan@uvt.nl

Martin Smits
University of Tilburg, m.t.smits@uvt.nl

A STRATEGY DEVELOPMENT PROCESS FOR ENTERPRISE CONTENT MANAGEMENT

Ramon O'Callaghan and Martin Smits. Center for Research on Information Systems and Management (CRISM), Tilburg University, PO box 90153, The Netherlands. E-mails: r.ocallaghan@uvt.nl, m.t.smits@uvt.nl

Abstract

Today, many organizations maintain a variety of systems and databases in a complex ad-hoc architecture that does not seem to fulfill the needs for company-wide unstructured information management in business processes, business functions, and the extended enterprise. We describe a framework to implement Enterprise Content Management (ECM) in order to address this problem. ECM refers to the technologies, tools, and methods used to capture, manage, store, preserve, and deliver content (e.g. documents, graphics, drawings, web pages) across an enterprise. The framework helps to select content objects that can be brought under ECM to create business value and guide the IT investments needed to realize ECM. The framework was tested in a large high tech organization.

Keywords: content management; document management; content strategy; ECM.

1. INTRODUCTION

Most organizations today generate information at such a rate that the challenge is putting this information in a format and in a place and where it can be found again, when needed. The Gartner Group estimates that most of the data (75%-80%) in organizations is unstructured and not in such a format that it can be found, when needed. Today, many organizations maintain a variety of systems and databases in a complex ad-hoc architecture that does not seem to fulfil the needs for company-wide unstructured information in business processes, business functions, and the extended enterprise. ERP systems and other systems do not exchange workflow information easily, and as a result a lot of the unstructured data is re-keyed manually leading to errors, inaccuracies and duplications [Oesterle, 2000; Weill et al, 2002].

A key challenge for organizations is to decide which unstructured data should be put under some kind of management control and which part can be left unstructured and unmanaged [Gupta et al, 2001]. The problem really has two facets: business issues, and technology issues. The business issues involve the analysis of the information and content needs in different organizational units and deciding on which content will be managed and how it will be managed. The technology issues involve the analysis of the existing systems providing content management solutions and their potential interconnection in an overall architecture to satisfactorily address the content management needs. This paper tackles the above problems from a business and a technology perspective and provides a framework to develop content management strategies and the underlying architectures.

We first define ECM (section 2), develop our framework and methodology (section 3), test the framework and methodology in a large high tech manufacturing organization (section 4), and provide conclusions and directions for further work (sections 5 and 6).

2. DEFINING ENTERPRISE CONTENT MANAGEMENT

Enterprise Content Management (ECM) has emerged as the convergence of two earlier technologies or solutions for managing unstructured information in organisations: Document Management (DM), and Content Management (CM). These terms and their differences are explained below.

2.1 DOCUMENT MANAGEMENT

Document Management (DM) is well established and has enjoyed considerable success in the 1980's and 1990's (Wei et al, 2002). DM systems aim to organize and make files accessible. DM systems have the following parts:

- File storage: The system knows the physical location of each file that it tracks but doesn't require the end user to know that location.
- File categorization: The user can assign file types and groups based on the criteria that he chooses.
- Metadata services: The user can attach any kind of extra data to a file (such as owner, status, create date, and so on) based on its type.
- Collaboration services: The user can check files in and out of the system and jointly edit them.
- Workflow services: The user can route files from worker to worker in an organized way.
- Versioning services: The user save a historical series of files and can retrieve them when required.
- Access services: The user can find files through tables of contents, indexes, and full-text searches.

DM systems do not necessarily deal with "documents". Although users often target systems toward word-processing and other office files, DM systems have no restrictions about the type of files that one puts in them. Thus, they should more accurately be called *file-management systems* (Boiko 2002).

2.2 CONTENT MANAGEMENT

Content management (CM) was originally associated with Web-development projects. As the Web moved past small, informally designed sites and into large, rapidly changing sites, the need for strong management tools became pressing. Systems vendors addressed this need and began offering *content management systems* (Hackos, 2002). Websites, however, are simply one of many outlets. And, as the amount of outlets and the information shared between them grow (e.g. multiple Web sites, print publications and anything else that can be created from the same information, especially in e-business networks), the need for a general approach becomes essential. CM is an overall process for collecting, managing, and publishing content to any outlet.

CM systems and DM systems have a lot in common. Both systems seek to categorize information, apply metadata to it, organize its creation through workflow and collaboration, and give end users complete access to it. On the other hand, the two systems differ in the following significant ways (Boiko 2002):

- *DM systems deal with files,* while CM systems deal with content components. Files don't make good containers for content. They can hold only one component per file if their metadata is to prove at all useful for finding individual components.
- *DM systems were invented to manage files that other applications create*. These systems make no attempt to open any of the files under their control and work with what's inside the files. CM systems, on the other hand, are directly concerned with creating content as well as managing it. In both systems, the user controls the creation of metadata for storage and access, but only in the CM system does the user control the creation of the content as well.
- *DM systems are to provide access to the files under its control*. The purpose of a CM system is to create publications that are a combination of the components under its control. To create publications, CM systems need all the access power of DM systems plus an extra capability to automatically construct a publication out of the components that it finds.

The focus of CM on content components represents an important paradigm shift: content re-use. Rather than write entire documents, authors create elements ("content objects") that can be assembled in different "information products" (e.g. a brochure, a press release, a presentation), for a number of different "delivery methods" and audiences ("target users"). There are multiple advantages to content re-use. Costs can be saved by writing once and using the information in two places, eliminating not only duplicate writing tasks but also reducing the amount of time needed for reviews. Once a module is written, it can be reviewed and tested and readied for use in multiple documents. The cost reduction associated with reuse is augmented by the ability to ensure that information is correct and complete

and appropriate to the needs of the users. For example, Dell Computer Corporation is able to create manuals for a wide variety of related products by creating inter-changeable modules. Variations in the information needed to support individual products can be tagged and selected. In this way, the author of a manual for a desktop computer uses the same core modules as the author of a manual for a laptop computer. Where the information differs, the authors write unique modules or modify existing modules by labelling the details that are different (Terra 2003).

2.3 ENTERPRISE CONTENT MANAGEMENT

As discussed above, the concept of ECM encompasses both Content Management (CM) and Document Management (DM). ECM capabilities manage traditional content types (images, office documents, graphics, drawings, and print streams) as well as the new electronic objects (Web pages and content, email, video, and rich media assets) throughout the lifecycle of that content. The definition of ECM has expanded beyond core library services to include imaging and workflow as well as records management, enterprise report management/computer output to laser disc, collaboration, and Web content management. ECM has become the umbrella term for a technology category for managing unstructured content. The term ECM today refers to the technologies, tools, and methods used to capture, manage, store, preserve, and deliver content (e.g. documents, graphics, drawings, web pages) across an enterprise. At the most basic level, ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists.

ECM involves the following tasks related to the development of content from its creation (from sources) towards use (in one or more audiences), represented in Figure 1 (Gupta, 2002):

- Create or acquire information from an existing source.
- Structure the information in the system by editing it, segmenting it into chunks (or components), and adding appropriate metadata.
- Create a repository that consists of database records and/or files containing content components, documents, and metadata.
- Extract components out of the repository and produce targeted publications such as Web sites, printable documents, and e-mail newsletters.
- Make the content available, i.e. easily accessible through a search engine or push it to key users

One way to enable ECM is to create a central repository that unifies content structure and content access (search, retrieve, version, index, archive). This includes the definition of content structure standards (types, metadata) and content organization standards (taxonomy, indexes, cross-references). To function as a central repository, the ECM system must extend the following qualities over all

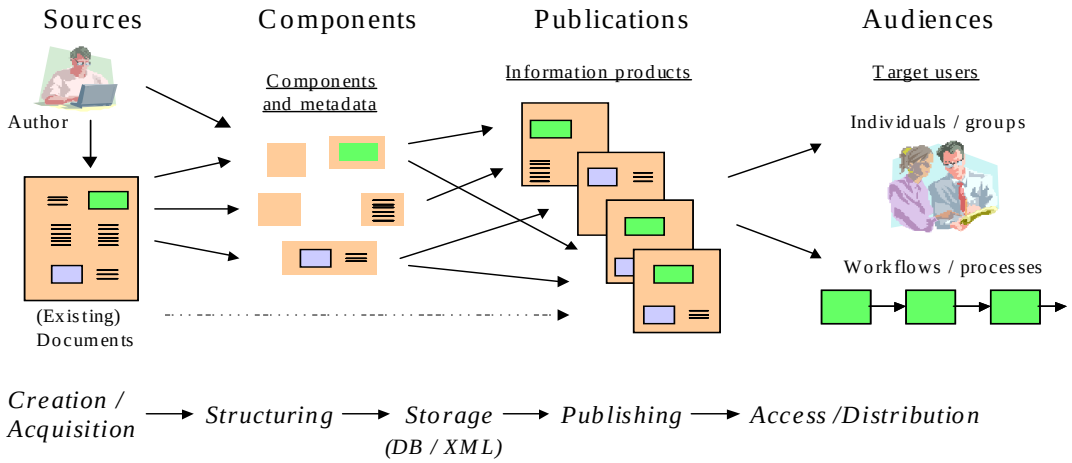

Figure 1. Stages in the process of content creation. (Based on Gupta, 2002)

content (Rockley 2003):

- Unified content structure. The way content is chunked and tagged (components and their elements) needs to be standardized across all data sources.
- Unified organization. The hierarchies and other organizational schemes you use to categorize and get to your content need to extend to any place where the content is stored.
- Unified access. The way you query and make use of the content you retrieve needs to be the same across all data sources.

The management of content is based on metadata. Metadata is information about the data: e.g. the instructions that come with the data. Metadata exists in addition to or after the data. It adds context and a wider interpretation to the data. The metadata isn't the content. It exists apart from the content (Everett et al, 2002). Metadata is also a set of standards that groups agree to for information definitions. Standards, which are the basis of any kind of data sharing, bring the possibility of large-scale efficiencies in information interchange among groups that don't even know one another. In the content management context, the standards may be mostly internal today, but they serve the same purpose. Standards ensure that others can automatically reuse the efforts of one person or group if they all follow the same standards (Stuckenschmidt and Van Harmelen, 2004).

To manage content, a choice must be made on what information must be separated out and made into metadata and what remains part of the content. This distinction is a practical, not a philosophical one. Managing content is managing metadata. Metadata provides the capability to share data across applications. In a content management context, metadata enables publications that need a somewhat different form of the same data to draw from a common repository (Wei et al, 2002).

We developed a 'framework for ECM strategy' to support this selection process and to identify content that is worth managing on an enterprise wide scale.

3. FRAMEWORK FOR ECM STRATEGY

An ECM strategy aims at reducing the costs of creating, managing, and distributing content, and ensuring that content effectively supports organizational needs. An ECM strategy is a method to identify content requirements, creating consistently content for reuse, managing that content in a definitive source, and assembling content on demand to meet organizational and customers needs. The ECM strategy development process starts by analysing existing needs and how these are being met. In other words: who needs what information, how information supports users, and how it is produced, i.e. what processes and technology are currently used to create content.

The ECM strategy includes three components: a content management (CM) system, reusable content, and collaborative CM processes.

- CM system: An ECM strategy requires a robust CM solution that manages content in a single source. Most CM systems provide traditional document management functionality, such as secure access to content (check-in/ check-out), revision control, reporting, powerful search and retrieval mechanisms, and metadata. However, ECM is not just about technology, it is about the interaction of business with content, people, processes and tools. Authors need CM to assist them in authoring. They may need help to find and distribute or publish content, and to ensure that the content they are distributing is accurate and appropriate.
- Reusable content: Content reuse means writing once and reusing it many times. Traditional documents are written in files that consists of sections. Reusable content is written as objects or elements, not documents. Documents are made up of content objects that can be mixed and matched to meet specific information needs. Reusable content is broken down into smallest reusable object (section, paragraph, sentence). When information is broken down at this level it is easy to select an element to reuse or repurpose it. However, even though content elements are re-used, copying and pasting is eliminated. Instead, elements are stored in the database or CM system and are referenced (pointed to) for inclusion in a virtual document. In this way, the element can appear in multiple places, but reside in only one.

- Processes: An ECM strategy also involves people and collaborative processes. The processes must create a collaborative environment in which authors share in the development of content to create a single definitive source of information. The ultimate goal in defining unified processes is to ensure that all departments are aware of what content exists, all authors can reuse existing content, and all processes are repeatable and transparent, regardless of which department and which authors are following them.

Figure 2 shows our framework for developing an ECM strategy. The following sections describe the framework, the methodology, and the steps involved.

3.1 CONTENT AUDIT

Determining ECM needs starts with a *content audit*: an accounting of the information in the organization. The purpose of a content audit is to analyse how content is used, reused, and delivered to its various audiences. The audit is also used to understand how information (and the processes that create it) can be unified, eliminating the "cut and paste" method many authors use. The following questions are pertinent:

- How much content is available for a given organisational unit? How fast is it growing?
- How many types of information content are there?
- Who manages which content? Who owns the content? Who uses the content?
- How does content get re-used and re-purposed?
- What content must be saved, in what form, and for how long?
- What solutions / systems / methods are currently being used?

After the content audit has been performed, the organization has an overview of all the content and how it is used. Such analysis reveals those content objects that are potentially more relevant from the point of view of ECM. These will typically be the more frequently used, or those used in important management decisions or key business processes. Since the process of assessing whether it is worth (or not) to put a given content object under ECM requires quite some effort, this preliminary step has a very practical purpose: to reduce the number of content objects that will have to be scrutinized. The point is that the audit may reveal thousands of documents or content objects in total for a given business unit or process, but only a fraction of those should be really analysed in depth.

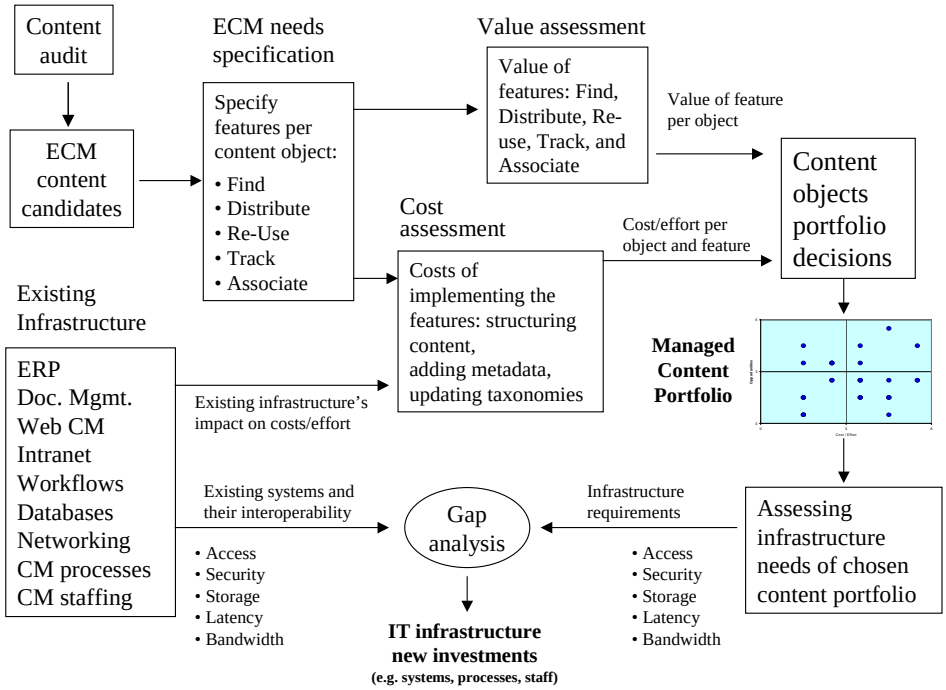

Figure 2. Overview of the ECM strategy framework (see text).

3.2 SPECIFYING ECM NEEDS

As discussed above, ECM is about facilitating the creation, storage, revision, re-use, distribution, search and retrieval of content. Thus, for the purposes of this research, we have conceptualised ECM in terms of five features: Find, Distribute, Re-use, Track and Associate. We call these "basic ECM features". The idea in this step is to review each content object in order to discuss what it means to make the object "findable", "distributable", "reusable", "traceable", and "associable". Focus group sessions with key informants (people knowledgeable about the content of a given organizational unit), should thus specify the desirable features and their level for each content object.

One of the key decisions concerning reuse is the level of "granularity". Granularity determines the smallest piece of information that is reusable. For an existing document, re-use may imply breaking it into smaller pieces. The level of granularity can change throughout the content. In one instance, large sections may be reused unchanged; in others, one may reuse content at the sentence or even the word level. Different levels of granularity can be used for authoring, for reuse, and for delivery.

3.3 VALUE ASSESSMENT

The next step is to assess the value of putting the content objects candidates under ECM. This is done object by object, by analysing the potential value of each feature separately. In other words, what is the value of making a given object "findable", "distributable", "reusable", "traceable", and "associable"? In practice this value assessment can be done by surveying a number of potential users (representative of the key audiences) or by convening a focus group with key informants (typically managers familiar with the information used in the main business processes), and asking them to rate the perceived value of each feature, i.e. give a score on a scale of 0 to 10. Based on our preliminary experience, we have developed some guidelines that can be used to initiate the value discussion in the focus groups:

- *Find*: The value of making a given object "easy to find" is a function of number of factors such as: the number of users or potential users (audience size), the importance of the tasks where it will be used, the criticality of the problem being addressed, and the status of the author/provider of the information.
- *Distribute*: The value of making a given object "easy to distribute" depends on: the defined target audience that "needs to know", the importance of the content object for their task, or the importance of the object as informational input to a given business process.
- *Re-use*: The value of making a content object "re-usable" is a function of the diversity of authors, information products, delivery methods, users, and processes in which a given content object may be re-used.
- *Track*: The value of making an object "easy to track" relates to the size of the potential audience, role of the user, criticality of the process/ task, and need for compliance (e.g. Sarbanes Oxley legislation in the US).
- *Associate*: The value of "associating" a given content object with other content refers to search situations in which the user does not know exactly what he / she is looking for ("fuzzy requests"). The value will thus depend on the frequency / need for such "fuzzy" inquiries. These tend to be more common in decision-making audiences, in collaborative networks, and in dealing with information from multiple sources.

A short questionnaire is used for each of these attributes. These questions can be used both in focus group discussion and for a survey of a large number of content users.

3.4 COST/ EFFORT ASSESSMENT

The cost of putting a content object under ECM is directly related to the effort and complexities associated with the way the above features (find, distribute, reuse, etc.) are implementation. Here, we should note that these costs are dependent on the existing infrastructure and on how content is currently being handled. For example, it makes a difference whether a document resides in the hard disk of a desktop computer, or whether it is already being handled by an enterprise system or a document management system. Thus, we should emphasize that both value and cost should be

assessed on a relative basis, i.e. with respect to the existing situation (which depends on the existing tools, methods, technology infrastructure and information architecture).

As in the case of the value assessment, it is useful to assess the five ECM attributes. Unlike, the value assessment, the cost assessment should be conducted by *information management specialists* in combination with key informants from the business units or processes. For the most part, the cost will be associated to adding and managing the metadata that will allow content to be managed in an effective manner. These steps concerning the implementation of the ECM features influence cost:

- *Find*: Making a content object "findable" implies adding content related metadata. To do that, one must take into account the terms that may be used by authors and the terms that may be used by users to classify topics and look-up. One way to implement this is to consider a controlled list of keywords, associate them to the relevant business processes, and put them in a taxonomy. Examples of retrieval metadata would be: Title/ subject, Author, Date (creation, modification), Keywords, Security level (who can view the content), abstract, related assets such as projects / process, places in the taxonomy.
- *Distribute*: Making a content object "distributable" implies adding user profile metadata for targeting (pushing content) to different users or processes. It also requires developing and managing a user-oriented taxonomy, and grouping (clustering) content for a given audience. It is worth point out the differences between Find and Distribute, and the implications for the respective metadada: In "Find", users find the content, e.g. when a user queries a content database and retrieves the selected content. In "Distribute", content "finds" the users, e.g. when the system queries a persons database and sends out a message to the selected people based on their profile and targeting criteria.
- *Re-use*: Making content "re-usable" implies breaking it into smaller documents / components ("bursting"), putting these in a content repository instead of formatted files, and adding metadata to each component for subsequent retrieval and use. As discussed above, the level of granularity is a key decision, and this one has major impact on costs and effort: The more granular the content, the greater the complexity of modelling, authoring and managing the content. Yet if content is not granular enough, one can compromise the ability to reuse information. Regardless of the level of granularity, however, authors still write complete documents, not elements. Authors write documents and assign the required granularity to elements (as defined in the information model) as they write. The main difference for authors is in following the assigned structure and in assigning or selecting metadata. In essence, the granularity defines how the completed document is broken down, tagged, and stored for reuse; but it does not define the authoring processes.
- *Track*: Making a content object "easy to track" implies adding metadata for review and approval workflow as well as content tracking. Status metadata, for example, could specify whether a document is: in draft version, ready for review, in review, final, in approval, approved. Content tracking would add version control metadata such as: Who created the content (author), when it was created/ modified (date), who reviewed / approved (reviewer/ approver), how long it took to create/modify/review (time), where it has been reused (channel, information product). In addition, when content has been changed, (re-)users may want to be notified through, for example, through alert messages.
- *Associate*: Making a content object "easy to associate" implies finding related content automatically. This can be done by adding links to other content objects, and by grouping or clustering related content. One can also use taxonomies to find related content. Whereas in "Find", people find content, and in "Distribute" content find people; here, in "Associate", content finds content.

3.5 CONTENT PORTFOLIO DECISIONS

The outcome of the preceding steps (i.e. value assessment and effort assessment) is a series of scores for every content object and the five ECM features for each object. The users surveyed will have rated (e.g. on a scale from 0 to 10) their perceptions concerning the relative value of implementing the ECM

features for all content objects under consideration. Similarly, the information specialist will have rated the perception of the marginal effort/cost associated with the implementation of the ECM features for all the content object candidates. The total number of data points is thus five times the number of objects. The next step is to plot these data points using a scatter diagram as shown in figure 3, which illustrates a simple example with 6 content objects and 30 data points.

In this diagram each dot represents a decision: whether or not to implement an ECM feature for a given content object. The position of the dot in the diagram reveals its relative attractiveness. For example, dots that involve high value and low effort (i.e. in the top left quadrant labelled DO) are clear candidates for implementation. Those in the bottom right (in the DON'T quadrant) require a lot of effort and the perceived value is low. These should not be implemented. Dots in the upper right quadrant (CONSIDER) represent decisions that would involve a lot of effort, but potentially a high value. These should be carefully analysed as the effort may well be worth it (high risk, high gain decisions). In principle, the diagonal line separates the DOs from the DON'Ts. The lower left quadrant includes the low cost, low value points. Here the decision can go either way because the costs involved are relatively low. In practice, the final decision may be influenced by the potential joint value of having several features or objects implemented together as a set even though initially they have been assessed independently. The outcome of this step is a *managed content portfolio* reflecting the decisions on all the candidate content objects, i.e. which objects will be put under ECM and which features will be implemented.

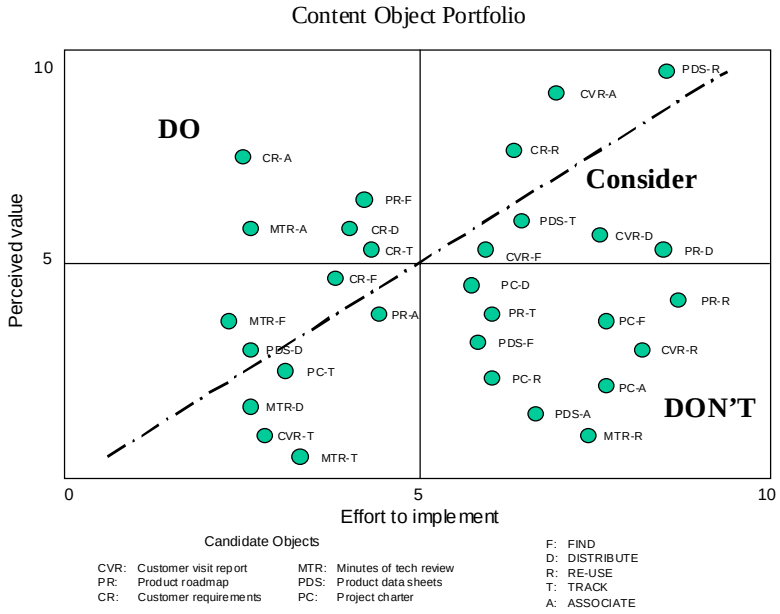

Figure 3. An example of the Managed Content Portfolio

4. APPLYING THE FRAMEWORK

The methodology was tested by applying the steps described in section 3 to a large manufacturing firm in the high-tech sector. The company has about 30,000 employees and has four key business processes and a large IT infrastructure consisting of a single kernel ERP application with several dozens (legacy) applications. The scope of the exercise was limited to the two business processes that had expressed a greater need for ECM: the product development process (PD process) and the marketing strategy process (MS process).

4.1 INTRODUCTION TO THE CASE

The PD process covers all business activities ranging from determining the specification of the products needed by the customers (world-wide) leading to the manufacturing of the products in a handful high tech production plants and the delivery of the finished goods. The MS process covers all

business activities regarding market development, product strategy development, and customer relationship management. Over the past decade, both processes have developed into a well defined set of activities, including well defined procedures for reporting and information exchanges, to enable strict process control. However, most control activities focus on one process, while coordination across the (four) business processes is still in its infant stages. Much information (and content) is created and exchanged within each process.

Various actors, including the CIO and the supply chain manager have recognized that information exchange between processes should now be improved. In 2004, the company decided to improve the linkages between processes by starting an enterprise content management program. Key objective of the program is to provide access to valuable content, available in the many web sites (intranet, extranet, internet), the many group pages, and the overwhelming amount of Powerpoint (and other Office) documents.

Based on about thirty interviews and the numbers of employees, applications, and business activities, we estimated that the PD process creates an enormous amount of informal information products (besides its well known 40 formal documents, including for instance product data sheets, typically consisting of up to 1400 pages each). Informal documents include e.g., news and brochures for about 250,000 people, 15 million intranet pages created by 500 people (which is about 60% of all Intranet pages in the company). Data to create reports, pages and Office documents come from many different groups and systems: ‘Getting a business plan from a business unit is a nightmare’ and ‘putting content in the Autonomy application is not possible because Autonomy is badly implemented’. Formal ‘golden documents’ are stored in well developed planning systems taking about 20 gigabytes per year (project milestone dossiers are 40mb each; 20 projects per business line (25) = 500 projects/yr). On the other hand, informal project content covers 20 Terabytes per year (up to 80Gb per project).

4.2 TESTING THE METHODOLOGY

The initial content audit involved interviews with key managers and information specialists in each domain. We call these informants “domain experts”, as they were knowledgeable about the business processes and the information created and used in these processes. These interviews allowed us to gain insight in their business processes, the roles involved (as authors and users), the information needs, the information being generated and used by the different activities. Together with the managers responsible for these processes and also with the help of some of the domain experts, were able to identify a short list of key content objects. These would be subjected to an in-depth assessment.

Value assessment was conducted in a focus group meeting with key representatives from the two business processes mentioned above: marketing planning and business development. The session was used to discuss what the implementation of ECM would imply for the selected content objects, and what the value of each feature (find, distribute, reuse, track, associate) would be for each object. After agreeing on the specification of the need and the value assessment, the group converged and rated the features per object using a score (0-10 scale) as described above.

The effort assessment was conducted with the additional help of IT staff and the chief IT architect. They used their knowledge of the existing CM tools and the new ECM solution that could be contemplated in order to assess the potential difficulties in the implementation of the content objects in question. As described earlier, the ratings reflected the relative advantage and the relative effort (i.e. additional gain and efforts with respect to the solutions already in place).

In the end, some 20 objects and 100 data points were assessed, and a scatter diagram with the 100 decision points for consideration. Following the guidelines described above, the diagram provided the basis for a discussion on what to do with every content object and ECM feature. A number of points were chosen. This lead to a preliminary managed content portfolio as a first step in their journey towards an ECM strategy. The focus of our analysis, however, is not on the characteristics of the resulting content portfolio (the outcome of the process) but rather on the assessment of the process itself. The next section discusses the conclusions reached in such assessment.

4.3 CONCLUSIONS

The outcome of the methodology outlined above is a *managed content object portfolio.* The portfolio reflects the choices regarding what content should be put under ECM, and as such it represents the Content Management Strategy. The tests conducted in the case have lead to the following conclusions:

- The method had conceptual and face validity. The majority of those involved understood and accepted the rationale and the concepts underlying the framework when it was explained to them.
- The usability of the method, however, was not straightforward. The team concluded that a briefing would be required if the method were to be applied in other parts of the organization.
- The relative value and the effort/cost of implementing the ECM features required discussion. A list with some questions and examples would help clarify the concepts and how to score them.
- Since the whole process takes a long time, it is better to split it in two sessions: one for agreeing on a short list of content objects, and another session for assessing their value/ costs.
- Further tests may need to be conducted by different groups for different clusters of content objects (different user communities) and/or a survey based on a structured questionnaire.

5. FURTHER CONSIDERATIONS

As discussed earlier, the value and cost assessment for developing the content portfolio has been done on a relative basis, i.e. taking into account the current situation, which is influenced by the existing IT tools, methods, and infrastructures. Indeed, the development of an ECM strategy in an organisation is not a "green field" exercise. On the contrary, it must take into account the existing infrastructure and assess which elements will enable or constraint ECM implementation. The ECM strategy process must, in the end, also identify elements of the IT infrastructure that need to be changed, extended or complemented with new systems / tools and how these will be integrated with the existing base.

5.1 INFRASTRUCTURE IMPLICATIONS

Although the focus of this paper is essentially on developing an approach for content management strategy (i.e. deciding on the ECM content portfolio), information managers will, in fact, care about the subsequent steps that an organization must undertake in the area of software tools and IT infrastructure. As the organisation moves towards implementation, a number of questions will emerge. For example: What does the chosen content portfolio require in terms of tools and systems? How does the existing IT infrastructure support these needs? What should be modified or added?

In addressing these questions, it is important to consider the perspectives of both content authors and users. In other words, one should consider the functionality needed to support authoring needs, on the one hand, and the desired vision of users' experience with information resources, on the other hand. This means that, in addition to the basic functionality of library services[1] and the additional functionality involved in bursting and reuse, the system must be robust enough to support the different user environments and growth scenarios.

The implementation of an ECM strategy should address issues such as: interoperability of information and document repositories with applications, consistency across processes to facilitate access and sharing of information, and modification of company's information systems architecture to accommodate ECM. In particular, the implementation of the ECM features discussed earlier (i.e. Find, Distribute, Reuse, Track, Associate), will impose different demands in terms of infrastructure needs. At a basic level, these needs can be articulated in terms of five technology domains: Access, Security,

[1] Typical library services include (i) a single point of access through a standard login, (ii) login security that ensures content is accessible only to those authorised users, (iii) check-in / check-out security to ensure that a document cannot be edited by more than one person at a time, (iv) version control that stores incremental changes to documents so that older versions are accessible if needed, (v) keywords that facilitate search and retrieval, (vi) optional workflow services that track a document trough a number of persons who are required to take some action such as review, edit or approval.

Storage, Latency, and Bandwidth. Some of these refer to the need of making content available so that it can be easily searched and found (e.g. via a URL) in a secure way (only for those authorized to see a given content object and excluding others). Others -like storage, latency and bandwidth- refer to physical properties of the network and storage capabilities which will affect network configuration and response time.

A distinction can be made between metadata and the content object as in many cases it is more important to access and manage the metadata than the object itself. For example, when a user looking for some content initiates a query, he/she will want to know quickly what object(s), if any, are available and proceed either to refine the search or download the found content. In this scenario, the speed at which the content (full text) is delivered is less important than the speed at which the metadata (e.g. document title, author, date, keywords, abstract) can be accessed and searched.

5.2 ECM GOVERNANCE

Having a powerful ECM system and/or IT infrastructure is useless if the content is inaccurate, outdated, or irrelevant for improving users' ability to perform. Organizations need to be disciplined about the publishing process and the management of taxonomies.

Taxonomies are necessary to tag the documents created. The tagging (adding metadata) is important to ensure that search engines will find the requested documents and for distribution based on personalization rules. Tagging can be a labour-intense process, and requires the cooperation of content authors. The capture of metadata should ideally occur right after content creation through a combination of automatic (author, name, date, etc) and manual processes (keywords, categories). (Terra and Gordon, 2003). But, the higher value metadata should be provided by the authors themselves and includes detailed description of the document (e.g., summary, context of creation, purpose, contributing, authors, etc.).

A related issue is that different people develop different taxonomies. This may require multiple ways to find the same information. But, who maintains the integrity of the taxonomies as they grow over time? Ongoing maintenance of the information architecture requires assigned responsibilities and not just the use of sophisticated tools. It should be noted here that the maintenance of a taxonomy is not something that is done automatically, it requires considerable human effort. This is typically the work that librarians have been doing for centuries. An alternative to taxonomies is full text search. In this case, keywords clusters can be generated automatically (e.g. through frequency count or cluster analysis). Research has shown, however, that automatically generated taxonomies perform worse than taxonomies managed by professional editors/librarians (Terra and Gordon, 2003). The choice of a taxonomy system, therefore, has implications beyond tools and infrastructure, as it affects procedures, roles and staffing.

Developing and keeping the information architecture requires dedicated personnel. For example, a taxonomy could start with some inputs developed jointly by content "owners" or domain experts and information architects. Non-expert users will need help to understand a specific domain, i.e. how content is organized and linked to other content. The implication is that organizations should develop a set of rules for ECM governance outlining roles and responsibilities, specially concerning the management of metadata and the set-up and maintenance of taxonomies. Although many ECM solutions provide automation features, organizations should not underestimate the critical role of editors, librarians, information architects, and designers.

Finally, it should be pointed out that the information architects, domain experts and business analysts interested in implementing a content-management solution should defer to the expertise of in-house information technology (IT) specialists. IT specialists also need to clarify their requirements and work closely with ECM vendors to ensure that their needs are well understood and supported by the system's capacity. At the same time, IT specialists should ensure that they understand the requirements of the authoring and publishing communities, and represent their needs effectively to the vendors.

6. CONCLUSIONS AND FURTHER RESEARCH

We defined CM and ECM and developed a framework and a method to select content objects to be put under ECM. Applying and testing the method and framework in a large firm shows that content objects can be listed per business (sub) process and then selected based on their added value and specific costs. Added value can be determined using five dimensions: 'easy to find', 'easy to distribute', 'easy to re-use', 'easy to track', and 'easy to associate'. Specific costs of bringing content objects under ECM can be determined using 'costs of structuring the content', 'costs of adding metadata', and 'costs of updating taxonomies'. Added value and specific costs per content object can be used to decide on a content portfolio to be put under ECM, resulting in optimal content values.

The costs of ECM depend, to some extent, on the existing IT architecture. Also, the ECM portfolio decision can be used to determine further investment planning in IT architecture. In this way, the ECM framework and methodology presented in this paper help to manage and bridge the complex relationships between IT architecture, business processes, information, and enterprise content. This is in line with the resource based view of the firm: investments are never made from scratch, but build on the existing technology, knowledge, and competency resources of the firm (Barney, 1991).

The framework and methodology need to be tested in more depth and in other organizations. We consider the work presented in this paper a preliminary step towards building a sound ECM strategy development methodology.

LITERATURE

Barney, J. (1991): "Firm resources and sustained competitive advantage". Journal of Management, Vol. 17, No. 1, pp. 99-120.

Boiko, B. (2002): "Content Management Bible". New York: Hungry Minds.

Davenport, T. and L. Prusak (1999): "Working Knowledge: How organizations Manage What They Know". Boston: Harvard Business School Press.

Davenport, T., et al. (2001): "Data to knowledge to results". California Management Review, 43 (Winter).

Everett J.O., Bobrow D.G., Stolle R., Crouch R., Paiva V. de, Condoravdi C., Berg M. van den, Polanyi L. (2002): Making ontologies work for resolving redundancies across documents. Communications of the ACM (45) 2: 55-60.

Gupta V.K., Govindarajan S., Johnson T. (2002): Overview of Content Management Approaches and Strategies. Electronic Markets (11) 4: 281-288.

Hackos, J. T. (2002): "Content Management for Dynamic Web Delivery". New York: John Wiley & Sons, Inc.

Oesterle H, Fleisch E, Alt R (2000): Business networking, shaping collaboration between enterprises. 2nd revised and extended edition. Springer, Berlin.

Rockley, A. (2003): "Managing Enterprise Content". Indianapolis: New Riders – Pearson Education.

Stuckenschmidt, H., Harmelen F. van (2004): Generating and managing metadata for web-based information. Knowledge based systems, (17) 5/6: 201-206.

Terra, José Claudio, and C. Gordon (2003), "Realizing the Promise of Corporate Portals: Leveraging Knowledge for Business Success". Burlington, MA: Butterworth-Heinemann (Elsevier Science).

Weill, P., Subramani, M., and Broadbent, M. (2002): "Building IT Infrastructure for Strategic Agility," Sloan Management Review (44) 1: 57-65.

Wei C.P., Hu P.J., Dong Y.X. (2002): Managing document categories in e-commerce environments: an evolution-based approach. European journal of information systems, (11) 208-222.