UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

An activity-based model for managing unstructured content

Mini-dissertation by

Izak Francois Jakobus Roos
(91133204)

submitted in partial fulfilment of the requirements for the degree

MASTER OF INFORMATION TECHNOLOGY

in the

SCHOOL OF INFORMATION TECHNOLOGY

of the

FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY
UNIVERSITY OF PRETORIA

Supervisor: Prof TJD Bothma **April 2009**

Acknowledgements

I would like to thank the following people

- Prof Theo Bothma for his encouragement, advice and patience.
- My wife Melanie Roos and my son Marcell Roos for their encouragement and support.

Table of Contents

List of Figures

An activity-based model for managing unstructured content

Abstract

An increase in the volume of documents, images and other unstructured content in organisations has resulted in several new approaches to the management of unstructured content. These approaches are however focused on specific problem areas and do not address the problem of how to manage all unstructured content in a unified manner.

This research investigates the requirements for managing unstructured content from a records management, knowledge management and information security perspective. To address the management requirements, a model for managing unstructured content by using organisational activities as a basis for categorising the content into a taxonomy is proposed. The management requirements for unstructured content are then determined and applied on the taxonomical node level.

The research showed that a strong relationship exists between organisational activities and the content management requirements of unstructured conent. It was also shown that using an activity-based model for classifying unstructured content, can be used as the basis for managing unstructured content from a records management, knowledge management and information security persective.

Keywords: Knowledge Management, Records Management, Document Management, Content Management, unstructured content, acivity-based, framework, model

Chapter 1 - Introduction

Managing unstructured content stored on corporate networks is becoming increasingly more complex for organisations. This is partly due to an increase in the amount of data stored on corporate networks coupled with new regulatory requirements concerning the management of corporate data. Unstructured content includes e-mail, documents, spreadsheets, PDF files, images and digital movies.

The increasing amount of data stored can be attributed to more powerful 'user productivity' tools, enabling users to view and manipulate larger data sets, increasingly richer digital content such as movies and higher quality images and the pervasiveness of electronic communication. This incessant storage requirement drives the development of higher capacity storage media at lower costs, which in turn removes the user's constraints on storing the data.

Managing unstructured content is necessary to comply with the regulatory requirements focusing on the management of electronic records and to ensure that the knowledge and information contained within documents are managed as a corporate asset. Current practices for content management include records management, documents management, enterprise content management and knowledge management. There is, however, little guidance for organisations on implementing any or all of these tools to ensure that the same set of rules is used for managing the entire collections of unstructured content within the organisation.

Documents are created and used during the execution of organisational processes and activities (Strong, 1999, p22). In each of the above-mentioned areas of document management the process / activity, oriented approach to managing content has been researched and shown to be effective. It thus seems appropriate to suggest that this approach be employed to manage the entire collection of unstructured content in organisations. This research will propose a model for managing organisational content, based on the processes and activities where the content is created and used.

1.1 Problem Statement

No unified content modelling approaches are currently in use for managing the unstructured content in organisations in terms of records management, knowledge management and information security.

1.1.1 Research question

To what extent does an activity-based taxonomy satisfy the records management, knowledge management and information requirements for managing unstructured content?

1.1.1.1 Sub-questions

- What are the requirements for managing unstructured content?
- What is the relationship between organisational activities and the management requirements for unstructured content?
- How does an activity-based taxonomy satsify the management requirements of unstructured content?

1.2 Research Method

The approach that will be followed in order to answer the research question is to construct a model for managing unstructured content that uses organisational activities to determine the management requirements for the content.

Olivier (2003, p51) states that the construction of a model is often used as the primary research methodology. Frank (1998, p112) argues that the creation of conceptual models by using a constructive research approach may constitute a research goal. Kasanen et al. (1993, p245) describe constructive research as an approach to solving managerial problems by constructing entities, which provides solutions to these problems. According to (Schwarz et al., 2007, p33-p34) a framework can be used “to synthesize previous research in an actionable way for practitioners”.

March and Smith (1995, p252-254) define two types of scientific interests in the field of Information Technology, namely descriptive and prescriptive. Descriptive interests deal with understanding IT phenomena and artefacts while prescriptive ones tackle solutions to IT problems. The descriptive interests tend to follow a more natural science approach where the aim is to develop and prove theories. Prescriptive

research on the other hand follows a design science approach with the aim of building constructs, models, methods and implementations by applying existing knowledge. Design science consists primarily of two activities, building and evaluating, while the primary activities of natural science are to theorize and justify theories during research. Constructive research can be viewed as a design science, which means that creation and evaluation of artefacts should be the primary research activities.

Kasanen et al. (1993, p246) describes the constructive research process as follows:

"1. Find a practically relevant problem which also has research potential.
2. Obtain a general and comprehensive understanding of the topic.
3. Innovate, i.e., construct a solution idea.
4. Demonstrate that the solution works.
5. Show the theoretical connections and the research contribution of the solution concept.
6. Examine the scope of applicability of the solution."

This approach, combined with that of Robinson (2006, p795), yields the following steps for model construction, which will be used as the research approach:

- Clearly define the purpose of the model
  This has been addressed in the section regarding the problem statement.
- Conduct a literature survey to collect information on the problem domain.

  The literature survey should, according to Mouton (2005, p87,90), be used to construct an in-depth understanding of the problem domain. This corresponds with the approach described by Kasanen et al. (1993, p246) that should be followed when constructing a model.
  There is a substantial amount of literature on the management of documents and records in organisations, with research having been undertaken on records management, electronic document management, knowledge management and access control. This body of knowledge will be used to determine the content management requirements and current approaches in managing unstructured content. Specific areas that will be examined are the following:

- Content management requirements for the areas mentioned.
- Current techniques used to manage documents in respect of the above.

The researcher will

- Identify the components of the model and the relationships between them;
- Construct the model;
- Clearly specify any assumptions;
- Show the theoretical connections and the research contribution of the solution concept;
- Examine the scope of applicability of the solution.

The model will be evaluated by assesing the model's ability to satisfy the content management requirements. This will be done by first by evaulating the requirements for content management from a stakeholder perspective and subsequently showing how the requirements can be met by using the proposed model. A stakeholder in this context can be defined as any user or group of users who have a requirement for content management

1.3 Summary

This chapter explained the problem of managing unstructured content in an organisational environment. A research question with sub-questions was formulated to direct research into this area. The researcher explained that a model will be constructed to answer the research question and sub questions. The approach in doing so is to undertake a literature survey, followed by the model's construction and an evaluation of it. A review of the literature is to be found in Chapter 2, the purpose being to determine the requirements for managing unstructured content as well as to identify any direct or indirect relationships between these requirements, the unstructured content and the organisational activities. Chapter 3 proposes and describes a model for the management of unstructured content, while the proposed model is evaluated in Chapter 4.

Chapter 2 - Literature Survey

2.1 Introduction

This chapter will examine the published literature to determine the requirements for managing unstructured content from a records management, knowledge management and information security perspective. Relationships between the content management requirements and organisational activities will also be examined. The outcome of this literature survey will then be used in the construction of the content management model in Chapter 3.

2.2 Records Management

Wiggins (2000, p62) describes records as recorded information created during the activities of the business which, when viewed over time, provide a view of the continued business activities of the organisation. Penn et al. (1994, p3) describes a record as “any information that is:

- recorded on any form of physical medium
- generated by a business enterprise as evidence of its organization, functions, policies, decisions, procedures, operations and internal or external transactions
- valuable of the information it contains”

The primary aim of records management is to ensure that records are retained for evidentiary or historic purposes (Jimerson, 2003, p136). A more detailed definition of records management is provided by Penn et al. (1994, p7) as comprising “a logical and practical approach to the creation, maintenance, use and disposition of records and, therefore, of the information that those records contain”.

When considering the relationship between documents and records it should be noted that some documents might be records, which means that an integrated approach should be followed when managing both (Wiggins, 2000, p66).

Records have traditionally been created on paper, which from a retrieval perspective ensured that even misfiled records could be retrieved by going through all the paper records, provided that the record had not been destroyed. Electronic records, however, can be incorrectly updated, thereby destroying the original, or can be lost if the electronic record keeping system is not adequately protected by means of

backup copies and other data security practices. They can also be misfiled to the extent where they may be virtually impossible to locate. Since records are created and kept as evidence of the organisation's activities, losing them could result in negative legal and financial consequences for the organisation (Wiggins, 2000, p65). This means that all records should be managed in a consistent manner, regardless of the medium on which the record is stored (Penn et al., 1994, p9, 10).

Several recent changes in legislation worldwide have brought the issue of electronic records management to the forefront (Middleton and Smith, 2002, p335). In South Africa specifically, the ECT Act of 2002 (SA 2002) specifies that electronic messages can be regarded as legal documents; this Act also goes further, to set specific rules for the management of electronic messages, containing a broad guideline as to what such a message is: "...'data message' means data generated, sent, received or stored by electronic means". Owing to such a broad definition, little guidance is given to organisations in terms of what should be retained and what not.

2.2.1 Requirements for records management

According to Cisco and Strong (1999, p5) (citing Silver, 1998) the main features of records management are:

- Record selection - is the document an official record or not?
- Classification - This is the assignment of categories in the organisational filing system to the record.
- Retention - The period for which the record must be kept.

To select, classify and specify retention periods for a small collection of documents could already represent daunting tasks; when this process is extended to an organisation receiving thousands of e-mails messages and that contains a document archive of millions of documents the task becomes impossible. This means that the end user will have to be involved in classifying his or her own documents (Cisco and Strong, 1999, p6). Determining the retention time is, however, not always easy or clear for the end user (Middleton and Smith, 2002, p336). To assist end users in the process of selection, classification and retention, an indexing system that is simple and to which the end users can relate is needed (Cisco and Strong, 1999, p6). The problem of document retention is further complicated by the fact that documents cannot be destroyed if litigation, where a legal requirement to provide the documents

might exist, is pending or in progress. This means that the relationships between electronic documents and other information sources also need to be specified. These could consist of other documents and structured data in formal databases.

From the above a ‘retain everything approach’ might seem to be the best option, especially with the continual decrease in the cost of electronic storage media. This would however leave organisations open to the risk of "standard disclosure", where documents that did not need to be retained can be discovered during legal action and negatively impact on the organisation (Middleton and Smith, 2002, p334).

2.2.2 Records management lifecycle and records continuum

The two main approaches to records management are the lifecycle approach and the records continuum approach.

2.2.2.1 Records lifecycle approach

Penn et al. (1994, p12) describe an approach to records management based on the lifecycle of the information. Their lifecycle consists of three main phases, namely: the creation phase, the maintenance and use phase and the disposition phase, each with specific records management activities associated with it. The creation phase deals with the identification and creation of required records. The maintenance and use phase consists of record storage, retrieval and record scheduling according to a retention schedule. The disposition phase deals with the way the record is handled once it becomes inactive, which could be destruction or archival. Wiggins (2000, p68) takes this approach further by expanding the records lifecycle to acquisition, creation, approval, retention, indexing, storage, retrieval, communication, revision, utilisation and destruction.

2.2.2.2 Records continuum approach

The records continuum is the basis for records management in Australia (Sletten, 1999, p1) and is aimed at providing an integrated approach between records management and archives management.

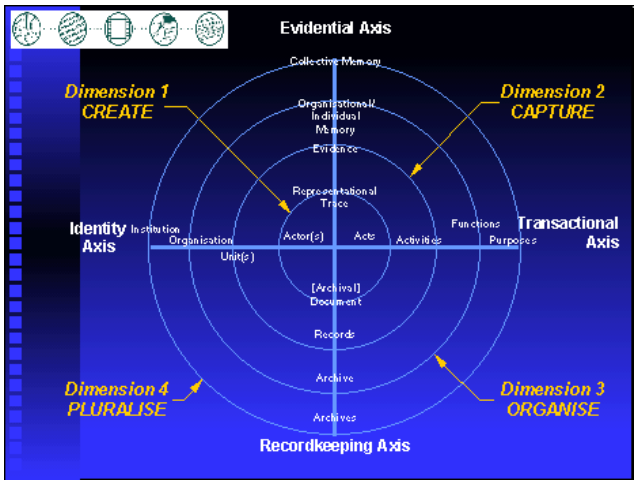

Figure 1 - Records Continuum Diagram (source: Upward, 1996)

The records continuum model (see figure1) as described by Upward (1996) consists of four dimensions, which roughly correspond to the life cycle of the records: create, capture organise and pluralise. The biggest difference between these dimensions and the records lifecycle is that a record can belong to more than one dimension in the continuum, whereas a record in the lifecycle approach can only be located in one specific phase of the lifecycle at any specific point in time.

The model also contains four intersecting axes, which are used to describe the state, properties and management of the record, based on the dimension where the record is located. These axes are the following:

Record Keeping Axis: This axis deals with the way that records are stored and aggregated into collections of records according to a specified classification scheme.
Evidential Axis: This axis describes the role the record plays in providing evidence of actions and its further role as a source of individual, corporate and collective memory.
Transactional Axis: The transactional axis describes the record in terms of the acts within specific activities that are documented by the records. The aggregation of activities and the records related to those activities serves as a record for the functions carried out in the organisation. Functions are aggregated into purposes, with the related records or archives or records describing these purposes.
Identity Axis: This axis provides a hierarchical functional organisational view which corresponds with the acts, activities, functions and purposes of the transactional axis.

2.2.3 Components of records management

2.2.3.1 Records inventory

The records inventory is created to identify all record series and how they are used.

Specific elements of the records inventory as described by Penn et al. (1994, p93) are:

- The use and flow of records in the organisation;
- Types of records and record series;
- Classification scheme used;
- Activity level of records;
- Periods covered by records;
- Period in active use;
- Archival / historical value;
- Vital record classification;
- The department with which the record is associated;
- Record origination;
- Record's retention;
- Storage media for record.

Penn et al. (1994, p93) describe the process of creating a records inventory as beginning with interviews or questionnaires to determine the type of records and their use. They also suggest a physical inventory compiled by a team of trained records personnel.

Both of the above approaches require that a company first go through the laborious process of identifying the actual records before starting to structure them. These approaches might have been sufficient before the move to electronic media for records keeping, but in the current organisational environment it is unlikely that any one person in a department will have access to or knowledge of all the records in the department. A classification of all records will first be required before the inventory can be done. This classification should include the following:

- A standardised classification scheme that allows users to store records according to an acceptable records series classification.

- Users' understanding of the classification scheme and commitment to store their records according to the standardised scheme.

The difference in this approach from previous ones concerns the sequence of the steps; most records management implementation starts with creating the records inventory and then determines the classification scheme from the inventory. This approach will require that the classification scheme first be created and that all records are classified, after which the inventory can be easily created.

2.2.3.2 Record classification

A classification scheme should be used to organise the records into records series. This scheme can be compiled either in terms of organisational functions or by using record series classifications per department or a combination of these (Penn et al., 1994, p116-127). A classification scheme that is based on the organisational functions will ensure standardization across the organisation. One way of doing this is to follow a hierarchical classification scheme (Wiggins, 2000, p75):

- *"Level 1: Function within the organisation*
- *Level 2: Activity within function*
- *Level 3: Subject within activity*
- *Level 4: Topic within subject"*

Scupsky (1999, p31) describes what he calls "The functional relational records retention schedule". This approach is based on classifying documents according to business function and sub-function. According to Scupsky (1999, p32) records related to the same sub-functions have similar retention periods. The functions and sub-functions that Scupsky (1999, p32) refers to are the processes and activities of the organisation; he uses as an example the accounting function, with sub-functions: general ledger, capital properties and accounting management information, which are all activities that are executed in the larger financial process of the organisation. This view is also supported in the records continuum approach (Upward, 1996).

A further motivation for employing the activities and functions of an organisation as the basis for the classification scheme is that records are used to document the actions and transactions of an organisation. Linking the records to the activities

furnishes the ability to determine accountability for the actions and records as well as the access requirements of the records (McKemmish et al., 2000).

The classification scheme can also assist with the retrieval of the records within the records series (Wiggins, 2000, p81).

2.2.3.3 Records appraisal

Records appraisal comprises the process of determining the value of records for archival and historical purposes. According to Penn et al. (1994, p107) appraisal should be done at the time of record creation. Appraisal forms the basis for records retention scheduling, which is a critical activity in ensuring that the records are kept according to legal and organisational requirements.

For conducting such an appraisal the function of the records must be taken into consideration; three basic functions are identified by Penn et al. (1994, p107) namely:

- Evidence of transactions;
- Compliance with governmental or professional regulations;
- Reference material.

Records could also fulfil more than one function: an example would be a mortgage contract, which must be kept for evidentiary and compliance reasons.

Appraisal is concerned with determining the value of the record. To do this the primary and secondary values as specified by Penn et al. (1994, p110) and the "rules of worth" as specified by Megill and Schantz (1999, p24-28) need to be considered.

2.2.3.3.1 Primary values

Records are created during the execution of organisational activities and it is this purpose which defines their primary value. This can be divided into three categories, namely:

- Administrative value;
- Fiscal value;
- Legal value.

2.2.3.3.2 Secondary values

The secondary values of records deal with their future use, as opposed to the way they are used during the active phase of their life. These values can be described as evidential or informational.

2.2.3.3.3 Rules of worth

Megill and Schantz (1999, p24-28) define the following five “rules of worth” that may be used for determining the value of documents and records:

1. Information used by important people (senior management and other important people) has higher value;
2. Information takes on value when it is shared;
3. Vital records are part of corporate memory (vital records are those required for organisations to continue operating);
4. Historical information is valuable. The historical information can be used to develop and preserve the corporate culture, it helps in decision-making and assists in determining past trends. Examples of these are minutes of board meetings, annual reports, organisational charts and articles of incorporation;
5. Information required to be retained by law or regulation is valuable for the duration of the specified retention.

Megill and Schantz (1999, p28) also state that the value of information is to a large extent determined by its use and because the way documents are used alters as time passes, the value of the documents will subsequently also alter.

2.2.3.4 Records retention

The main aim of the records retention schedule is to ensure that all records are kept while they are required for administrative, legal, fiscal, reference or evidentiary purposes and that they will be destroyed when they no longer serve any of these purposes. The retention schedule should be applied to the categories or series listed in the master classification plan. Certain events such as audits and litigation may cause a change in the retention requirements of specific records or records series and require a process for ensuring that these records will not be destroyed while the audit or litigation is in process (Penn et al., 1994, p116-127).

2.2.3.5 Protecting vital records

Vital records are those which are required for an organisation to be able to continue functioning as well as records that protect the rights and interests of the organisation and those entities that the organisation deals with. Organisations must ensure that the vital records are properly protected to ensure that they will be available after a disaster occurs; this normally involves catering for a disaster recovery plan, which will entail the preservation of duplicates of the vital records (Penn et al., 1994, p129-131).

2.3 Knowledge Management

The knowledge management literature distinguishes between two categories of knowledge, tacit and explicit (Nonaka, 1991, p4; Grover and Davenport, 2001, p7). Tacit knowledge is knowledge that is not articulated: it is the knowledge that a person possesses of a specific domain, gained through experience and the internalisation of information. Explicit knowledge on the other hand has been codified in some form to allow for its expression. When knowledge is expressed as explicit knowledge, it can be viewed as information that requires interpretation by a person to be once again turned into tacit knowledge (Nonaka, 1991, p4; Grover and Davenport, 2001, p7). When considering documents we are dealing with codified information and data, which places documents in the realm of explicit knowledge. Data can be aggregated and interpreted in a specific way in order to be once again used as a source of information. Explicit knowledge is information that can be interpreted and used in a specific context so as to add to the tacit knowledge of the user of the information (Grover and Davenport, 2001, p6).

In order to consider the management of unstructured content from a knowledge management perspective it is necessary to determine the role of unstructured content in knowledge management. Bellaver and Lusa (2002, p5-11) explain knowledge in terms of six components:

- ***Wisdom:*** Wisdom is the ability to apply the understanding of information during the decision making process.
- ***Understanding:*** Understanding is the ability to interpret and apply information. Understanding is also the ability to understand the data in order to turn it into information.
- ***Information:*** Information is data that is presented in an understandable way to assist in decision-making or communicating ideas.
- ***Data:*** Data are facts that are structured and organised in a specific way.
- ***Facts:*** Facts represent phenomena in codified form which describes the phenomena explicitly.
- ***Phenomena:*** Bellaver and Lusa (2002, p8) describe this as follows: “Phenomena appear as knowledge in perception, as they are perceived”.

Unstructured content can thus be viewed as information or data, which in terms of the above description forms the codified components of knowledge. Tiwana (2002, p37) describes information as actionable knowledge. Consequently, when considering the management of unstructured content from a knowledge management perspective it is the management of explicit knowledge that should be considered.

2.3.1 Knowledge processes

The main objective of knowledge management is to use the knowledge that is available in the organisation to assist in achieving the goals of the organisation and to identify areas where the current knowledge should be expanded so as to enable the organisation to achieve its goals (Zack, 1999, p135).

When considering the management of organisational knowledge the knowledge-information-data model as described by Braganza (2004, p348-350) should be taken into consideration. This model suggests that the data, information and knowledge requirements of an organisation should be specified by first determining the knowledge requirements. The next step is then to determine the information requirements from the knowledge requirements and the data requirements from the information requirements. By looking at knowledge requirements in this way it is immediately evident that the organisational functions, processes and activities determine the knowledge requirements. This implies that the information and data requirements are also dependent on the functions, processes and activities of the organisation.

Grover and Davenport (2001, p12) refer to two processes of knowledge creation and management in organisations; these are deliberate and emergent. Deliberate processes aim to satisfy clearly defined knowledge management strategies while emergent ones are aimed at creating knowledge to support and structure organisational processes and activities. Grover and Davenport (2001, p13) add that knowledge management processes should be integrated with organisational strategy and with the processes and activities of the organisation, in other words the convergence of deliberate and emergent knowledge processes.

Grover and Davenport (2001, p7-8) describe the knowledge process as consisting of three sub processes:

- Knowledge generation: The development and acquisition of knowledge;
- Knowledge codification: Transferring tacit knowledge into explicit knowledge;
- Knowledge transfer: The movement of knowledge from the point of origination to where it can be used.

2.3.2 Knowledge life-cycle and knowledge management activities

Frank and Gardoni (2005, p57) regard the knowledge management life-cycle as consisting of the following activities: Identify, Acquire, Structure, Combine, Share, Distribute, Use, Preserve and Eliminate. These life-cycle components can also be viewed as activities within the knowledge management process. Satyadas et al. (2001, p431-432) define the lifecycle activities as create, capture organize and distribute/share; Davenport et al. (1996, p54) defines knowledge work activities as find, create, package and apply. Shankar and Gupta (2005) describe a knowledge management life cycle consisting of creation organisation, dissemination and analysis and use. Cisco and Strong (1999) depict the activities of the information management value chain as consisting of capture, transform, store, transfer and apply. These activities and the way they are described can be used to illustrate the required activities for managing information from a knowledge management perspective. These activities, although differently named by the different authors, refer to the same set of events that need to occur when managing unstructured content as knowledge. The activities as described by Cisco and Strong (1999) will be used to further describe the individual life-cycle activities since these describe a close relationship with those required for records management and as such will allow for better integration from a model perspective. Davenport et al. (1996, p54) found in their research that in some instances some organisations or departments in the organisation are more focused on one of the knowledge life-cycle activities than others; for example, physicians are expected to apply knowledge but not necessarily to create new knowledge, while pharmaceutical firms might concetrate more on creating new knowledge. This focus on specific knowledge activities may also be found in some departments and processes in an organisation; for example, the finance function might apply knowledge whereas the marketing function might be expected to create new knowledge about customers. For knowledge management to be effective, the knowledge life-cycle activities must be integrated into the daily activities of the employees of the organisation (Jenkins, 2004, p74). In other words,

the knowledge management activities need to be aligned to the process activity's information needs, be these to capture, transform, store, transfer or apply, or combinations of these.

2.3.2.1 Capture

Knowledge capture consists of document creation or acquisition; this is the point where the document enters the organisation. The capture component of knowledge refers to the creation of new knowledge by transforming tacit knowledge into explicit knowledge as well as organising data and information in such a way that it can be interpreted and used. Creation also includes the creation of new explicit knowledge from the combination of existing explicit knowledge (Handzic and Zhou, 2005 p87). To effectively explicate knowledge requires an understanding of the types of knowledge since these will drive the process of determining what knowledge should be explicated. Zack (1999, p46) refers to the following types of knowledge:

- ***Declarative knowledge*** builds a common understanding in the organisation by explicitly describing "things". These "things" can be concepts, categories or descriptors.
- ***Procedural knowledge*** describes how actions are performed and the dependencies of these.
- ***Causal knowledge*** describes the reasons for why things are done in a certain way. This can typically refer to why activities are executed in a specific way.

Knowledge can also be general or specific:

- ***General knowledge*** refers to knowledge that is applicable in most circumstances, which makes this easy to codify. Because general knowledge is not industry or context specific, it is often generally available and shared between organisations.
- ***Specific knowledge*** is specific to certain industries, circumstances or even to the organisation. Specific knowledge is often what sets organisations apart from competitors and is not readily shared between organisations.

The knowledge capture process should identity which knowledge can and should be explicated and whether it is general knowledge that can be acquired or specific

knowledge that must be created. Zack (1999, p47) argues that not all tacit knowledge can be explicated, which means that management must understand the knowledge used in the organisation to determine what can and should be made explicit. According to Handzic and Zhou (2005 p 88) one of the key factors of knowledge creation in organisations is by means of communication and interaction between individuals in the organisation, which points to the value of communities of practice and of a directory that can assist in finding subject experts. Jenkins (2004, p74) states that the information contained in the knowledge repository must be authentic and accurate, which means that a process for validating it must exist.

2.3.2.2 Transform

This activity involves the organisation of information into categories or taxonomies, indexing information for retrieval, appraising it to determine its value and cross linking information. Organising documents into a taxonomy aides in the retrieval of information and assists knowledge workers to know what information is available. The taxonomy also facilitates information reuse (Satyadas et al. 2001, p431-432).

A process for re-evaluating knowledge in the knowledge repository is also required, which includes the re-indexing and re-categorisation of knowledge as time passes to ensure that the knowledge is still valid (Dalkir, 2005, p32). The taxonomy employed for organising the knowledge must also be reviewed to ensure that obsolete categories are removed and new categories are added. This process might require the migration of information to a new category (Zack 1999, p55).

2.3.2.3 Store:

This refers to the following actions involved in storing the information

- Creating an inventory of information;
- Determining and assigning the retention and disposal requirements for the information;
- Storing the information on the correct media type;
- Implementing security requirements for the information such as confidentiality, integrity and availability;
- Adding any required metadata.

2.3.2.4 Transfer

Transfer involves the activities required to make information available to the users of the information. Handzic and Zhou (2005, p92) refer to this as “distributing the right knowledge to the right people at the right time”. This entails enabling users to browse or search for information as well as providing methods to push information to users.

Search engines and knowledge maps may be employed for assisting users to pull information from the knowledge repository (Satyadas et al., 2001, p432). Handzic and Zhou (2005, p94) regard the visibility of knowledge as a key factor in knowledge transfer; to make it visible they suggest the use of knowledge maps and directories containing details of subject experts.

“Knowledge portals, intelligent agents and recommendation systems” are all forms of pull technologies that can be used (Satyadas et al. 2001, p432).

The knowledge transfer activity also includes the identification of the frequency of knowledge delivery with respect to push technologies and information delivery rules such as copyright and the confidentiality of the information (Dalkir p31-32).

2.3.2.5 Apply

The information is used to enable decision-making and to execute activities

Application encompasses all of the applications of knowledge such as (Dalkir, 2005, p32-42):

- Standardisation of routine processes;
- Decision making, predictions and event analysis;
- Training and learning activities;
- Innovation of new products and or new knowledge.

The key aim of this activity is to ensure that a knowledge management environment exists which motivates users to employ the available information in the execution of their daily activities (Handzic and Zhou, 2005, p96). This requires that the knowledge management systems are up to date and that they are user-friendly (Handzic and Zhou, 2005, p97).

2.3.3 Knowledge classification

The knowledge management literature abounds with different knowledge categorisations, classification and indexing approaches. The most common to be found are taxonomies, controlled vocabulary, folksonomies and thesauri. All of these are related in the sense that they attempt to group knowledge according to some structure, although some such as taxonomies adopt a more systematic approach whereas folksonomies are almost totally ad hoc and chaotic. Zack (1999, p56) refers to two case studies where both organisations experienced difficulty in finding a standard classification schema for use in their knowledge repositories, and states that this is not an isolated situation. The use of a standard terminology for organising knowledge is essential for the success of a knowledge repository since the standard terminology is used to index and retrieve knowledge (Zack, 1999, p56). According to Kwasnik (1999, p24) the classification system itself is a means to show the relationships between the classified entities and as such could be used to communicate the knowledge structure of the classified entities. Kwasnik (1999, p24) also states that the classification schema can be used for knowledge representation and knowledge discovery.

To decide which approach an organisation should use requires an understanding of the different knowledge classification and representation approaches. Those which will be addressed in this section are taxonomies, folksonomies, thesauri and controlled vocabulary.

2.3.3.1 Taxonomy

Taxonomy is the science of classification (Bruno and Richmond, 2003, p45) which aims at categorising things. Taxonomies usually make use of a hierarchical structure that depends on a parent-child type relationship between a node and its sub nodes (Bruno and Richmond, 2003, p45-46). Kwasnik (1999, p24-42) describes a number of different approaches to creating taxonomies including hierarchies, trees and facet analysis.

2.3.3.1.1 Hierarchies

Hierarchies are a rigid form of classification where only generic relationships between super and subclasses are allowed (is-a relationship). This means that the top class includes all subclasses and that the subclasses inherit all properties of all

the superclasses above them. Subclasses of a superclass are related through the properties of the superclass but must be distinct in a predefined way; hence a predefined set of rules must exist to associate subclasses with a superclass and to differentiate between subclasses. An entity can only belong to one class and this association with the class must be based on specific predefined criteria.

The strict rules of hierarchies make them ideal for knowledge representation and discovery, but require an in-depth and complete knowledge of a domain to create the classification schema. Some knowledge domains are also better suited to categorisation based on relationships other than the generic relationship, or might need multiple hierarchies for complete representation (Kwasnik, 1999, p24-30).

2.3.3.1.2 Trees

Trees also use superclasses and subclasses but inheritance is not assumed in these classification schemas. Trees allow for relationships other than the strict generic relationship, which makes them more suitable to classification structures such as organisational functions. The rules for creating a tree are less rigid than for hierarchies, but a fairly complete knowledge of the entities is required for specifying the rules that determine the composition of the classes and the criteria for inclusion in a class. A drawback of trees is that, just like hierarchies, trees only allow for categorising information according to a single type of relationship (Kwasnik, 1999, p30-35).

2.3.3.1.3 Facet analysis

Kwasnik (1999, p39-42) describes facet analysis as a combination of multiple trees and / or hierarchies for classifying information. This is based on the concept that an entity can be viewed from more than one perspective. A primary facet is required with optional secondary facets to which an entity can belong.

2.3.3.1.4 Organisational taxonomies

Bruno and Richmond (2003, p48) list five types of taxonomies (figure 2) that are frequently used in organisations, together with their benefits and disadvantages. This indicates that the functional taxonomy is best suited to organisational knowledge organisation; taxonomies such as subject, product and department can then be used as supplemental taxonomies or secondary facets.

Functional: This type of taxonomy organizes itself along the different functions performed by an organization – both administrative and operational.

Pros	Cons
• is most in tune with organizational goals and business processes • reduces silos of information • reduces duplication • makes it easier to find the most recent official document • shows the flow of information • naming of headings is unaffected by department name changes • is recommended by ISO Technical Report 15489	• requires a new way of thinking about information • needs buy-in from everyone • requires one person to oversee the major shared "buckets" • requires a liaison within each department that contributes to the "bucket" • requires more training of employees

Department: This type of taxonomy is department-based and mirrors an organizational chart.

Pros	Cons
• is easy to build • is easy to understand • preserves the chain of command, avoiding internal "politics" • allows an individual to work in only one area of the taxonomy	• requires taxonomy headings to be changed frequently • department mergers and splits will force parallel changes • splits information on a project or topic across the taxonomy if two or more people from different departments contribute information • encourages a proprietary way of thinking • does not represent what an organization actually does • is difficult for new employees to use • requires the management of departing employees' documents/files

Subject: This type of taxonomy is based on the subjects of information with which an organization might deal.

Pros	Cons
• is appealing if need to classify a discrete body of knowledge • allows for greater depth if required • is an excellent application for an EDMS	• is limited to that one body of knowledge • may be difficult to select the terminology for the subject headings if the users of the taxonomy are both novices and experts in the subject field

Product/Services: This type of taxonomy is based on the products and services that the organization provides.

Pros	Cons
• provides good representation of information for product- or service-centered organizations	• is more of a stand-alone taxonomy rather than a way to represent an entire organization

Location: This type of taxonomy is based on the organization's geographic location.

Pros	Cons
• is ideal for large multinational organizations • allows for customization based on location • allows for the incorporation of customs, culture, and regulations that are specific to the location	• is challenging to split information between the corporate office and branch locations • requires specialist in each country to create the taxonomy because of language and cultural nuances • is difficult for centralized control

Figure 2 - Types of organisational taxonomies (Bruno and Richmond, 2003, p48)

Thesauri can be viewed as hierarchical taxonomies with the following added information (Garshol, 2004):

- A description of each class;
- A list that provides a class name for terms that are synonymous with the class;
- Related terms that are not synonymous with the class: for example, one could show that “ontologies” are related to but not synonymous with a class named “topic maps”.

2.3.3.2 Controlled vocabulary

A controlled vocabulary is a predefined list of terms that should be used when classifying knowledge. The controlled vocabulary may be a simple list of terms or it might comprise a list of terms with a description of the term (Garshol, 2004). Controlled vocabularies are required for the construction of taxonomies and thesauri.

2.3.3.3 Folksonomy

Folksonomies are created by allowing users to categorise information by assigning categories or tags to documents, the categories that are used are not pre-determined and it is up to the user to specify his own categories. This can be done as broad tagging where multiple users can assign multiple tags to a content object or narrow tagging where only one or a few users can tag the content object (Tonkin, 2006). Folksonomies do not make use of a hierarchical or tree type taxonomical structure (Hammond et al., 2005) where inheritance of rules or attributes can be inferred. Consequently, folksonomies by their nature are unreliable for creating categories that can be used to apply content management rules, such as information security and records management requirements. They are, however, an excellent additional level of metadata that can be used for information retrieval purposes (Hammond et al., 2005).

2.3.3.4 Knowledge map

Knowledge maps are visual representations of the knowledge landscape in an organisation, designed to assist employees to find and understand explicit and tacit knowledge and the sources of these knowledge (Eppler, 2001, p1; Handzic and Zhou, 2005, pp70). The purpose and benefits of knowledge maps do, however, extend far beyond this simple view and can be summarised as follows:

- Knowledge maps can be utilised to show the relationship between the components of a knowledge domain (Grey, 1999).
- Knowledge maps afford the knowledge user a view of how information flows though the organisation, i.e. where it is created, where it is used and for what purpose it is used (Grey, 1999).
- Knowledge maps can be used as navigational tools for users to locate and interpret knowledge (Grey, 1999; Eppler, 2001, p7)
- The use of knowledge maps can assist in identifying knowledge gaps and lead to the creation of a knowledge or skills acquisition roadmap (Grey, 1999; Eppler, 2001, p6).
- Knowledge maps can be employed to identify knowledge experts as well as knowledge users; this information may then be used to identify users who could be members of communities of practice (Grey, 1999; Handzic and Zhou, 2005, p71)

2.3.3.4.1 Types of knowledge maps

Various different types of knowledge maps may be created, depending on the requirements of the organisation. Handzic and Zhou (2005, p71) describe three categories of knowledge maps: procedural, conceptual and competency maps. Eppler (2001) names five types of knowledge maps: knowledge source maps, knowledge assets maps, knowledge structure maps, knowledge application maps and knowledge development maps. These types of maps could also be combined to offer knowledge users a different viewpoint from which to look at the knowledge. A brief description of some of the types of knowledge maps is given below:

- **Procedural knowledge maps** are maps where the relationships between knowledge and the organisational processes are depicted (Handzic and Zhou, 2005, p79).
- **Conceptual knowledge maps are** constructed in terms of the taxonomy used for categorising the knowledge in the organisation (Handzic and Zhou, 2005, p79).
- **Competency maps** provide a visual representation of the knowledge experts and skills in the organisation (Handzic and Zhou, 2005, p79).

- **Knowledge source maps** are similar to competency maps and are employed to find people who possess knowledge about a specific subject or process. These maps may also be integrated with other maps such as a geographic location or organisational structure (Eppler, 2005, p3).
- **Knowledge asset maps** are used to show the availability of knowledge and skills in the organisation: Eppler (2001, p3) refers to such a map as a "graphic balance sheet of a company's intellectual capital".
- **Knowledge structure maps** give a visual overview of a knowledge domain and illustrate how knowledge and skills are related within the knowledge domain. The knowledge structure map is also made use of to depict the relationship between different parts of a knowledge domain and even between different knowledge domains (Eppler, 2001, p3).
- **Knowledge application maps** depict the knowledge required for the execution of organisational activities and for decision making. These maps also indicate where the required knowledge can be found (Eppler, 2001, p3).
- **Knowledge development maps** offer a visual roadmap for developing specific skills and competencies (Eppler, 2001, p3).

Combinations of these maps may be used to enhance the user's experience when attempting to locate knowledge, as well as to identify knowledge gaps and the actions required to fill these.

Knowledge maps thus offer a visual representation of one or more of the categories used for knowledge classification in the organisation, with the added element of information regarding employees who possess specific skills and knowledge.

2.3.4 Sharing and transferring knowledge

2.3.4.1 Knowledge pull and push strategies

This section will discuss different knowledge management strategies that may be employed to facilitate the sharing and transfer of knowledge in organisations. Knowledge transfer can occur as either knowledge pull, where knowledge users have a need for knowledge and start to search for it, or as knowledge push where knowledge is delivered to knowledge users based on their knowledge preferences (Baird and Henderson, 2001, p85).

2.3.4.1.1 Structuring knowledge to enable knowledge pull

Baird and Henderson (2001, p61-63) describe knowledge push strategies as actions that aim to attract knowledge users to use the knowledge repository or knowledge base. For this to occur, they argue that the knowledge base must be up to date, well maintained and structured in such a way that it allows knowledge retrieval according to the knowledge user's needs. They also regard the structuring of the knowledge base as consisting of an administrative strategy and an active and passive frame.

2.3.4.1.1.1 Administrative strategy

The administrative strategy described by Baird and Henderson (2001, p62-66) consists of using focal units and communities of practice. Such units are groups of individuals who are responsible for creating and maintaining allocated portions of the organisational knowledge. Focal units are used where speed is required in creating and structuring the knowledge base, intellectual property rights and information confidentiality constitute a concern where expert knowledge is required for the construction of the knowledge base. Communities of practice are used where the knowledge is difficult to explicate, knowledge is dispersed amongst a large group and where continued commitment from a large group is required for a long period of time.

2.3.4.1.1.2 Active and passive frames

These frames refer to the extent to which the knowledge is structured for representation. The active frame structures the knowledge in a very specific way according to predefined categories: for example, process maps where the knowledge base is structured according to specific processes and activities within the processes. The passive frame on the other hand possesses minimal structure except for some metadata that might be captured when the knowledge is created or acquired. A typical example of the passive frame is a search engine that is used for information search and retrieval (Baird and Henderson, 2001, p67-74).

2.3.4.1.2 Targeting knowledge users and communities for knowledge push

The organisation might have a need to ensure that specific knowledge gets to the knowledge users in a timely manner. To do this requires the use of knowledge push strategies, which involve matching knowledge users' knowledge requirements with the available knowledge in the knowledge base and then sending it to the users

when it becomes available or when it is expected that they will require it. This requires discovering the knowledge preferences of the knowledge users, in terms of content, packaging and also of when they will require the knowledge (Baird and Henderson, 2001, p85-90).

2.3.4.2 Social network analysis

Social network analysis (SNA) involves mapping the knowledge flows between people in an organisation (Dalkir, 2005, p116). The aim of this is to determine how knowledge is exchanged in the organisation and between whom. This information can then be used to improve the knowledge flow within the organisation and to identify the barriers to knowledge sharing. SNAs are normally constructed by identifying people, teams, organisations and other information/knowledge processing entities, such as computers, as nodes and the information or knowledge exchange as relationships between these nodes. These links can then be weighted and measured to perform mathematical analysis of the patterns of interactions and the number of links between people (Dalkir, 2005, p116-118).

2.3.4.3 Community yellow pages

Community yellow pages are directories that are used by knowledge users to locate knowledge experts. These directories are constructed by one, or a combination, of the following (Dalkir, 2005, p119; Lamont, 2003):

- Analysing the contributions made by employees to document repositories.
- Using questionnaires or interviews to determine who the knowledge experts are.
- Analysing e-mails.

This information is then used to build a knowledge profile of the employee, which he or she can update and publish.

2.3.5 Version control

The purpose of version control is to ensure that users know which version of a document is the latest and to prevent multiple users from editing the same document at the same time and then erasing each other's changes when saving the changes. Version control should also provide an audit history of who made changes to a document and what those changes were (Jenkins, 2004, p92).

2.3.6 Process oriented knowledge management

A number of research projects and articles exist that focus on knowledge management from an organisational process and activity perspective (Remus and Schub, 2003; Davenport et al., 1996; Kim et al., 2003; Kang et al., 2003; Jablonski et al., 2001). The aim of this is to align knowledge management to organisational goals by ensuring that the knowledge required for executing the activities and processes that allows the organisation to achieve its goals is available (Remus and Schub, 2003, p238). Davenport et al. (1996, p54-55) view a process-oriented knowledge management approach as a means to structure knowledge work in order to increase productivity.

Zack (1999, p46) describes three types of organisational explicit knowledge namely:

- Descriptive knowledge: This consists of descriptions of objects used to create a shared understanding, which serves as the basis of knowledge sharing.
- Procedural knowledge: This is used to describe the actions and procedures in the organisation.
- Causal knowledge: This contains the reasons and motivations for the actions performed in the organisation.

A process-orientated perspective on knowledge management strives to categorise knowledge according to the activities and processes where the knowledge is created and used (Jablonski et al., 2001, p80). This ensures that a continuous loop is established whereby knowledge is created, categorised, used and improved upon (Remus and Schub, 2003, p239).

Jablonski et al. (2001, p79-80) define process-oriented knowledge management from a structure and storage perspective. Their approach treats information objects as knowledge carriers and the attributes of these knowledge attributes are called knowledge particles in their terminology. They furthermore define the categorisation of attributes such as processes, products, subjects, authors etcetera as dimensions to which the knowledge carriers can belong. The attributes are thus specific nodes in a taxonomy which makes up the dimension and the knowledge objects (knowledge carriers) are categorised according to the dimension/s to which they belong by assigning specific attributes to them.

Another approach to process-oriented knowledge management can be found in Kang et al.'s (2003) workflow-based knowledge map. This knowledge map extends beyond merely mapping the knowledge to processes. It includes establishing a relationship between the user, his/her designated activities which comprise his/her role and the knowledge associated with the process/activity. These relationships are then used to determine the user's knowledge requirements and the information access requirements. Knowledge in this knowledge map is also categorised according to how the knowledge is created and used, relative to a specific activity. The categories suggested by Kang et al. (2003, p286) are:

- Input Knowledge: Knowledge that is required to executed the particular activity.
- Output Knowledge: Knowledge that is either a direct result or by-product of executing the task
- Applied knowledge: Knowledge that is created by executing the activities which are dependent on the current activity.
- Parallel knowledge: Knowledge that is created in other activities executed in parallel with a specific activity. This means that the execution of the activity resulted in the creation of knowledge in another activity through collaboration.

From the preceding discussion, it should be clear that the process-oriented knowledge management approach is primarily a means of categorising knowledge in order to establish a relationship between knowledge and the organisational processes.

2.4 Information security

Ensuring that corporate information is secure is a critical component of managing Information systems and unstructured content is no exception. Ensuring that information is secure requires ensuring its confidentiality, integrity and availability (Bishop 2004, p1).

2.4.1 Confidentiality

Confidentiality refers to the concealment of data. This means preventing people who are not authorised to possess information, from accessing it or even knowing about its existence (Bishop, 2004 p2). Confidentiality is thus closely related to information sharing and knowledge management in that it represents the security mechanisms used to determine how people gain access to information. Some mechanisms for managing confidentiality are access control, permissions and encryption (Bishop 2004).Access control is one of the primary methods of ensuring confidentiality (Bishop 2004).

2.4.2 Integrity

Integrity deals with protecting data and information against unauthorised changes and modification (Bishop 2004). The aim of integrity is to ensure that information is trustworthy (Bishop 2004) and, as such, that it can be considered to support records management in order to ensure that organisational records can be trusted (Strong 1999; Skupsky 1999; Ekweozor and Theodoulidis 2004). The integrity of information is protected by using a combination of prevention and detection strategies (Bishop 2004). Prevention strategies include assigning permissions and using WORM (write once read many) technologies, whereas detection strategies make use of audit logs and intrusion detection systems (Bishop 2004; Ekweozor and Theodoulidis 2004). The decision regarding which documents should be protected against modification must be specified according to the records and information management requirements.

2.4.3 Availability

Availability deals with ensuring that the data and information are on hand for use when required (Bishop 2004). To protect information and simultaneously to ensure availability requires strategies to guarantee that the information system will not become unavailable as well as information recovery strategies (Bishop 2004).

2.4.4 Mapping security requirements to information

There are many specific technologies which assist in protecting information and computer systems. For the purpose of developing an Information Management model, of which security is but one component, the technologies are, however, less relevant than determining which information sources require protection and what level of protection is required. To do this will require mapping the security requirements to the information.

2.4.4.1 Access control

2.4.4.1.1 Overview

Several access control models exist today, of which the most well known are Mandatory Access Control (MAC), Discretionary access control (DAC), Access control lists (ACL), Role-based access control (RBAC) and Task-Role-based access control (T-RBAC) (Oh and Park, 2003; Essmayr et al., 2004). Both RBAC and T-RBAC models support the identification of tasks or activities in an organisation and assigning these to roles, which are then used for assigning permissions for controlling access to information resources (Oh and Park, 2003; Essmayr et al., 2004). The requirements for using tasks or activities to determine which access a user must possess to information sources is important because it allows a more controllable way of defining access, based on the user's business responsibilities, as opposed to MAC, DAC and ACL. It also facilitates the ability to allow inheritance of permission and the enforcement of separation of duties (Karjalainen et al., 2000; Essmayr et al., 2004). This section will specifically consider RBAC and T-RBAC because of the relationship between organisational roles / tasks and assignment of permission.

2.4.4.1.2 RBAC

RBAC is a model for using roles for managing access permissions to information resources. This is achieved by organising transactions into roles and giving users access to execute these transactions by assigning the users to the same roles. Transaction in this context refers to any read / write action within an information system (Ferraiolo and Kuhn, 1992 p4). Organisational functions and activities are relatively stable over time, which makes the management of access control easier when using roles to map users and permissions (Sandhu et al, 1996 p39).

According to Sandhu et al (1996, p40) a key requirement for role-based access control systems is that both user role membership and the permissions assigned to roles should be easy to determine. In addition, the management of role membership and role permissions must be centralised.

Sandhu et al. (1996) and the NIST model (Sandhu et al. 2002) describe RBAC using a group of models that build on each other to describe RBAC at different levels of sophistication.

These models provide a comprehensive access security model with the following characteristics:

- Permissions are assigned to roles in a many to many relationship.
- Users are assigned to roles in a many to many relationship.
- A user can only gain access to an object if the roles used during the session have the necessary permissions.
- Partial and full inheritance of permissions are allowed in role hierarchies.
- Constraints can be applied on:
    - Role membership
    - Hierarchical role inheritance.
- User-role review.
- Permission-Role review.

The aims of these models are to ensure that the security principles of least privilege and separation of duty are catered for in the model, while providing a dynamic access authentication model for managing access security in enterprises.

2.4.4.1.3 T-RBAC

T-RBAC extends the RBAC model by mapping permission to tasks and then assigning these tasks to roles. The general principle of T-RBAC is that tasks, not roles, should be used to define access rights in order to effectively cater for both tasks that directly relate to business process and tasks of a supervisory nature. Separating tasks from roles provides better support for the implementation of business rules such as separation of duty (Oh and Park 2003, p542, p551-556).

In this model a task is defined as the “fundamental unit of business work or business activity” (Oh and Park 2003, p536). Using tasks to assign access permission, and then assigning these tasks to roles, provides a more granular method to manage access to information resources. It also provides more flexibility in defining and managing roles, because roles are containers for tasks and do not have access rights without the defined tasks.

2.5 Summary

The literature survey demonstrated clearly that a strong relationship exists between the content management requirements of records management and the organisational activities, while a similar but weaker relationship also exists between the knowledge management requirements and organisational activities. The information security requirements are dependent on the relationship between the records management requirements and organisational activities on the one hand and the knowledge management requirements and the organisational activities on the other. In Chapter 3 these relationships, together with the content management requirements, will be used to construct a model for the management of unstructured content.

Chapter 3 - Activity-based Model for Managing Unstructured Content

3.1 Introduction

This chapter will describe a model for managing unstructured content, based on the content management requirements and the relationships between these requirements and organisational activities as described in the literature review. In order to do so, the requirements for management of unstructured content will first be defined, after which a model of the relationships between content users, organisational activities and content objects will be described. The activity-based model for managing unstructured content will then be described and a description will also be furnished of how the content management model enables the management of unstructured content.

3.2 Content Management requirements

3.2.1 Records Management

- Records should be clearly identified.
- The purpose and the value of the record must be specified.
- Records must be organised in a classification system that enables easy classification and retrieval.
- A system must be set in place that can be used to determine the worth of the record according to Megill and Schantz's (1999) rules of worth. This system must also assist in determining the primary and secondary values of the record.
- Records must be retained according to a documented retention schedule.
- The retention period of a record may be influenced by events such as litigation and the model must cater for this.
- A record can have informational value and evidentiary value: the retention period should be defined to consider both values.
- A system for identifying records must exist to ensure that records can be found when required.
- The integrity of records must be protected.

3.2.2 Knowledge management

- Knowledge must be organised in a classification system that enables easy classification and retrieval.
- A retention schedule must exist to ensure that knowledge is retained for the duration that it has value.
- A system must be set in place that allows for the identification of existing knowledge and the identification of knowledge requirements.
- Knowledge must be organised in such a way that it allows for:
    - The identification of the knowledge type as described by Zack (1999);
    - The identification of the business area where the knowledge was created and where it will be used;
    - Providing the right information to the right person;
    - Other knowledge management requirements such as communities of practice, knowledge yellow pages and knowledge maps must be able to benefit from the organisation / classification scheme.
- Knowledge must be protected from unauthorised disclosure.

3.2.3 Information Security

- The confidentiality of documents must be protected by ensuring that:
    - Only authorised personnel have access to confidential information;
    - Knowledge of the existence of confidential documents is only available to those who need to use it.
- The integrity of documents must be protected to ensure that:
    - Documents can be used for evidentiary purposes;
    - Information contained in documents is trusted as being reliable.
- The availability of documents must be protected to ensure that:
    - Records and knowledge are not lost to the organisation;
    - The unavailability of documents does not impact on the operation of the organisation or exert other negative financial influences on the organisation.

3.3 Content Relationship Model

To manage unstructured content requires an understanding of the relationships between it and the attributes related to the management of the content objects. This can be achieved by looking at the relationship between content users' organisational activities and content objects together with their attributes. Figure 2 provides a visual representation of these relationships. These form the basis of the proposed framework for managing unstructured content and will be described in terms of a model consisting of objects, object attributes and relationships.

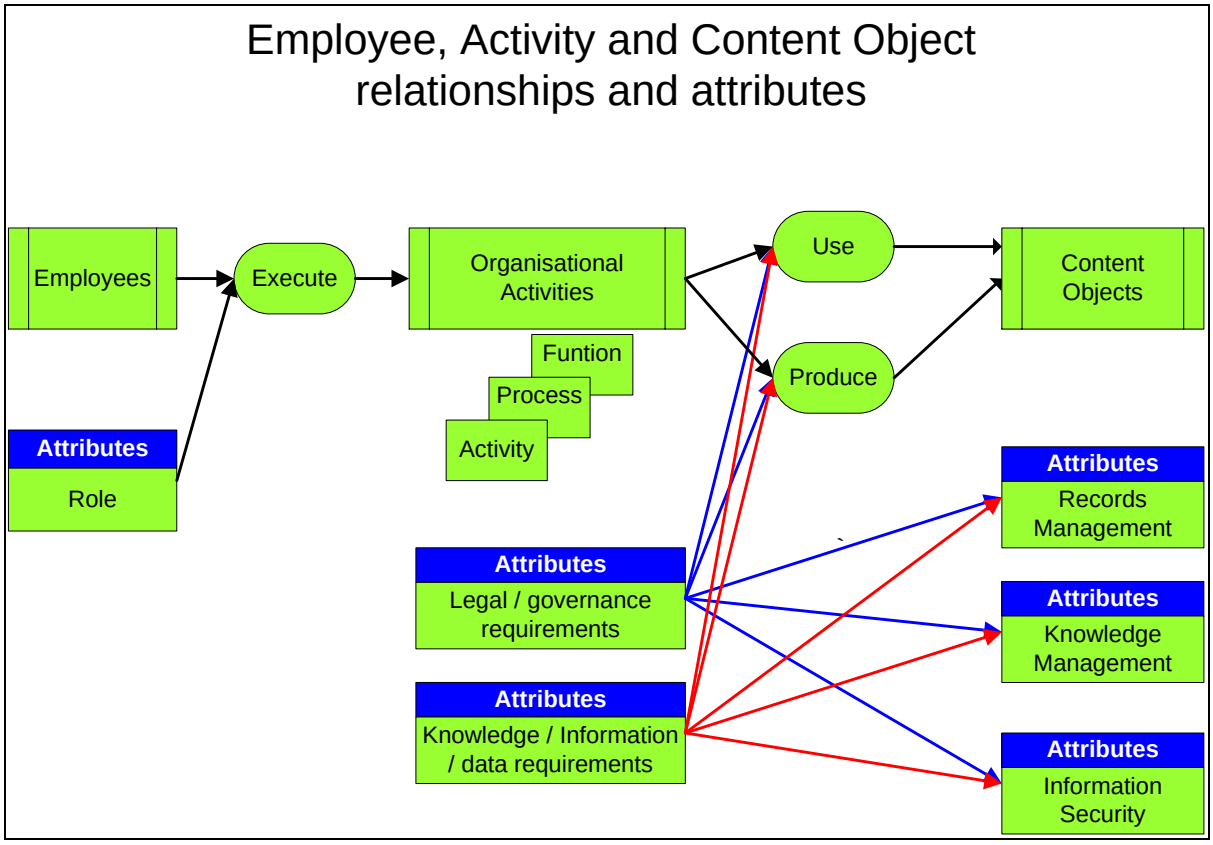

Figure 3 - Relationship model

3.3.1 Objects

3.3.1.1 Employees
These are the content users in the organisation

3.3.1.2 Organisational Activities
Organisation activities refer to the tasks that are carried out in the organisation. These can be viewed as a hierarchical class consisting of functions, processes and activities.

3.3.1.3 Content objects
Content objects refer to all unstructured content in the organisation and include documents, e-mail, images, movie files etcetera.

3.3.2 Object attributes

3.3.2.1 Employee attributes
Role: The role attribute is a specification of the activities executed by the employee. Roles are created by grouping the activities that the employee execute.

3.3.2.2 Activity attributes
Legal requirements and regulatory requirements: These refer to the legal and regulatory requirements for the execution of the activity and may include:

- The creation or acquisition of content objects as proof of the execution of the activity;
- The use of content objects in the execution of the activity;
- The creation or acquisition of content objects as proof of the manner in which an activity was executed.

Knowledge, information and data requirements: An employee might need specific knowledge, information or data to be able to execute an activity. This attribute refers to both the creation of the content objects containing the knowledge, information and data as well as the editing and use of these objects.

3.3.2.3 Content object attributes
Content object attributes refer to specific requirements regarding the management of documents. These attributes can be classified as belonging to one of the following categories:

- Records management;
- Knowledge management;
- Information security.

3.3.2.3.1 Records management attributes

Records management attributes consist of retention period, records series, record authenticity requirements and record integrity requirements.

3.3.2.3.2 Knowledge Management attributes

The knowledge management attributes comprise information sharing requirements, information users, information retention, information confidentiality, information integrity requirements and information availability requirements.

3.3.2.3.3 Information Security attributes

The information security attributes are concerned with the integrity, confidentiality and availability requirements as defined by the records management and knowledge management attributes. Access requirements are derived from the confidentiality, integrity and availability requirements, but are mentioned separately in the model because they form a key component in determining the relationship between employee, activity and content object.

3.3.3 Relationships

The following relationships exist:

3.3.3.1 Activity execution relationships

- A relationship exists between the employee role attribute and the activity, based on the execution of the activity. Therefore the employee executes the activity by using the relationship between the role of the employee and the activities assigned to the role.

3.3.3.2 Content creation and use relationships

- During execution of an activity, content objects can be created or used; this means that a relationship exists between the activity and the content object.
- The attributes of the activity such as the legal and governance requirements might require the creation or use of content objects during execution of the activity. In other words, a relationship exists between the attributes of the activity and the content objects created or used during the activity. The same relationship

also exists between the knowledge, information and data requirements attributes and the content objects.

3.3.3.3 Employee to content object relationship

- The employee has a relationship to the content objects via the activity execution relationship and the content creation and use relationship. Stated differently, the relationship between the role and the activity gives the employee a relationship with the activity while the relationship between the activity and the content objects extends the employee's relationship with the activity to one with the content object.

3.3.3.4 Relationships between attributes

- Legal and governance requirements attribute: This activity attribute is related to the security requirements attributes of the content object, since the legal and governance requirements may lead to specific requirements in terms of the confidentiality, integrity and availability of the content objects.
- Knowledge, information and data requirements attribute: This activity attribute has a relationship with the security requirements attributes of the content object because knowledge, information and data requirements may lead to specific requirements in terms of the confidentiality, integrity and availability of the content objects.

3.4 Content Management Model

The content management model is based on the relationships that are defined in the content relationship model.

A model for managing unstructured content requires, as a primary component, a method of organising the content in such a way that it enables the effective management of the content from a records management, knowledge management and security management perspective. Hence a taxonomy that can serve as the foundation for the management of the content is required. Organisation and categorisation of documents were discussed earlier for both knowledge management and records management. Security management concepts such as access control also require that the documents are organised or grouped in a way that allows an organisation to determine what the security requirements of a collection of documents are.

It should thus be clear that taxonomy lies at the heart of the management of unstructured content. The taxonomy must enable the organisation and classification of the information in such a way that it supports the management requirements of the unstructured content as defined in point 6.1. It is clear from the content relationship model that the organisational activities and the attributes of these activities are critical in determining the management requirements of unstructured content. This is further supported by the earlier literature survey which makes it clear that records management requirements, knowledge management requirements and information security requirements can be determined by using a functional or activity-based taxonomy. This is primarily because information is created or acquired in support of, or as a by-product of, the activities of an organisation (Ferraiolo and Kuhn, 1992; Davenport et al., 1996; Upward, 1996; Scupscy, 1999; McKemmish et al., 2000; Wiggins, 2000; Jablonski et al., 2001; Sandhu et al., 2002; Kang et al., 2003; Kim et al., 2003; Oh and Park, 2003; Remus and Schub, 2003; Bishop, 2004).

The model that will be described employs an activity-based taxonomy for organising the documents and then applies the content management requirements at a taxonomical level.

The model will describe the management of all unstructured content in the flowing sequence:

- Activity based content management model.
    - Classification of unstructured content;
    - Information security;
    - Document attributes.
- Specific focus on records management and how the model assists in this.
- Specific focus on knowledge management and how the model enables the management of knowledge.

3.4.1 Activity based content management model

This model's primary aim is the classification of all documents according to an activity-based taxonomy. The latter is then used to apply the records management, knowledge management and information security requirements of the documents.

3.4.1.1 Activity-based taxonomy

An activity-based taxonomy consists of the functions, processes and activities in the organisation. The taxonomy will place the functions at the highest node, with processes as child nodes of this node, and activities as child nodes of either processes or functions. The taxonomy is not limited to these 3 levels: the 3-levels are rather just a simple view of such a taxonomy. The taxonomy can also be created by using functions>processes>sub-processes>activities; once again activities might be found at function, process or sub-process level. This taxonomy also does not exclude the use of subject or other taxonomies, but these should be implemented as secondary taxonomies that can be used for facet analysis. Documents may be associated with any node in the taxonomy and could belong to multiple nodes at the same time.

Processes and activities may span departments, which mean that departments will not be used as part of the taxonomy. Departments, which more accurately reflect organisational structure, could be used as a secondary taxonomy.

3.4.1.2 Classifying unstructured content

The documents classification process can be viewed as consisting of organising documents by mapping the documents to the relevant taxonomical nodes, regardless of their status as records or knowledge objects. This classification process can be carried out as follows:

- Assign the document to the taxonomical node which indicates the activity that was executed during the creation or acquisition of the document.
- Assign the document to the taxonomical node that indicates the activities where it is or will be used:
    - This may occur at the stage of document creation if the use of the document is known, as in well-defined processes.
    - The assignment to taxonomical nodes can also take place at the time when the document is used.

3.4.1.3 Applying information security requirements

The information security requirements can be determined at taxonomical node level. This means that the confidentiality, integrity and availability requirements are

determined by the activities to which the content objects are mapped. Access control is the primary security mechanism of the proposed model; it serves both as a mechanism to apply security requirements but also as a way of connecting users to the content that they require.

3.4.1.3.1 Access control

Access to documents can be determined at the taxonomical node level by creating a relationship between users and roles and roles and taxonomical nodes. This means that the roles will have access to specific taxonomical nodes; these may be higher or lower level nodes depending on the role. The access could also be one of the different types of access, that is, read, modify, create etcetera. The users will then be assigned to roles depending on the activities that they execute or the processes / functions for which they are responsible. Using roles that are mapped to organisational activities is an effective way of applying access control and also provides for security requirements such as separation of duties and the principle of least access (Sandhu et al., 2002; Oh and Park, 2003),

3.4.1.3.2 Relationship between security requirements and organisational activities.

Information security requirements arise from managing the risk to an organisation in terms of the confidentiality, availability and integrity of information. These risks are determined by determining what the information is used for and what the legal and governance requirements are for managing the information. The value of the information in terms of intellectual capital should also be considered from both an information loss and information dissemination perspective.

Mapping documents to activities makes it easy to determine which organisational activities create and use the information and which legal and governance requirements apply to the documents, thereby determining the information security requirements of these documents.

Knowledge value and the risk of loss owing to the dissemination of knowledge can be managed by ensuring that only those employees who need sensitive knowledge to execute their assigned activities are given access to it.

The need for additional information security measures such as worm technologies or auditing, for protecting the integrity of documents, can then be determined by considering the information security requirements of the documents relating to a specific activity. The same applies for ensuring that backup copies of documents exist so as to ensure the availability of the documents.

3.4.2 Records management

The records management function primarily consists of:

- Classifying records;
- Specifying and managing the retention of records;
- Making records available for use.

Using an activity-based model for managing records will enable this in the following way:

3.4.2.1 Classifying records

The records classification process can be viewed as consisting of determining whether a document is a record or not and of then organising those documents that are records into record series. The model uses the taxonomical nodes to classify documents at activity, process or functional level. This includes all documents, regardless if they are organisational records or not. The records appraisal process is applied to determine whether the document is a record or not.

3.4.2.1.1 Records Appraisal

Records appraisal comprises the process of determining whether a document is a record or not and of establishing the value and worth of the records.

Records selection: The records appraisal processes needs to consider the activities where documents are created or used and to determine from that if the documents meet any of the following criteria:

- The documents are evidence of transactions.
- The documents are proof of compliance with governmental or professional regulations.

Since the above are activity related, they can be undertaken for all documents associated with a specific activity.

Record value: The value of the record is also determined by the activity where it is created or used.

This means that the appraisal should be carried out at the taxonomical node level and that all documents assigned to a specific node will have the same value as records.

3.4.2.2 Records Retention

Records retention includes the following activities:

- Specify the retention time for the record.
- Ensure that the record is available during the specified retention time.
- Ensure that the record's integrity and authenticity is protected for the time that it must be retained.
- Ensure that a process exists which allows for the identification of records for which the retention time must be adjusted, as the result of specific events such as litigation.
- Destroy or archive the record when it is no longer required.

The retention time is determined by the legal or compliance requirements relating to the specific activity where the record was created or used. This means that the retention periods must be specified at the taxonomical node level and that all documents assigned to that node will be accorded the same retention periods.

All documents related to a specific taxonomical node will also display the same requirements in terms of protecting the integrity and authenticity of the records, which relates to their information security requirements.

Events such as litigation can be viewed as an organisational activity since a specific set of activities occurs when litigation is in progress. In terms of the model, the records will be assigned to the taxonomy nodes of the activities, thereby ensuring that the retention requirements of these activities will also apply to the relevant records.

The destruction or archival of records after the expiry of the retention period will depend on the records' value for reference purposes: this will be discussed as part of the knowledge management component of the model.

3.4.2.3 Records use

Making records available for use will require that the records can be retrieved and that access to records is managed. The activity-based taxonomy may be used to locate and retrieve records. The relationship between the taxonomy, user roles and user can be used to determine who uses the records and for what purposes.

3.4.3 Knowledge management

The knowledge management activities that relate to the management of unstructured content are those of capture, transform, store, transfer and apply.

3.4.3.1 Capture

This is the entry point of knowledge into the organisation; the model assists this by identifying knowledge requirements on a per activity, process or function level.

3.4.3.2 Transform

Documents are categorised using the activity-based taxonomy. This taxonomy may then be used to retrieve the knowledge for a specific function, process or activity.

3.4.3.3 Store

- The taxonomical organisation of the documents could be viewed as an inventory of the information required to execute and manage the organisational activities.
- In order to define the retention requirements of knowledge, the value of the knowledge should be considered as well as the duration of this value. To undertake this, Megill and Schantz's (1999, p24-28) rules of worth suggest that information used or created by important people is important and could even have historical value. This means that the users of knowledge should be considered in determining the retention of the knowledge. The knowledge is, however used to execute specific activities in the organisation and as such, the duration for which it will be valid for this purpose should also be considered. The retention of knowledge used for executing activities requires that the value that the knowledge contains for the execution of an activity needs to be constantly evaluated on a per-activity basis. Knowledge that contains no value for the execution of the activity should be disposed of or archived.
- Storing the information on a specific media type relates to the integrity and availability of the information; these are security requirements. Since the information is employed in the execution of organisational activities, information relating to the same activity will be accorded the same security requirements.

3.4.3.4 Transfer

- Assigning access to information is an information security requirement. Access control in this respect is dependent on who will create the knowledge and who will use it. Since it is created and used during the execution of organisational activities, the access required will be determined by the relationship between users and the activities that they perform.
- The relationship between the activities and the users can be used to identify users who will require knowledge regarding specific activities, which can then be used for knowledge push initiatives.
- Users can browse the knowledge in the organisation by making use of the taxonomy. This will allow easy identification of knowledge that relates to specific functions, processes and activities. The activity role-based access control can be utilised to present only those taxonomical nodes to which the user is given access, thereby assisting his/her searches.

3.4.3.5 Apply

Organising the knowledge according to the organisational functions, processes and activities ensures that users can find and apply the knowledge that they require for the execution of these activities.

3.5 Summary

The proposed model is based on the relationships between the management requirements of unstructured content and the organisational activities. This chapter explained the nature of these relationships and also showed how an activity based classification of unstructured content may be used as the basis for satisfying the records management, knowledge management and information security requirements of unstructured content. The applicability of using the proposed model for managing unstructured content will be evaluated in the next chapter.

Chapter 4 - Model Evaluation

4.1 Introduction

To evaluate the value of the model proposed in Chapter 3 requires an analysis of the requirements for managing unstructured contents. The model then needs to be evaluated to determine if it can deliver these requirements. This will be done applying the following approach:

- Using the earlier literature survey to identify the requirements for content management; this will be done from a stakeholder perspective. A stakeholder can be viewed as any person or group of persons with specific content management requirements.
- Determining how the proposed content management model can be used to deliver the identified content management requirements.

4.2 Evaluating stakeholder requirements

The unstructured content is created, updated and used by the employees in an organisation in support of or as a result of the execution of organisational activities. In some instances content that is created or updated, needs to be approved before it can be added to the content repository. To facilitate this requires another stakeholder, who is a content approver. Because unstructured content can be used as organisational records or explicit knowledge, the roles of records manager and knowledge manager are also required as stakeholders.

4.2.1 Requirements common to all stake holders

Access control and a taxonomy for categorising information are required by all stakeholders in a content management environment. These requirements will be discussed first and then a more detailed requirement analysis will be undertaken for each stakeholder.

Stakeholder requirements

a) A classification system to which the employees can relate and understand is required to classify all unstructured content.
b) A process for controlling employee access to content objects and the classification system is also required.

Model evaluation

a) The primary classification system specified in the model is an activity-based taxonomy which can be constructed by analysing the functions, processes, and activities in the organisation. These functions, processes and activities will then be used to create the activity-based taxonomy. Activities could reside below a function for activities that are directly related to the function itself: an example of these might be strategic planning activities. The activity may also reside below a process for process-oriented activities. Content objects can be classified as belonging to any node in this taxonomy. Some examples follow: policies related to a function will be linked to the taxonomy node representing that function. Documents explaining a process will be linked to the taxonomical node representing the processes while documents that are created or used when executing a specific activity will be linked to that activity.

b) The model proposes that an activity-based access control system be used. This can be implemented by assigning users to roles, adding activities that can be executed to the roles and then linking content objects to the activities by employing the activity-based taxonomy. Doing this will provide the following:

- Access in order to change the taxonomy can be controlled on role level, meaning that specific roles can be given access in order to make changes to specific taxonomical nodes. These can be top level or lower level nodes, depending on the access accorded to the role. Using the roles to control access to the taxonomy means that some users can be given access to change or update certain portions of the taxonomy. The role-based access can also be utilised to ensure that users can only view the specific taxonomical nodes to which they require access, thereby reducing the complexity of the classification system to which they are exposed.
- Access to the content objects may be controlled by specifying the content object access, according to the activities which are added to the roles to which the user belongs. Hence the content object access is managed on taxonomical node level.

In other words, both access to the taxonomy itself and access to the content object are managed on a taxonomical node level when employing activity-based access control and an activity-based classification system. There will, however,

be a need to specify the access to the taxonomy and the access to the content objects separately, because some users might be given access to create or modify content objects, but not to change the taxonomy.

4.2.2 Content creator

Content creators are responsible for acquiring or creating new organisational content in the organisation. The records management and knowledge management activities that will be executed by the content creator are:

4.2.2.1 Records management specific requirements

Stakeholder requirements

a) Content creation/acquisition involves the identification of the records that need to be kept and the identification of knowledge that needs to be explicated. Records identification involves the records appraisal activity, which means that the content creator needs to determine if the content possesses administrative, legal, financial, evidential or informational value. To do this requires an understanding of the activity where the content was created and of the legislation, organisational policies and corporate governance requirements relating to the activity.

Model evaluation

a) Newly created or acquired content is assigned to the taxonomical node representing the activity where it was created. The value of the content is then determined on the taxonomical node level, based on records selection rules specified by the records manager for each taxonomical node. The benefit of this is that the content creator does not require any knowledge of the records selection process, but only of the classification system.

4.2.2.2 Knowledge management specific requirements

Stakeholder requirements

a) Content creation/acquisition involves identifying knowledge gaps and the creation of new, explicit knowledge to fill these gaps. For this purpose, the content creator

will require a method to evaluate the knowledge that currently exists in the organisation against the knowledge requirements of the organisation.

b) The content creator will require access to current knowledge to be able to re-use existing knowledge and to ensure that knowledge is not unnecessarily duplicated.
c) A means of identifying people with tacit knowledge that needs to be explicated to fill the knowledge gaps is required.

Model evaluation

a) The activity-based taxonomy will allow the content creator to evaluate the knowledge related to specific activities, processes or functions on the activity, process or functional level. Secondary taxonomies and metadata used for faceted classification can then be used to determine the knowledge requirements for other dimensions such as specific knowledge subjects on a per activity basis.
b) Access is controlled by the relationship between activity and content and activity and user roles. This means that the user will be able to access the content related to the activities, processes and functions that are linked to the roles to which the user is assigned. Where this level of access is insufficient, the content creator will be given access to knowledge yellow pages in order to locate people with the required knowledge or who have access to required knowledge.
c) Knowledge yellow pages will be required to locate knowledge experts and holders of tacit knowledge so as to assist a content creator to create new content that is currently only in tacit form. The model supports this by creating a relationship between the taxonomy and the content creators, which may then be used to identify those users who create, edit or approve content.

4.2.2.3 Stakeholder requirements for both records management and knowledge management

Stakeholder requirements

a) The content creator needs to categorise the new content according to the organisational taxonomy. This requires an understanding of the classification scheme.
b) Additional metadata needs to be captured for the content; this could include the following:

- Author or acquirer details such as name, location, department etc.;
- Product to which the content is related;
- Client linking information for content related to a client of the organisation;
- Supplier details when the content concerns records such as invoices;
- Asset information for content that concerns the management of assets in the organisation;
- If the content contains an informational or knowledge component then subject related classification might be necessary.

c) Created or acquired content needs to be stored in the content repository; the content creator will require access rights that allows this.

Model evaluation

a) Employees can easily be trained to identify the taxonomical nodes relating to the activities that they execute.
b) Author/acquirer details could be captured by the content creator or may be derived from the security credentials that the content creator used when storing the content. Product, client, supplier and asset information can be captured by specifying required metadata on a per activity basis. This means that activities related to asset acquisition will typically require asset detail information. The same principle applies for activities related to dealing with vendors, such as the creation of a purchase order, and also for product and client related activities.
c) Saving new content to the content repository is related to the access security that is applied to the documents; this is done on the taxonomical node level according to the relationship between the activity and the user roles.

4.2.3 Content Editor

The content editor needs to edit the content currently in the repository; this may take the form of updating documents, or of changes to classification related metadata. In order to do this it will be necessary that the content editor can search for and retrieve current documents and be given access to edit the contents of the documents and the classification related metadata.

Stakeholder requirements

a) The content editor will need to be able to find and retrieve the documents that need to be edited.
b) Version control needs to be in place to ensure that changes to the document are made in a controlled manner; this will call for the following:
    - A process of controlling who is editing the document to prevent multiple updates to the same document at the same time.
    - An audit process to track the changes to documents and who made those changes.
    - The ability to identify the latest version and all previous versions of a document.
c) The content editor needs to be able to update the document with new classification metadata. This may involve assigning the document to additional nodes in the activity-based taxonomy or assigning the document to nodes in secondary taxonomies that are used as part of a facet classification approach.
d) Access security is necessary to ensure that only authorised people are allowed to make changes to documents.

Model Evaluation

a) The finding of documents is supported in the model by navigating or searching the taxonomy. This allows users a known starting point for searching for documents, unlike keyword searches, which require an understanding of the content of a document before searching. The model does not exclude the use of secondary taxonomies or keyword indexing, which could be used in conjunction with the taxonomy when searching for documents, thereby yielding more accurate search results.

    Retrieval of documents is related to the access security applied on the taxonomical node level.
b) Versioning is supported as follows:
    - The proposed model does not prevent multiple users from updating the same document at the same time. This is a technological requirement that is already part of most content management products.
    - Auditing changes to documents is supported in the following manner:

  - The relationship between documents and activities allows for the application of auditing rules on the taxonomical node level. This means that the auditing requirements for documents can be assessed and applied on a per activity basis.
  - The access security, which is core to the proposed model, can be used to extract information regarding the user who altered a document.
  - Detailed information concerning what was changed is a technological requirement and is not catered for in the model.

- Two components are necessary for finding the various versions of a document and determining the correct version or the version for which the user is looking. The first is a versioning or numbering system while the second is the ability to locate all the different versions. The first of these requirements is not addressed by this model. The second component is supported by the model through the ability to link a document to multiple nodes in the taxonomy, thereby eliminating the need to duplicate the same version of a document. This means that a document and its related versions should be linked to the same taxonomical nodes.

c) Updating the taxonomy requires access permissions to alter the document's metadata and an understanding of the taxonomy. Both aspects were discussed earlier. Additional metadata is supported by the model.

d) Access is managed via the relationship between activities and taxonomical nodes.

4.2.4 Content user

The content user uses documents that are in the content repository. These can be either organisational records or information that is needed for the execution of organisational activities.

Stakeholder requirements

a) The content user must be able to locate and retrieve documents required for the execution of the organisational activities for which s/he is responsible. To assist the content user in finding document the following are also needed:

- Suggestions on how the user can narrow or refine the search;

- The user might require an environment that is personalised to his/her information requirements.

b) Information regarding the use of the document must be stored for auditing purposes.

c) The user might want information about others with the same information needs in order to be able to locate people who might possess knowledge that s/he requires.

Model evaluation

a) Locating content objects refers to the ability to search for and find information; the model supports this by making use of the activities in the taxonomy as search terms or by browsing the taxonomy.
   - Searches can be refined by showing the users to which activities the documents that are returned by a search belong, thereby affording the user the opportunity to search in other activities containing similar information. Because secondary taxonomies are allowed, the secondary taxonomical nodes to which documents belong can also be displayed to assist the user in narrowing or expanding his/her searches.
   - Personalisation can be achieved by only showing the user those taxonomical nodes to which s/he has access. This function could also be used to push new information to where it is required.

b) Auditing is supported by recording the user's security credentials that were used to access the documents.

c) Information regarding who uses what information can be deduced from a combination of showing which users have access to which taxonomical nodes, based on the relationship between activities and user roles, and from the document auditing information.

4.2.5 Content approver

The content approver needs to validate new content and updates to content in order to ensure that documents meet specific requirements in terms of records or knowledge management. It is possible that only some of the content in the content repository needs approval.

Stakeholder Requirements

a) A process is necessary to determine what types of documents need to be approved.
b) A process is also needed for ensuring that content which requires approval goes through an approval process.

Model Evaluation

a) The model does not support the content approval process, since documents that require approval and those that do not require approval can be created or used in the same organisational activity. Catering for this requirement will call for introducing additional metadata which supports the identification of documents that require approval. The required metadata will depend on the organisational rules that determine which documents need approval.
b) The processes required for document approval do not form part of the proposed content management model. It is, however, possible to specify the content approver for a specific taxonomical node.

4.2.6 Records manager

The records manager is responsible for the overall process of managing organisational records according to legal and organisational governance requirements.

Stakeholder requirements

a) The records manager must provide users with guidance on identifying and selecting records.
b) The records manager must specify the retention periods of all organisational records. This requires organising the records into a classification system that groups records with similar retention times together.
c) Events such as litigation, changes in legislation and changes in organisational policy can result in changes to the retention times of records. A process for updating the retention times is therefore called for. This could be located on the record series level when legislation alters or on the individual record level when litigation is in progress.

d) International organisations could stipulate different retention requirements for records in different countries owing to differences in legislation. A process for adhering to these different retention periods must be devised.

e) A process is necessary for archiving or destroying older versions of documents, in terms of the records management or knowledge management value of the document versions.

f) The records manager must be able to determine the value of records on the record series level, which means that a way to create records series that corresponds with the value of records is essential. Records value should be determined by taking primary and secondary value into consideration and applying the rules of worth as described by Megill and Schantz (1999, p24-28).

g) The records manager needs to be able to find records when these are necessary for discovery, due to legislation or other investigations. This might involve tracing all records pertaining to communications or transactions between certain entities or finding documents that describe certain processes or policies. Hence the discovery is not necessarily based on the organisational activities.

h) The records manager needs to keep a records inventory containing the following information:

- The use and flow of records in the organisation;
- Types of records and record series;
- Classification scheme used;
- Activity level of records;
- Periods covered by records;
- Period in active use;
- Archival / historical value;
- Classification of vital records;
- The department with which the record is associated;
- Record’s origination;
- Record’s retention;
- Storage media for records.

Model Evaluation

a) Employing the proposed model, the records manager can specify the records management requirements on taxonomical node level, which means that the same records management requirements will apply to all documents that are classified as belonging to that node. This removes the need to train the content creators, editors and users.
b) The records retention time is specified on the taxonomical node level, meaning that all documents belonging to a specific node will have the same retention time.
c) Changes in retention periods are caused by specific events:
    - Litigation and other investigations: When these are in progress the documents will be used by the activities involved in litigation or investigation, which means that the document classification will be updated to indicate that the document now also belongs to the new taxonomical nodes representing these activities. The retention periods of the litigation related activities will then also apply to the documents. Rules will be required to specify which retention periods apply when the document is linked to multiple taxonomical nodes that have different retention requirements. It is possible that the classification entry linking the record to a specific taxonomical node may be removed when the retention time for that node is reached, while the classification entries are retained for those taxonomical nodes where retention time still requires the record to be stored.
    - Changes in legislation or corporate policy. These changes can be applied on taxonomical node level, which means that the retention requirements of all documents linked to a taxonomical node will be updated.
d) Catering for different retention times, owing to changes in legislation affecting different countries, will necessitate that information is available in the form of document metadata, regarding the country where the documents are created or used. Retention can then be specified by using the combination of the country information and the taxonomical node. Country information could be derived by using the relationship between the content users and the taxonomical nodes, provided that country information is available for the users. In other words, the users of the content determine the country whose legal retention requirements must be adhered to.

e) Determining which document versions can be destroyed or archived depends on the legal and regulatory requirements and the knowledge requirements of the activity where the documents are created or used. This information could be added as an attribute of the taxonomical nodes, which means that the rules for archiving and destroying documents are applied on taxonomical node level. These rules then apply to all content linked to a specific taxonomical node.

f) The value of records can be determined by analysing the activities represented by the taxonomical nodes as well as the users who created, modified or used the records.

g) The model does not fully cater for records discovery requirements, because not all record discovery initiatives will be based on the activities where records are created or used. The introduction of additional taxonomies or metadata for classifying the information as well as keyword indexing might be required for records discovery.

h) The information that is required to compile a records inventory can be obtained from the following areas in the model:

- The activity-based taxonomy, together with access security that allows for a distinction between content creators, editors and users, can be used to determine where documents originated and where documents are used. This could then be employed as a starting point for determining the flow of records.
- Activity level of records. Auditing that is necessary for ensuring the integrity and confidentiality of the content can be used to determine the activity level of the records.
- The classification scheme used is the activity-based taxonomy.
- Periods covered by the records. This information is a combination of the creation date of a document and the retention times specified for the document.
- Period in active use can once again be determined from the auditing information. This could also be specified on taxonomical node level as part of the retention and archiving requirements for a specific activity, process or function.
- The historical value of a document is determined by appraising the record's value, which in the proposed model, occurs on taxonomical node level. This is

also part of managing the document as containing explicit knowledge or information.

- Vital records are those documents that are required for the continued functioning of an organisation. Identifying them is based on identifying those activities where the availability of records or information iscritical to the functioning of the organisation. This can be undertaken on the taxonomical node level.
- The department with which the record is associated could be determined by assessing which departments execute the activities to which the record is linked and also establishing which users created, edited or used the records and to which departments they belong. The department could also be used as a secondary taxonomy; the indexing can then be carried out using the above mentioned relationships between the document, activities, users and departments.
- The record originator is not necessarily the content creator; for example, an invoice received by the accounting department. While the content creator captured the document, it originated in another organisation. Originator information has to be captured as additional metadata and is not supported by activity-based classification.
- Retention is specified on taxonomical node level, which means that all records belonging to a specific node will have the same retention times.
- The storage media on which documents are stored are determined by a combination of the record's retention requirements, the integrity, confidentiality and availability requirements or the document and the organisational performance required for the activity where the documents are created or used. All this information can be gathered from the attributes of the objects in the model;,consequently the model can be used for specifying the required storage media.

4.2.7 Knowledge manager

The knowledge manager must ensure that the knowledge required for achieving the organisational goals is available. This requires structuring this knowledge and creating processes that aid in the creation, distribution and use of knowledge.

Stakeholder requirements

a) A classification system for structuring the organisational knowledge is necessary.
b) A process to identify the organisational knowledge requirements is essential.
c) The knowledge manager needs to specify the retention requirements for explicit knowledge so as to ensure that old knowledge is removed when it does not have any remaining value and also to ensure that essential knowledge is not deleted.
d) A process for determining the security requirements for information must be devised in order to ensure that critical or sensitive information and knowledge is treated as such. The integrity requirements of the knowledge must be managed to ensure that users can trust the knowledge in the repository. Availability is important to ensure that knowledge which is critical to the execution of organisational activities is available when it is required.
e) The knowledge manager needs to ensure that critical knowledge and information reach the people who must use or be aware of the knowledge and information, when they need it or when it becomes available. Therefore knowledge push strategies that can be customised according to user, group or organisational requirements need to be implemented.
f) Knowledge communities and yellow pages of knowledge experts are required to improve knowledge creation and transfer.

Model evaluation

a) The proposed model is based on using an activity-based taxonomy to classify and structure all organisational content. This taxonomy can also be utilised as the basis for creating a knowledge map that shows the explicit knowledge of the organisation on a functional, process or activity level. When combined with additional taxonomies, as would be the case in facet analysis based classification, the activity-based taxonomy could be used to determine the facets that are applicable for specific functions, processes or activities. Hence, when a subject-based taxonomy is used as an additional facet for classification an organisation could determine which knowledge subjects are relevant for which functions, processes or activities.
b) The activity-based taxonomy provides an organisational perspective on the available knowledge per function, process and activity. These same functions,

processes and activities can be analysed in order to determine the organisational knowledge requirements which may then be compared to the available knowledge to determine where knowledge gaps exist.

c) Knowledge retention is based on the length of time the knowledge is of value to the organisation, which means that documents containing explicit knowledge and information are needed if the knowledge is used for the execution of organisational activities or if the knowledge possesses historical value. To determine whether the knowledge is still being used, the following is necessary:
   - The document must be versioned; this could then be applied to allow the archiving or deletion of older versions of a document. The model does not cater for this, although versioning can be viewed as a standard component of most document management systems.
   - The knowledge contained in documents must be reviewed so as to assess the value of the knowledge. The model supports this by enabling a per activity evaluation of documents; if the document has no value for the execution of an activity its classification can be modified so that it is not assigned to that activity. Once the document is not assigned to any activity, it will have no value as a source of process related knowledge.
   - The knowledge security requirements can be determined from the organisational activities where the knowledge is created and used. This is true for the confidentiality, integrity and availability requirements in the following way:
     - Confidentiality: The activities where the knowledge objects are created or used determine the content of the knowledge in the documents, which means that the confidentiality requirements can be determined on the activity level. Examples of this include HR processes and activities that create and use confidential employee information, CRM processes and activities that deal with confidential client information and strategic planning processes and activities that deal with information and knowledge about the future and plans of the organisation. Activity role-based access security can be applied to ensure that confidential information can only be accessed by those employees who need the knowledge for executing organisational activities.

- Integrity: The amount of accuracy required for the execution of an activity will determine the integrity requirements for the knowledge used in the execution of that activity. Integrity can then be managed by identifying who should create new or modify existing knowledge related to a specific activity. A process for reviewing new and modified knowledge can be implemented by assigning a content approver on a per activity basis.
- Availability: Knowledge availability can be determined by assessing the role that the knowledge plays in the execution of the activity; if the knowledge is a prerequisite for executing the activity and the activity is crucial in ensuring the smooth operation of the organisation the knowledge will have a high availability requirement.

d) The relationship between content, organisational activities and the content users, as described in the content relationship model, can be used to determine which users should receive specific knowledge. Once new information becomes available, the identified content users may then be notified or the content may be distributed to these users. Information search strategies can also be designed to ensure that knowledge users can find the knowledge related to a specific activity when they need it.

e) Creating knowledge yellow pages and communities of practice requires that knowledge experts and knowledge users be identified. This may be carried out by using the following relationships:

- The relationship between content creators or approvers and the organisational activities can be used to identify the knowledge experts on a per activity basis. When secondary taxonomies are employed, the relationship between content objects and the content creators or approvers can be utilised to identify the knowledge experts according to the secondary classification scheme: as an example, when a subject-based taxonomy is used. This presents a dynamic way of identifying knowledge experts according to the taxonomies that are used in the organisation.
- The relationships between all content users (creators, modifiers, users, and approver) and the content objects, in conjunction with the relationship between the content objects and the taxonomy or taxonomies that are in use can be emlpoyed to identify possible members of communities of practice.

4.3 Summary

In this chapter, the model that was proposed in Chapter 3 was evaluated to determine whether it could satisfy the needs of the organisational content users and content managers. It was shown that the model does satisfy most of the listed requirements and that the areas where the model is lacking could be bridged by the introduction of additional taxonomies, or by using the technological features already available in most content management systems. This evaluation does, however, need to be expanded to include implementation in an actual organisational environment to determine any additional gaps in the abilities of the model, so as to satisfy the management requirements for unstructured content. A real world implementation is also required to measure social acceptance and challenges of the proposed approach to managing unstructured content.

Chapter 5 - Conclusion and further research

5.1 Introduction

This chapter will re-evaluate the research problem and show how this research contributed to the resolution of this problem, by examining the stated research question and the sub questions and then answering them in the context of the research conducted in the previous chapter. The limitations of the model will also be discussed and recommendations for further research will be provided.

5.2 Re-evaluating the research problem

The aim of this study was to determine the applicability of using an activity-based model to satisfy the management requirements of unstructured content for the purposes of records management, knowledge management and information security. To determine this, the following questions were formulated (see section 1.1.1.1):

- What are the requirements for managing unstructured content?
- What is the relationship between organisational activities and the management requirements for unstructured content?
- How does an activity-based taxonomy provide for the management requirements of unstructured content?

5.2.1 What are the requirements for managing unstructured content?

The requirements for managing unstructured content are listed in Chapter 3 section 0 and, in more detail, as stakeholder requirements in Chapter 4. These can be summarised as follows:

5.2.1.1 Records management

- Records should be clearly identified;
- The purpose and the value of the record must be specified;
- Records must be organised in a classification system that enables easy classification and retrieval;
- A system must be in place that can be used to determine the worth of the record according to Megill and Schantz's (1999) rules of worth. This system must also assist in determining the primary and secondary values of the record;
- Records must be retained according to a documented retention schedule;

- The retention period of a record may be influenced by events such as litigation and the model must cater for this;
- A record can have informational value and evidentiary value: the retention period should be defined to consider both values;
- A system for identifying records must exist to ensure that records can be found when required;
- The integrity of records must be protected.

5.2.1.2 Knowledge Management

- Knowledge must be organised in a classification system that enables easy classification and retrieval;
- A retention schedule must exist to ensure that knowledge is retained for the duration of the period that it has value;
- A system must be in place that allows for the identification of existing knowledge and the identification of knowledge requirements;
- Knowledge must be organised in such a way that it allows for:
    - The identification of the knowledge type;
    - The identification of the business area where the knowledge was created and where it will be used;
    - Providing the right information to the right person;
    - Other knowledge management requirements such as communities of practice, knowledge yellow pages and knowledge maps must be able to benefit from the organisation / classification scheme;
- Knowledge must be protected from unauthorised disclosure.

5.2.1.3 Information Security

- The confidentiality of documents must be protected by ensuring that:
    - Only authorised personnel are given access to confidential information;
    - Knowledge of the existence of confidential documents is only available to those who need to use it.
- The integrity of documents must be protected to ensure that:
    - Documents can be used for evidentiary purposes;
    - Information contained in documents is trusted as being reliable.
- The availability of documents must be protected to ensure that:
    - Records and knowledge are not lost to the organisation;

- The unavailability of documents does not impact on the operation of the organisation or lead to other negative financial effects on the organisation.

5.2.2 What is the relationship between organisational activities and the management requirements for unstructured content?

The organisational activity where unstructured content is created or used determines most of the management requirements of the unstructured content. It specifically affects the areas of records management, knowledge management and information security in the following ways:

5.2.2.1 Records management

- Identifying and apprising records requires that the legal and evidentiary value of documents be considered. The activities where the documents are created or used will determine whether the documents are required to be managed as organisational records for legal or governance requirements (see sections 3.3.2.2, 3.3.3.4, 3.4.2.1).

- The records management requirements with regard to retention of records is determined by taking into account the legal or governance requirements imposed on the activity (see sections 3.3.3.4, 3.4.2.2, 4.2.6).

5.2.2.2 Knowledge management

- Identify organisational knowledge requirements: Knowledge is required to execute organisational activities, which means that a relationship exists between the activity and the knowledge requirements of the organisation (see sections 3.4.2.2, 4.2.7).

- Create or acquire knowledge required by the organisation: This is determined by the organisational knowledge requirements. Knowledge can also be created as a result of executing organisational activities (see section 3.3.3.2).

- Transforming knowledge: Not related to activities.

- Storing knowledge: Not related to activities.

- Transferring knowledge: Knowledge transfer requires that users possess the knowledge required to execute their assigned activities. The relationship here is between the employee who requires knowledge and the activity that needs to be

executed, as well as between the activity and the knowledge related to that activity (see sections 3.3.3.3, 3.4.3).

- Knowledge application: Knowledge is applied during the execution of organisational activities (see section 3.3.3.2).

5.2.2.3 Information Security

Information security displays several relationships with organisational activities:

- Users require access in order to retrieve or create documents during the execution of the assigned tasks. This means that user access to unstructured content needs to be based on providing access to the content objects required to execute organisational activities (see section 3.3.3.3, 3.4.1.3.1)

- Information confidentiality requirements are based on the organisational impact if confidential information is disclosed within or outside of the organisation. The activities where knowledge is created or used determine the confidentiality requirements of the knowledge (see section 4.2.7).

- Information integrity requirements are determined by the value of the document as either an organisational record or an important source of organisational knowledge. The required integrity of the record is determined by the legal and regulatory requirements set for the execution of the activity. The required integrity of knowledge objects is determined by the organisational impact if employees are not in possession of the correct information when executing organisational activities (see section 3.3.3.4, 3.4.1.3.2).

- Information availability requirements are determined by the impact on the organisation if the records or knowledge required to execute organisational activities are not available (see section 3.3.3.4, 3.4.1.3.2.)

5.2.3 How does an activity-based taxonomy provide for the management requirements of unstructured content?

The proposed model is based on an activity-based taxonomy; the applicability of the model using the model for managing unstructured content was discussed in detail in Chapter 4. From the model description in Chapter 3 and the model evaluation in Chapter 4 it is should be clear that such a taxonomy provides the following:

- An ability to determine the content management requirements of unstructured content, from the activity where the content was created or used.
- The ability to specify the content management requirements on the activity level in the taxonomy and then applying these to all content associated with that activity.
- A way of managing employee access to unstructured content based on the activities that the employee needs to execute in the organisation.

5.2.4 Answering the research question

The research question posed in section 1.1.1 was “To what extent does an activity-based taxonomy satisfy the records management, knowledge management and information requirements for managing unstructured content?”. The proposed model and the evaluation of the model showed that an activity-based classification system satisfies all of the records management requirements as well as all of the information security requirements associated with the management of organisational records. The knowledge management requirements are satisfied for the management of functional, process and activity related knowledge; however, subject related knowledge will necessitate the addition of secondary taxonomies.

5.3 Limitations of the proposed model

Although the use of an activity-based taxonomy might make classification easier for content creators and users, the burden of assigning documents into pre-existing content repositories cannot be automated. This means that implementing this model will require a huge undertaking to index or re-index existing content according to the activity-based taxonomy.

Legal discovery might require that related records are grouped in such a way that they can be easily found during discovery. The model caters for this eventuality insofar as the activity where the documents are used, or the content creator, content editor or content user, can be utilised to establish a relationship. There might however be a requirement to use criteria such as product type, vendor or client for establishing the relationship. The model does not support this, although the use of secondary taxonomies and key word search technologies may possibly satisfy these requirements.

Knowledge types such as procedural, declarative, causal and general knowledge are not specified as part of the classification process in the suggested model. This could however be carried out by requiring additional metadata to be captured during content creation or use.

Cross-linking information can be undertaken on the activity, content creator and content user level, but the model does not support cross-linking of information at subject level. This could be achieved by using additional taxonomies for classifying information.

5.4 Further research

This model has not been evaluated in an organisational setting, which means that the user experience of such a model was not tested, it is also possible that an implementation of the model might identify more limitations that need to be addressed. The model also displays certain limitations in terms of using the activity based taxonomy for specifying and applying some content management requirements (see sections 5.3). Specific areas that require further research are:

- Implementing the model in an organisational setting to assess the social acceptability and also to determine the limitations of the model.
- A more in-depth analysis regarding which secondary taxonomies and other metadata are required to address the limitations of the proposed model in managing unstructured content.
- This model has been build to manage unstructured content; hence research into the applicability of the model for managing structured content is required.

5.5 Conclusion

The proposed model and the evaluation of the model demonstrated that using organisational activities for organising unstructured content provides a sound platform from which the content management requirements of such content can be determined. The research also indicated that the records management, knowledge management and information security requirements of an organisation are closely related to the organisational activities where records and knowledge are created or used.

Bibliography

1. Baird, L., Henderson, J. 2001. *The knowledge engine: How to create fast cycles of knowledge-to-performance and performance-to-knowledge*. San Francisco: Barrett-Koehler.
2. Bellaver, R.F., Lusa, J.M. 2002. *Knowledge management strategy and technology.* London: Artech House.
3. Bishop, M. 2004. Introduction to computer security. Addison Wesley Professional. Viewed 23 May 2008, <http://safari.informit.com/0321247442>
4. Braganza, A. 2004. Rethinking the data-information-knowledge hierarchy: Towards a case-based model. *International Journal of Information Management*. 24:347-356 . Viewed 5 November 2007, <http://www.sciencedirect.com>
5. Bruno, D., Richmond,H. 2003. The truth about taxonomies. *Information Management Journal.* 37(2): 44-53. Viewed 24 April 2005, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdlink?did=320636601&sid=1&Fmt=4&clientId=15443&RQT=309&VName=PQD>
6. Cisco, S.L., Strong, K.V. 1999. The value added information chain. *Information Management Journal* 33(1): 4-15. Viewed 11 May 2005, <http://0-search.ebscohost.com.innopac.up.ac.za:80/login.aspx?direct=true&db=buh&AN=2333081&site=ehost-live&scope=site>
7. Dalkir, K. 2005. *Knowledge management in theory and practise*. Elsevier: Butterworth-Heinemann.
8. Daum, P. B. 1997. Technology and the four-level information hierarchy. *Records Management Quarterly*. 31(4): 8-14. Viewed 26 July 2007, <http://findarticles.com/p/articles/mi_qa3691/is_199710/ai_n8758411>
9. Davenport T.H., Jarvenpaa S.L., Beers M.C. 1996. Improving knowledge work processes. *Sloan Management Review*.37(4):p53-65. Viewed 3 February 2008, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdweb?did=9891109&sid=1&Fmt=2&clientId=15443&RQT=309&VName=PQD>

10. Ekweozor, U., Theodoulidis, B. 2004. Review of retention management software systems. *Records Management Journal*. 14(2):64-77. Viewed 8 January 2006, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdweb?did=679673871&sid=2&Fmt=4&clientId=15443&RQT=309&VName=PQD>
11. Eppler, M.J. 2001. Making knowledge visible through intranet knowledge maps: Concepts, Elements, Cases. *System Sciences. Proceedings of the 34th Hawaii International Conference on System Sciences*. p1-10. Viewed 12 September 2008, <http://0-ieeexplore.ieee.org.innopac.up.ac.za/stamp/stamp.jsp?arnumber=00926495>
12. Essmayr, W. ,Probst, S., Weippl, E. 2004. Role-based access controls: Status dissemination and prospects for generic security mechanisms. *Electronic Commerce Research.* 4(1-2):127-156. Viewed 5 August 2007, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdweb?did=501147881&sid=5&Fmt=2&clientId=15443&RQT=309&VName=PQD>
13. Ferraiolo D., Kuhn R. 1992. Role-based access control*. Proceedings of the 15th National Computer Security Conference.* 1992. Viewed 12 May 2008, <http://csrc.nist.gov/rbac/ferraiolo-kuhn-92.pdf>
14. Frank, C., Gardoni, M. 2005. Information content management with shared ontologies: At Corporate Research Centre of EADS. *International Journal of Information Management.* 25, (1):55-70. Viewed 23 April 2005, <http://0-dx.doi.org.innopac.up.ac.za/10.1016/j.ijinfomgt.2004.10.009>
15. Frank, U.1998. Essential research strategies in the information systems discipline – reflections on formalisation, contingency and the social construction of reality. *The Systemist*, Vol. 20:98-113. Viewed 18 February 2008, <http://www.wi-inf.uni-due.de/FGFrank/documents/Zeitschriftenartikel/Systemist.pdf>
16. Garshol, L.M. 2004. 'Metadata? Taxonomies?Topic maps! Making sense of It all'. *Proceedings of XML Europe 2004*. Viewed 2 September 2008, <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>
17. Grey, D. 1999. Knowledge mapping: A practical overview. Viewed 12 September 2008, <http://www.impactalliance.org/file_download.php?location=S_U&filename=10383546681Knowledge_Mapping.htm>

18. Grover, V. Davenport, T.H. 2001. General perspectives on knowledge management: Fostering a research agenda. *Journal of Management Information Systems.* 18(1):5-21. Viewed 28 January 2008, <http://0-search.ebscohost.com.innopac.up.ac.za:80/login.aspx?direct=true&db=buh&AN=4753177&site=ehost-live&scope=site>
19. Hammond, T. Hannay, T. Lund, B. Scott, J. 2005. Social bookmarking tools(I). *D-Lib Magazine* 11(4). Viewed 4 September 2008, <http://www.dlib.org/dlib/april05/hammond/04hammond.html>
20. Handzic M., Zhou A.Z. 2005. *Knowledge management*. Oxford: Chandos Publishing.
21. Jablonski S., Horn S., Schlundt M. 2001. Process oriented knowledge management. *Proceedings of the Eleventh international Workshop on Research Issues in Data Engineering.* 77-84. Viewed 5 February 2008, <http://0-ieeexplore.ieee.org.innopac.up.ac.za:80/stamp/stamp.jsp?arnumber=916494&isnumber=19789>
22. Jenkins, T. 2004. *Enterprise content management*. Canada : Open Text Corporation.
23. Jimerson, R.C. 2003. Deciding what to save: *OCLC Systems & Services*. 19(4):135-140. Viewed 17 April 2005, <http://www.emeraldinsight.com/10.1108/10650750310508108>
24. Kang, I., Park, Y., Kim,Y. 2003. A framework for designing a workflow-based knowledge map. *Business Process Management Journal*. 9(3):281-294. Viewed 15 May 2005, <http://www.emeraldinsight.com/10.1108/14637150310477894>
25. Karjalainen, A., Paivarinta, T., Tyrvainen, P., Rajala, J. 2000. Genre-based metadata for enterprise document management. 2000. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*.2:4-7. Viewed 1 May 2005, <http://0-ieeexplore.ieee.org.innopac.up.ac.za:80/stamp/stamp.jsp?arnumber=926696&isnumber=20043>
26. Kasanen, E, Lukha, K, Siitonen, A. 1993. The constructive approach in management accounting research. *Journal of Management Accounting Research* 5: 243. Viewed 12 February 2008, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdweb?did=7573831&sid=1&Fmt=3&clientId=15443&RQT=309&VName=PQD >

27. Kim S., Hwang H., Suh E. 2003. A process-based approach to knowledge-flow analysis: A case study of a manufacturing firm. *Knowledge and Process Management.* 10(4):260-276. Viewed 30 October 2007, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdweb?did=516159791&sid=1&Fmt=2&clientId=15443&RQT=309&VName=PQD>
28. Kwasnik, B.H. 1999. The role of classification in knowledge representation and discovery. *Library Trends* 48(1):22-47. Viewed 16 June 2005, <http://0-search.ebscohost.com.innopac.up.ac.za:80/login.aspx?direct=true&db=aph&AN=2412873&site=ehost-live&scope=site>
29. Lamont, J. 2003. Expertise location and the learning organization. *KM World Magazine.* Viewed 14 September 2008, <http://www.kmworld.com/Articles/Editorial/Feature/Expertise-location-and-the-learning-organization-9410.aspx . >
30. March, T.S., Smith, G.F. 1995. Design and natural science research on Information Technology. *Decision Support Systems* 15(4): Viewed 12 February 2008, <http://www.sciencedirect.com/science/article/B6V8S-3Y5FNDY-1/2/2ea948195f581aa5f63b92459958d987>
31. McKemmish, S., Acland, G., Reed B. 2000. Towards a framework for standardising recordkeeping metadata: The Australian recordkeeping metadata schema. Viewed 21 January 2008, <http://www.sims.monash.edu.au /research/rcrg/publications/framewrk.html>
32. Megill K.A., Schantz H. 1999. *Document management: New technologies for the information services manager*. London. Bowker Saur.
33. Middleton, R., Smith, H. 2002. Data retention policies after Enron: Damned if you do, damned if you don't? A look at data retention policies in the aftermath of Enron. *Computer Law & Security Report*. 18(5): 333-337. Viewed 23 April 2005, <http://0-dx.doi.org.innopac.up.ac.za/10.1016/S0267-3649(02)00910-X>
34. Mouton, J. 2005. *How to succeed in your Master's & Doctoral studies*. Pretoria: Van Schaik Publishers.
35. Nonaka, I.1991, The Knowledge-creating company. *Harvard Business Review*, 69(6):96-104. Viewed 16 October 2007, <http://0-search.ebscohost.com.innopac.up.ac.za:80/login.aspx?direct=true&db=buh&AN=9201061306&site=ehost-live&scope=site>

36. Oh, S., Park, S. 2003. Task-role-based access control model. *Information Systems.* 28(6):533-562. Viewed 6 January 2006, <http://0-dx.doi.org.innopac.up.ac.za/10.1016/S0306-4379(02)00029-7>
37. Olivier, M.S. 1999. *Information Technology research: A practical guide*, Published by author.
38. Penn, I.A., Pennix G.B., Coulson J. 1994. *Records management handbook*. Second Edition. Aldershot. Gower.
39. Remus U., Schub S. 2003. A blueprint for the implementation of process-oriented knowledge management. *Knowledge and Process Management*. 10(4):p237-253. Viewed 30 October 2007, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdweb?did=516159821&sid=1&Fmt=10&clientId=15443&RQT=309&VName=PQD. >
40. Robinson, S. 2006. Conceptual modeling for simulation: Issues and research requirements. Proceedings of the 2006 Winter Simulation Conference.792-800. Viewed 11 September 2007, <http://0-ieeexplore.ieee.org.innopac.up.ac.za:80/stamp/stamp.jsp?arnumber=4117684&isnumber=4117570>
41. Sandhu R., Ferraiolo D., Kuhn R. 2002. The NIST model for role-based access control: Towards a unified standard. In Proceedings of the Fifth ACM Workshop on Role-Based Access Control. Viewed 4 May 2008, <http://doi.acm.org/10.1145/344287.34430>
42. Sandhu R.S., Coyne E.J., Feinstein H.LO. Youman C.E. 1996. Role-based access control models. *IEEE Computer*, Vol29(2):38-47. Viewed 25 March 2008, <http://csrc.nist.gov/rbac/sandhu-ferraiolo-kuhn-00.pdf>
43. Satyadas, A. Harigopal, U. Cassaigne, N.P. 2001. Knowledge management tutorial: An Editorial Overview. IEEE Transactions On Systems, Man And Cybernetics-Part C: Applications And Reviews. 31(4):429-437. Viewed 27 January 2006, <http://0-ieeexplore.ieee.org.innopac.up.ac.za:80/stamp/stamp.jsp?arnumber=983926&isnumber=21210>

44. Schwarz, A., Mehta, M., Johnson, N., Chin, W.W. 2007. Understanding frameworks and reviews: A commentary to assist us in moving our field forward by analyzing our past. SIGMIS Database. 38(3):29-50. Viewed 16 November 2007, <http://0-doi.acm.org.innopac.up.ac.za:80/10.1145/1278253.1278259>
45. Shankar R., Gupta A. 2005. Towards framework for knowledge management implementation. *Knowledge and Process Management*. 12(4):259-277. Viewed 3 October 2007, <http://dx.doi.org/10.1002/kpm.234>
46. Skupsky, D.S. 1999. Applying records retention to electronic Records. *Information Management Journal.* 33(3):28-35. Viewed 8 March 2005, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdlink?did=44789031&sid=1&Fmt=3&clientId=15443&RQT=309&VName=PQD>
47. Sletten, L. 1999. Lesson from down under: Records management in australia. *Information Management Journal*. 33(1):26-30. Viewed 5 November 2007, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdlink?did=44816077&sid=2&Fmt=3&clientId=15443&RQT=309&VName=PQD>
48. South Africa. 2002. Electronic Communications and Transactions Act. No. 25 of 2002. *Government Gazette*, 446(23708). Viewed 22 September 2004, <http://www.info.gov.za/gazette/acts/2002/a25-02.pdf>
49. Strong K.V. 1999. Integrating EDMS functions & RM principles. *Information Management Journal.* 33(3): 18-27. Viewed 24 April 2005, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdlink?did=44788875&sid=1&Fmt=4&clientId=15443&RQT=309&VName=PQD>
50. Tiwana, A. 2002. *The knowledge management toolkit : Orchestrating IT, strategy and knowledge platforms*. Upper Saddle River, NJ: Prentice Hall.
51. Tonkin, E. 2006. Folksonomies: The fall and rise of plain-text tagging. *Ardiadne*. 47. Viewed 11 September 2008, <http://www.ariadne.ac.uk/issue47/tonkin/>
52. Upward F. 1996. Structuring the records continuum - Part One: Postcustodial principles and properties. *Archives and Manuscripts*. 24(2). Viewed 23 January 2008, <http://www.sims.monash.edu.au/research/rcrg/publications/recordscontinuum/fupp1.html>
53. Wiggins B, 2000. *Effective document management: Unlocking corporate knowledge.* Aldershot: Gower

54. Zack M.H. 1999. Managing codified knowledge. *Sloan Management Review*.40(4).45-58. Viewed 28 January 2008, <http://0-proquest.umi.com.innopac.up.ac.za:80/pqdweb?did=43482351&sid=4&Fmt=6&clientId=15443&RQT=309&VName=PQD>