ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

A Review of Strategic Decision Making Tools: Towards an Efficient Enterprise Content Management System

Ajeet Ghodeswar[1], Sachin Sonawane[2], Mahendra Patil[3], Sandeep Chavan[4]

Assistant Professor, Dept., of Computer Engineering, Atharva College of Engineering, Mumbai, Maharashtra, India[1, 2]

Head of the Department, Department of Computer Engineering, Atharva College of Engineering, Mumbai, India[3]

Head of the Department, Department of Computer Engineering, Shree Datta Polytechnic College, Kolhapur, India[4]

Abstract: Driven by the need to control the content chaos that pervades local drives, file shares, email systems and document stores, organizations large and small are looking to impose order through Content Management. There are two types of content – structured and unstructured. Unstructured content includes: email, PowerPoint presentations, images, videos, audio recordings, documents, records and other files. ECM focuses on unstructured content to make it structured. Data mining is a tool used to extract important information from existing data and enable better decision-making throughout the banking and retail industries. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analysed and the information that is captured is used throughout the organization to support decision-making. Data Mining techniques can be very helpful to the banks for better targeting and acquiring new customers, fraud detection in real time, providing segment based products, analysis of the customers purchase patterns over time for better retention and relationship. Those banks that have realized the usefulness of data mining and are in the process of building a data mining environment for their decision-making process will obtain huge benefit and derive considerable competitive advantage in future.

Keywords: decision making, ECM, data miming, knowledge mining, OLAP.

I. INTRODUCTION

Enterprise content management, as a form of content management, combines the capture, search and networking of documents with digital archiving, document management and workflow. It specifically includes the special challenges involved in using and preserving a company's internal, often unstructured information, in all of its forms. Therefore, most ECM solutions focus on business-to-employee (B2E) systems.

As ECM solutions have evolved, new components have emerged. For example, as content is checked in and out, each use generates new metadata about the content, to some extent automatically; information about how and when the content was used can allow the system to gradually acquire new filtering, routing and search pathways, corporate taxonomies and semantic networks, and retention-rule decisions. Email and instant messaging are increasingly employed in decision-making processes; ECM can provide access to data about these communications, which can be used in business decisions. [04]

Solutions can provide intranet services to employees (B2E), and can also include enterprise portals for Business-to-Business (B2B), Business-to-Government (B2G), or other business relationships. This category includes most former document-management groupware and workflow solutions that have not yet fully converted their architecture to ECM, but provide a web interface. Digital asset management is a form of ECM concerned with content stored using digital technology

II. BACKGROUND

A. Need

In any large business, information is created, shared, edited and stored at a fast pace. With this heavy flow of content, it is easy for documents to become lost, deleted, changed or distributed without proper approval. An enterprise content management system lets you manage all of your company's content in one central location, making it easy to share, store, protect and control important information. With an enterprise CMS, you can also convert physical stacks of paper into digital documents. [07]

Most enterprise CMS solutions let you label documents using keywords so you can more easily find them amid the sea of your company's content. Using full-text search capabilities, you can easily find any document you are looking for. Security measures are also put into place to ensure your content is kept safe from anyone who hasn't received special access

B. Basic Concepts

Enterprise content management, as a form of content management, combines the capture, search and networking of documents with digital archiving, document management and workflow. It specifically includes the special challenges involved in using and preserving a company's internal, often unstructured information, in all

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

of its forms. Therefore, most ECM solutions focus on business-to-employee (B2E) systems.
As ECM solutions have evolved, new components have emerged. For example, as content is checked in and out, each use generates new metadata about the content, to some extent automatically; information about how and when the content was used can allow the system to gradually acquire new filtering, routing and search pathways, corporate taxonomies and semantic networks, and retention-rule decisions. Email and instant messaging are increasingly employed in decision-making processes; ECM can provide access to data about these communications, which can be used in business decisions.
Solutions can provide intranet services to employees (B2E), and can also include enterprise portals for business-to-business (B2B), Business-to-Government (B2G), Government-to-Business (G2B), or other business relationships [12]. This category includes most former document-management groupware and workflow solutions that have not yet fully converted their architecture to ECM, but provide a web interface. Digital asset management is a form of ECM concerned with content stored using digital technology.

C. Applications
ECM applications are delivered in three ways: on-premise software (installed on the organization's own network), software as a service (SaaS) (web access to information that is stored on the software manufacturer's system), or a hybrid solution composed of both on-premise and SaaS components.
ECM aims to make the management of corporate information easier through simplifying storage, security, version control, process routing, and retention [09]. The benefits to an organization include improved efficiency, better control, and reduced costs. For example, many banks have converted to storing copies of old checks within ECM systems versus the older method of keeping physical checks in massive paper warehouses. Under the old system a customer request for a copy of a check might take weeks, as the bank employees had to contact the warehouse to have someone locate the right box, file and check, pull the check, make a copy and then mail it to the bank who would eventually mail it to the customer [16]. With an ECM system in place, the bank employee simply searches the system for the customer's account number and the number of the requested check. When the image of the check appears on screen, they are able to immediately mail it to the customer—usually while the customer is still on the phone.

III. LITERATURE SURVEY

Since ECM is a relatively new concept within the field of IS, the existing body of literature is small and a common definition of ECM has not yet been developed. In this section, we shortly present the position of ECM as an IS research field, specifically in relation to knowledge management. Then, we analyze the current ECM literature. Finally, we review the definitions of ECM in the currently available literature and propose a more consistent and comprehensive definition.The position of ECM as an IS research field.
Tyrväinen et al. (2006) position ECM as a field of IS research, aggregating research results of diverging subjects such as retrieval algorithms, usability issues, or implementation methods. Nordheim and Päivärinta (2006) and Päivärinta and Munkvold (2005) regard ECM as a sub-field of knowledge management, since ECMSs can be used to capture and utilize content that contains explicit knowledge in repositories or to manage organizational knowledge resources. However, even Munkvold et al. (2006) and Päivärinta and Munkvold (2005) themselves argue that ECM incorporates fields that are distinctly different from knowledge management, such as the long-term storage of content or managing scanned invoices. By definition, this kind of content is not organizational knowledge which only exists in the heads of humans. Although ECMSs can be used for supporting knowledge management, it seems likely that ECM and knowledge management are in fact different fields of research that partly overlap (Dilnutt, 2006; Herschel and Jones, 2005; Kuechler and Vaishnavi, 2006), but which should not be subordinated to each other.

A. Literature Filter Criteria
We conducted a systematic literature review using the methods described by Webster and Watson (2002), who focus on the structure of a literature review paper, and Okoli and Schabram (2010), who focus on the process of conducting a systematic literature review. For identifying academic papers on ECM, we searched for papers in the following databases: ACM Digital Library, EBSCO, Google Scholar, IEEE Xplore, ProQuest and World cat. Additionally, we searched the databases of a number of well-respected conferences in the Information Systems field, including AMCIS, ECIS and ICIS.
We conducted full text searches up until 2009 using "Enterprise Content Management" and the combination of these terms, i.e. "Enterprise" AND "Content" AND "Management", as our search keys. Furthermore, we followed the practice described by Webster and Watson (2002) to look for citations in the found papers to check for earlier appropriate papers ("going back-ward"). Where possible, the databases were also used to look for papers which cite the found papers ("going forward"). This search process resulted in a collection of several hundred academic papers.
During the collection of the academic papers, we applied a practical screen to determine which papers should be kept for further study (Okoli and Schabram, 2010). Applying the screen has been alternated with the literature search in order to limit the amount of work for "going backward and forward". A rather tolerant screen was used since obtaining a broad overview of papers published in this domain was the goal. For example, limiting the selection to top outlets only would potentially exclude ECMS functionalities. This is particularly applicable for emerging research fields like ECM where the discussion is likely to occur in many different outlets. Screening the papers consisted of checking whether they just accidently

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

contained the words 'enterprise', 'content', and 'management' and whether they really addressed the topic of ECM. During screening, we developed a more elaborate understanding, resulting in continuous iterations while going through the literature. After screening our database of identified literature, 32 academic papers remained (see Table 1). The papers range from theoretical explorations of the ECM concept to empirical studies and summaries of practitioner case studies. We also decided to exclude Tyrväinen, Salminen, and Päivärinta (2003) as well as Salminen, Tyrväinen, and Päivärinta (2005) because they are only one page long introductions to the ECM-minitracks at HICSS 2003 and 2005.

We used two different methods for data extraction. To address the first research aim, all papers were scanned for definitions or descriptions of ECM. To address the second research aim, a more comprehensive method for data extraction was followed based on coding techniques from grounded theory research. The technique that was used is very similar to open coding. This is an analytical technique in which phenomena are named and categorized through close examination of data (Strauss and Corbin, 1990). This is also known as "labelling": taking a sentence, conceptualizing the general idea of this part of text and giving it a name. Labels which pertain to similar types of phenomena can then be grouped into categories, which can then be re-grouped by comparing them semantically. As described in more detail below, this coding finally resulted in the FEF.

When we analysed these academic papers during the creation of the FEF, we felt that the studied papers did not cover the full range of ECMS functionalities. Therefore, we also added practitioner papers to our database and included them in the data extraction. Including non-scientific literature represents the "clinical perspective" described by Schein (1987), who states that there often is a difference between what is being described in scientific literature and what practitioners believe to 'really be going on' (Schein, 1987, p. 13). We included practitioners' papers from a broad range of sources to limit bias, e.g. by including only a single vendor. Even though we could find a large number of sources, we chose those papers which we considered to have a certain quality. This resulted in the inclusion of nine practitioner papers for creating the FEF. They are from an often-referenced industry association (the Association for Information and Image Management, AIIM) and two major ECMS vendors who had provided more detailed descriptions of their ECMSs. We also included the observations of an ECM-consultant and descriptions from market researchers. As can be seen from Table 2 in the Appendix, the inclusion of the clinical perspective proved to be useful since several functionalities were either exclusively mentioned in practitioners' literature or are mainly derived from it.

B. Structuring the Literature

To structure our literature research, we used a framework for research on ECM defined by Tyrväinen et al. (2006), which has also been adopted, by other researchers (cf. vomBrocke and Simons, 2008). Using this framework, we categorized the identified papers according to the four perspectives used in the framework:

(1) *Content perspective* is composed of three views:

A. the *information view* is concerned with the semantics of the content and how it can be represented for different purposes;

B. the *user view*, which elaborates on how content should be presented in or-der to be interpreted correctly and fit the specific needs of the users; and

C. the*systems view* focuses on systems as containers of the content, which are accessed by the users.

(2) *Technology perspective* addresses the basic technologies used for ECMSs including, for example, hardware, software, and standards.

(3) *Enterprise perspective* "considers organizational, social, and business issues" (Tyrväinen et al., 2006, p. 630).

(4) *Process perspective* subsumes research about both the development and the deployment of ECMSs [14]

We added "Research Field" as the seventh category for structuring the literature review, which denotes papers that also comment on ECM as an IS research field, e.g. by providing a definition of ECM or by categorizing ECM research. This paperfor example can be categorized into the systems view and as commenting on ECM as a research field.

C. Defining ECM

In the identified papers, several definitions of the contemporary perception of ECM can be found. The definitions can broadly be divided into two different groups. The first group of papers (papers 2 till 8) concentrates on the content and technology perspectives, focusing "on the premise that all forms of content or unstructured data should be managed in a repository, independent of the applications utilizing the information. These concepts parallel first principals [sic] of structured data management and database systems" (Reimer, 2002, p. 18). These papers describe general functional requirements and the technologies needed for integrating content. Reimer (2002) clearly separates ECMSs from structured data management systems such as relational databases. We also found this perception in the evaluated practitioners' literature. For example, the Association for Information and Image Management (AIIM, 2005), an international industry association focusing on ECM, perceives ECM from a content and technical perspective, but already broadens its definition to include related methods and adds a strategic notion, by defining ECM as "the technologies used to capture, manage, store, pre-serve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists. [...] Content must be managed so that it is used to achieve business goals. Central to this strategy

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

are the tools and technologies of ECM, which manage the complete lifecycle of content, birth to death" (ibid).

The second group of identified articles defines ECM from an enterprise or process perspective (papers 16 till 28). Tyrväinen, Salminen, and Päivärinta (2003) describe ECM as focusing "on the management of textual and multimedia content across and between enterprises, emphasizing the coexistence of technical and social aspects within the content management. Method sand techniques applicable for managing textual and multimedia information with all sizes of content units, ranging from XML and database structures through web pages and documents to document collections, are studied as well as approaches focusing on specific content structures" (ibid, p. 2). Smith and McKeen (2003, p. 647-648) similarly define ECM as an "integrated approach to managing all of an organization's information including paper documents, data, reports, web pages and digital assets" and "the strategies, tools, processes, and skills an organization needs to manage all its information assets (regardless of type) over their lifecycle". This definition is also used by Reich and Behrendt (2007) and Rückel et al. (2007).

The process perspective has gained considerable attention from several researchers. Nordheim and Päivärinta (2006) define ECM as representing "a modern concept of Information Re-source Management in general, addressing the integration of semi- and unstructured data with the management of formal databases" (ibid, p. 649). Tyrväinen et al. (2006) further specify the content lifecycle to include "activities such as content creation and capture, content editing, review, approval, content indexing, classifying and linking, content distribution, publication and use, update, preservation, format transformation for long-term archival, and retention" (ibid, p. 631).

A notable difference in definitions is related to the use of structured data on the one hand and unstructured or semi-structured data and/or information on the other. The AIIM limits ECM to the management of "unstructured information" (AIIM Europe, n.d.). The view of limiting ECM to unstructured or semi-structured data is not only shared in the practitioners' literature. Scientific literature such as Reimer (2002), O'Callaghan and Smits (2005) and Andersen (2008) also limit ECM to unstructured information. However, there are numerous academic papers which also consider structured data as content that can be managed by ECMSs. We decided to follow this perception, since it emphasizes that the scope of ECM covers all the "information assets (regardless of type)" of an organization (Smith and McKeen, 2003, p. 648). The broad range of subjects covered by ECM shows that it is more than "the latest buzzword" (Mescan, 2004, p. 55). However, previous definitions vary and are at times contradictory. Therefore, we propose a more consistent and comprehensive definition of ECM:

"Enterprise Content Management comprises the strategies, processes, methods, systems, and technologies that are necessary for capturing, creating, managing, using, publishing, storing, preserving, and disposing content within and between organizations."

The proposed definition summarizes all relevant perspectives of ECM which have been mentioned in the first seven years of research on this topic and provides a common conceptual basis for further research in this field. It points out that ECM is not limited to technologies, but rather covers a wide range of subjects so that this definition is aligned well with the notion of ECM being its own field of IS research. Finally, the definition also includes a specification of the content lifecycle mentioned in previous definitions and therefore further illustrates the breadth of this concept.

D. Functional ECM Framework

To our knowledge, there is no overview available in literature that summarizes the functionalities that an ECMS can provide. If contributions can be found, for example Grossniklaus and Norrie (2002), they do not have a focus on ECM, but a more narrow focus, such as (web) con-tent management. The reviewed articles listed in the previous section mostly have a more in-formational or technological focus on content management. Developing an overview of ECMS functionalities addresses several scientific and practical purposes. First of all, a consensus about ECMS functionalities will serve to understand and communicate about ECMSs. Secondly, an overview of ECMS functionalities can direct future research concerning specific functionalities or concerning ECMSs in general. It can for example be used as a reference in case study research to describe which ECMS functionalities are used by a particular organization. Thirdly, it can be the basis for a more formal reference architecture that lays the foundation for designing and building ECMSs. Finally, the FEF can be used as an assessment tool by practitioners for comparing the functionalities provided by existing ECMSs.

We refer to the ECMS functionalities overview as the Functional ECM Framework (FEF). The term 'functional framework' has been carefully selected. We prefer it over other terms such as taxonomy, reference architecture or reference model, because these terms have specific connotations in related research domains. A taxonomy is by definition characterized by a strict hierarchical decomposition of elements (cf. Nagra, Thomborson and Collberg, 2002). As will be shown later, the FEF is based on a number of guidelines derived from literature which only partially suggest a hierarchical decomposition. The terms reference architecture and reference model are established concepts in the field of software architecture and software development. While a reference model shows the functional requirements of an ECMS (Soft-ware Engineering Institute, n.d.), a reference architecture rather presents a technical implementation of functionalities in software components (Angelov, 2006; Bordegoni et al., 2002; Grefen and Vries, 1998; Mellish et al., 1996; Software Engineering Institute, n.d.).

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

As one of our foremost aims is to support the understanding of and the communication about ECMSs, we present a functional framework that leaves out the details of implementation. These details are mainly of interest for developers of an ECMS rather than for example for managers in an organization who need to make decisions about required ECMS functionalities. To summarize, the FEF is a representative of Orlikowski and Iacono's (2001) 'tool' view, i.e. we aim at describing what an ECMS is and what it is intended to do.

E. Defining Functionality

The terms 'functionality' and 'function' (which we perceive as synonyms in the context of information systems) are often used when describing information systems, but their definitions left implicit most of the time. However, we need a definition for labeling particular pieces of text as ECM functionality while scanning the literature.

Definitions for the term function can be found in general IS literature as well as in the Enterprise Modeling (cf. ARIS, DEM, Zachman Framework) literature. Generally spoken, a function transforms inputs into outputs (Davis, 2001; Zachman, 1987). In the information systems literature, a function is typically referred to as a capability of an information system (Rolland and Prakash, 2001). This means that certain information inputs are transformed into information outputs (Stair and Reynolds, 2006), for example customer order data that is being transformed into a customer invoice by the invoice function. In the enterprise modeling literature, a (business) function typically relates to a particular business process or a cluster of business processes that can be broken down in activities and tasks (Davis, 2001; Turban et al., 2005; Zachman, 1987). Functions of information systems and business functions are directly related since an information system function should execute or support specific business functions (activities or tasks) for intended users (Nickerson, 2000). Summarizing the above for this research, we define functionality as a capability of an information system referring to a particular business function(s) that needs to be executed by the ECMS.

F. Creation Guidelines

Considering the previously described aims of a FEF, we formulated design guidelines to guide the development of the FEF which are inspired by literature on reference architectures. Authors in this research field have been explicit about such guidelines, which is not the case for the (also less abundant) literature on functional taxonomies or frameworks.

The first and most important guideline is that the FEF is *comprehensible and usable*. The FEF needs to be "understandable and usable by the communities targeted" (Bernus and Nemes, 1996, p. 180), which is in line with our main aim. In this case, the targeted communities are the ECM research and practitioner communities as they should be able to understand and use the FEF. Secondly, the FEF needs to be *complete*, i.e. it has to include all possible functionalities currently described and has to be composed independently of the thought that there needs to be a concrete ECMS that includes every possible functionality (Angelov, 2006; Grefen and Vries, 1998; Williams, 1994). Only if the FEF is able to describe a wide range of existing ECMS implementations, it is of use for both researchers and practitioners. Thirdly, the FEF should be *generic* enough to be used for making comparisons among different ECMSs by making it supplier and solution independent. Consequently, it should hide low-level and sup-plier-specific details (Grefen and Vries, 1998; Williams, 1994). Fourth, it needs to be *distinguishing* enough so that differences between implementations at different organizations can be captured and analyzed (Grefen and Vries, 1998). Fifth, the FEF needs to be *future-proof* and should be constructed in such a way that it is able to accommodate future developments (Bernus and Nemes, 1996). It should in particular support the integration of new functionalities that can be expected because the field is still in development.

G. Creating the FEF

We started the creation of the FEF by extracting a list of functionalities from the literature and as mentioned before, we also included practitioners' literature in this extraction process. For the extraction, we used a technique similar to open coding. We scanned the collected literature, extracted pieces of text describing functionalities and labeled them. The label of each functionality is a description that best summarizes the functionality as interpreted by the re-searchers. For example, the text fragments "many methods of imputing data and information into an application or database. Other methods include: on-line entry, input from other applications, Electronic Data Interchange, […]" (Dilnutt, 2006, p. 78) was labeled as 'capture: digital forms', 'capture: digital sources', and 'client application integration'. Another example is the sentence "effective content creation and capture from heterogeneous external and internal sources (integrated production environments, scanning and imaging, conversion of file formats, forms-based data capture)" (Päivärinta and Munkvold, 2005, p. 4), which was la-belled as 'imaging', 'capture: digital sources' and 'capture: digital forms'. Labelling is typically an iterative process where labels are adapted if new labels are being defined. After label-ling the potential functionalities, we eliminated duplicates and further harmonized the names. This resulted in the list as presented in Table 2 in the Appendix. However, a list by its very nature does not provide an intuitive overview and can also be impractical for comparing different ECMSs. For producing an overview of functionalities that fulfils the guidelines as mentioned above, we decided to proceed with a graphical approach.

Our next step was to divide the functionalities into main categories. We grouped labels pertaining to similar types of phenomena into categories, which we then re-grouped by comparing them semantically. This was an iterative process in which we divided functionalities into preliminary categories, changed categories and re-assigned functionalities again until we felt we had an understandable and representative categorization.

DOI 10.17148/IJARCCE.2015.4314

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

We were also inspired by literature describing models for presenting functionality of information systems in layers. Fowler (2003) describes a four layered presentation consisting of four layers: presentation (handling interaction with users), service (contains for example transaction control), domain (performs processing of input and generates outputs), and data source (communication with for example a database). Similarly, McKeever (2003) presents a WCM hierarchy consisting of the following four layers: Audience (groups of people interacting with the WCMS), Outlet (types of outlets through which the content can be accessed), Activity (activities involved in managing content), and Content (types of content).

H. Additional Observations during Literature Review

During the literature review, we encountered three issues which are worth mentioning here.

1. ECMS as Middleware

In the interest of completeness, it is noteworthy to mention that a different view on the main purposes of an ECMS can be found in literature. Although Bandorf et al. (2004) agree that an ECMS includes, for example, the functionalities portal integration and information retrieval, the authors consider ECMSs to be mainly a middleware infrastructure for content. According to their perception, the ECMS itself only stores metadata which describes content that is stored in repositories managed by other systems (e.g. file systems, data bases, and applications). The ECMS integrates the various repositories and provides other applications with access to these 'content stores' (Bandorf et al., 2004). Related to the functionalities mentioned above, this point of view would mean that the *EAI interface* functionality is the central element of an ECMS.

2. Management of Structured Data

As described above, ECM(S) is/are also concerned with the management of so-called structured data. The functionalities *business intelligence* and *business activity monitoring* are often used for analysing structured data about workflow initiations and therefore ECMSs already contain certain functionalities for managing structured data. However, hardly any source further elaborates on this topic, for example by explaining how other sources of structured data are accessed and managed in practice. Smith and McKeen (2003) mention that databases and data warehouses can be used for managing structured data, but do not define whether these databases are part of the ECMS or whether an ECMS needs additional functionalities besides the ones already included in the FEF. Even exponents of the inclusion of structured data in ECM state that "ECM and ERP systems [containing structured data, eds.] clearly represent two different approaches to such issues as workflow and data management" (Nordheim and Päivärinta, 2004, p. 7) or describe a project where the structured data remains in the Enterprise Resource Planning (ERP) system and is only linked to content stored in the ECMS (Päivärinta and Munkvold, 2005). These observations confirm the previous conclusion that the *EAI interface* functionality presumably plays an important role in the management of structured data. It is also assumed that this functionality needs to be combined with a special functionality for managing structured data, e.g. for extraction, transformation and load operations (Doculabs, 2004).

3. ECMSs as Multi-Product Software

The dissimilar characteristics and the diversity of functionalities offered by an ECMS lead to the conclusion that although ECMSs are marketed under a single term, an ECMS implementation will be an integration of multiple software products (Reich and Behrendt, 2007). ECMSs can offer functionalities which are very different, if not entirely opposite, from each other. The management of electronic records, for example, requires protecting content from changes; potentially for an unlimited period of time and even if the content has become outdated in the meantime. When managing electronic documents, changes are allowed to occur, but normally need to be tracked so that old versions of a document can be reverted to. Digital assets such as for example video files require different user interfaces than text files. This view is supported by the history of ECMSs, which are rooted in different classes of information systems: some suppliers of traditional Electronic Document Management systems have extended their products with functionality for managing web content by acquiring specialized companies and their products, whereas some traditional suppliers of Web Content Management systems have done the opposite (Dilnutt, 2006). Böhm (2007) even argues that certain functionalities are not covered by products from the ECMS vendors, but rather by specialized third-party products. Therefore, a parallel can be drawn with the best-of-breed approach for implementing ERP systems (ERPSs): instead of introducing only a single ERPS that needs to cover all functionalities, several standard and customized products are integrated with each other. It has been shown that this approach requires a different implementation process than the introduction of a single system (Light, Holland and Wills, 2001) and it can be assumed that this is also true for ECMSs. However, the current literature on the implementation of ECMSs (e.g. Nordheim and Päivärinta, 2006; Scott et al., 2004) does not explicitly analyse the process of 'multi-product implementation' within the context of ECMSs and therefore, this topic remains open to further research.

I. Applying the FEF in Practice: Case Study Research

Before creating the FEF, we had defined five guidelines, namely that the FEF is (1) *comprehensible and usable,* (2) *complete*, (3) *generic* enough to be used for making comparisons, (4) is *distinguishing* enough for capturing differences, and (5) is *future-proof*. After creating the FEF, our next step was to conduct initial practical tests whether it actually embraces these guidelines. Therefore, we applied the FEF in three case studies in which we studied the impacts of implementing ECMSs in organizations. Applying the FEF to actual ECMS implementations also provided us with an improved insight on potential

DOI 10.17148/IJARCCE.2015.4314

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

interactions among the different (groups of) functionalities.

In the case study research (Grahlmann et al., 2010), we used the FEF to capture the functional scopes of ECMSs, being defined as "the range of business functions" (Karimi, Somers, and Bhattacherjee, 2007, p. 105) offered by the ECMS. ECMSs at three organizations (A, B and C) were evaluated by conducting semi-structured interviews and by studying available documentation. Based on this information, the functionalities used in a particular organization were coded into the FEF by highlighting the applicable boxes. These codifications have been reviewed by the respondents, serving as a coding check. Two examples of the results are depicted in Figures 2 and 3.

IV. EXISTING SYSTEM

1. IBM Enterprise Content Management

IBM Enterprise Content Management System Monitor (ECM System Monitor) proactively monitors, automates and reports on ECM-based IT services and infrastructure for improved efficiencies. It can improve uptime and service-level delivery to help reduce costs associated with ECM platform operations. IBM ECM System Monitor integrates with your existing enterprise system management infrastructure for more expansive system monitoring functions.

2. Microsoft Office 365

Office 365 is a subscription-based online office and software plus services suite which offers access to various services and software built around the Microsoft Office platform.

Serving as a successor to Microsoft's Business Productivity Online Suite, the service was originally designed to provide hosted e-mail, social networking and collaboration, and cloud storage to teams and businesses. As such, it first included hosted versions of Exchange, Lync, SharePoint, Office Web Apps, along with access to the Microsoft Office 2010 desktop applications on the Enterprise plan. With the release of Office 2013, Office 365 expanded to include new plans aimed at different types of businesses, along with new plans aimed at general consumers wanting to use the Office desktop software on a subscription basis.

OTHER TECHNOLOGIES

(i) HCL's Enterprise Content Management

HCL's enterprise content management practice gives its customers the advantage of a focused domain centric approach and platform competencies leading to strategic information management for the enterprise.

Managing content and information is one of the most crucial requirements for today's organizations. Content is clearly defined a king these days when it comes to an efficient operation and gaining a competitive advantage over others. Since unstructured information comprises of over 80% of an organization's information assets, on average Enterprise Content Management (ECM) emerged as one of the top 4 priorities for a vast majority of CIOs today. The two main reasons why ECM is becoming so important for CIOs are:

1. Enterprise Information Management - Ensures interoperability and extensibility - as need to maintain total visibility of global business becomes imperative
2. Compliance - It is the main driving force behind exploding growth in the ECM space. It helps avoid litigations leading to stunted business growth [06]

(ii) Microsoft Business Productivity Online Suite

Microsoft Business Productivity Online Suite is a set of Microsoft Hosted Massaging and Collaboration Solutions including Microsoft Exchange Online. Microsoft SharePoint Online, Microsoft Office Live Meeting and Microsoft Office Communications Online.These online services are designed to give your business streamlines communication with high availability, comprehensive security, and simplified IT management.

Microsoft Business Productivity Online Standard Suite is a set of messaging and collaboration tools, delivered as a subscription service, that gives your business rich capabilities without the need to deploy and maintain software and hardware on-premise. These online services are designed to help meet your needs for robust security, all-day every-day reliability, and user productivity

If you are a mid-level company and need to host your website on Microsoft servers rather than buying any other service provider or setting up your own web server, it's quite easy and cheap for those companies which are trying to get ahead by impressing their clients that they are using cloud services of Microsoft [13]

When your company subscribes to one or more of the services, it connects to the rich interactivity of on-premises client and server applications with the flexibility and scalability of Web-based services. Because Microsoft hosts the services, it reduces your company's need to maintain IT services on site.

Many organizations today are moving to this "software-plus-services" model, because it decreases the need to manage hardware and software.

V. OUR ANALYSIS

- Our system is basically a web application which will be authorised to be used by the Higher Authority/Manager of the corresponding organization.
- The application would extract the data for analysing from the banks database and then perform required operation as selected by the logged user.
- The application will be allowing following operation to be performed,

  - Customer Behaviour Analysis
  - Credit Assessment
  - Attrition Analysis

- Identify the common functionality used for the development of data mining solutions.
- Algorithms selected for the implementation of various operation,

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

➢ K-Mean and then applying Aproiri Algorithm for Customer Behaviour Analysis

- Support Vector Machines or Bayesian Network Algorithm for Credit Assessment
- Selective Naive Bayesian (SNB) or Boosted Naive Bayesian (BNB) or Neural Network for Attrition Analysis.

The aim of our system is to integrate the Enterprise or process oriented business with our application and maintain the soft copy of Data.In the analysing step information is retrieved and maintained from extracted data for decision support in an Enterprise.As the system is for an Enterprise we have large data to maintain and we have to retrieve at real time so implementing OLAP in the data warehousing.For implementing higher Security for user data we are implementing MVC in asp.net.Services should be provided to higher authority in an efficient manner. It can be reused in similar applications.

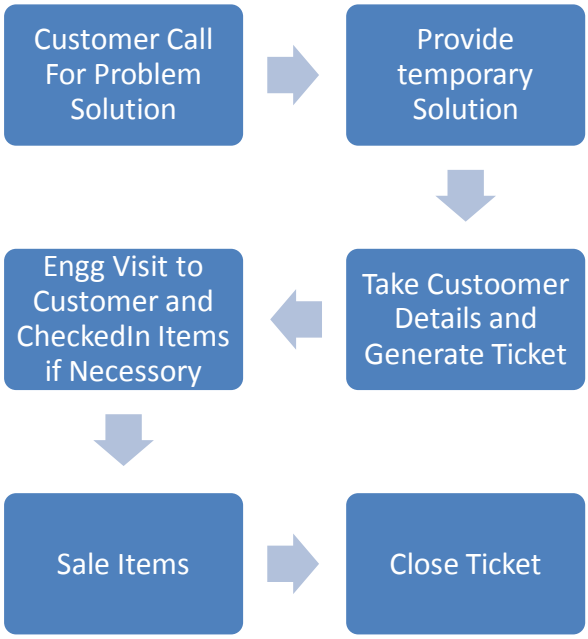

Fig.1 System Flow Diagram

VI. CONCLUSION

Data mining is a tool used to extract important information from existing data and enable better Content Management through Enterprise Content Management System. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed and the information that is captured is used throughout the organization to support Content Management.

Data Mining techniques can be very helpful to the Organizations for better targeting and acquiring new customers, fraud detection in real time, providing segment based products, analysis of the customers purchase patterns over time for better retention and relationship. Those banks that have realized the usefulness of data mining and are in the process of building a data mining environment for their decision-making process will obtain huge benefit and derive considerable competitive advantage in future.

A limitation of this research concerns the case studies in which the FEF has been applied. This is because we only studied two Dutch public agencies and one relatively small commercial organization with main operations in the Netherlands. None of them included a 'complete' ECMS installation and also the combined functional scopes of all three case studies did not cover all potential functionalities of ECMSs. For this reason, it is not entirely sure whether the FEF really contains a complete list of potential functionalities.

REFERENCES

[01] Bank Credit Risk Analysis with Bayesian Network Decision Tool, Berk Bekiroglu* et al. / (IJAEST) International Journal of Advanced Engineering Sciences and Technologies Vol. No. 9, Issue No. 2, 273 – 279

[02] Gajraj Singh and DR. P.K. YADAV, "Implementation of the Associative Classification Algorithm and Format of Dataset in Context of Data Mining", International Journal of Advanced Research in Computer Science, 4 (8), May–June, 2013,18-21

[03] Credit card fraud detection using anti-k nearest neighbor algorithm, VENKATA RATNAM Ganji (IJAEST) ISSN : 0975-3397 Vol. 4 No. 06 June 2012

[04] Application of data mining techniques in customer relationship management:A literature review and classification, E.W.T. Ngai , Li Xiu , D.C.K. Chau , E.W.T. Ngai et al. / Expert Systems with Applications 36 (2009) 2592–2602]

[05] Real Time Data Mining-based Intrusion Detection Wenke Lee, Salvatore J. Stolfo, Philip K. Chan, EleazarEskin , Wei Fan, Matthew Miller, ShlomoHershkop, and Junxin Zhang, DARPA (F30602-00-1-0603)

[06] Mining customer product ratings for personalized marketing Kwok-Wai Cheung, James T. Kwok , Martin H. Law , Kwok-ChingTsui, Decision Support Systems 35 (2003) 231– 243

[07] Knowledge Mining and Trifle Management for Predicting Heart Disease, Mr.Bensujin, Dr.Keziselvavijila and MrsCynthia Hubert, International Journal of Advanced Research in Computer Science, 3 (7), Nov.-Dec. 2012, 211-217

[08] Jianchao Han, "Learning Fuzzy Association Rules and Associative Classification Rules", IEEE International Conference on Fuzzy Systems, 2006, pp 1454-1459

[09] Mikut, R. &Reischl M. (2011) "Data mining tools", WIREs Data Mining and Knowledge Discovery, Vol. 1, No. 5, pp. 431-443

[10] Swapnil R. Isod and Amit M. Sahu, "Clustering Techniques", International Journal of Advanced Research in Computer Science, 4 (6) Special Issue, May 2013,5-9

[11] Gonzalo Mariscal, O´ Scar Marba´ N and CovadongaFerna´ Ndez, "A survey of data mining and knowledge discovery process models and methodologies", Vol. 25:2, 137–166. & Cambridge University Press, 2010 doi: 10.1017/S0269888910000032

[12] R.Rajalakshmi, Dr.K.Thangadurai, M.Uma and Dr.M.Punithavalli, "Impact of Datamining Techniques in Forecasting Plant Disease", International Journal of Advanced Research in Computer Science, 3 (7), Nov –Dec, 2012,187-190

[13] A. Jurek, D. Zakrzewska, Improving Naive Bayes Models of Insurance Risk by Unsupervised Classification: Proceedings of the International Multi- conference on Computer Science and Information Technology pp. 137–144

[14] A. ANBAR, Credit Risk Management in the Turkish Banking Sector: A Survey Study: ElektronikSosyalBilimlerDergisi ISSN: 1304-0278 Yaz -2006 C.5 S.17 (10-24)

[15] Q. Zhang and R. Segall, Web mining: a survey of current research, techniques, and software, International Journal of Information Technology and Decision Making, Vol. 7, Issue: 4, 683-720, (2008), DOI No: 10.1142/S0219622008003150

[16] Shiv Pratap Singh Kushwah, KeshavRawat, Pradeep Gupta, "Analysis and Comparison of Efficient Techniques of Clustering Algorithms in Data Mining", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-3, August 2012

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

[17] Dhanashree S. Deshpande, "A Survey on Web Data Mining Applications", Emerging Trends in Computer Science and Information Technology -2012(ETCSIT2012) Proceedings published in International Journal of Computer Applications® (IJCA)

[18] KDnuggets Software Poll, http://www.kdnuggets.com/polls/2012/analytics-data-mining-big-data-software.html [10-9-2013]

[19] Data mining Alexander Doug, http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/, [10-9-2013]