DEVELOPMENT AND IMPLEMENTATION OF AN ENTERPRISE CONTENT MANAGEMENT SYSTEM FOR THE CSIR-FOOD RESEARCH INSTITUTE

By

Kwabena A. Bugyei

Raphael K. Kavi

Stephen Nketia

October 2015

CSIR-Food Research Institute, P.O. Box M20, Accra

List of Figures

List of Abbreviations and Acronyms

AIIM – Association for Information and Image Management

AOP – Aspect Oriented Programming

API – Application Programming Interface

BI – Business Intelligence

CIFS – Common Internet File System

CM – Content Management

CMIS – Content Management Interoperability Services

CSIR- Council for Scientific and Industrial Research

CSS – Cascading Style Sheet

DAM – Document Asset Management

DM – Document Management

ECM – Enterprise Content Management

ECMS – Enterprise Content Management System

EDM – Enterprise Document Management

ERM – Enterprise Record Management

ERP – Enterprise Resource Planning

ES – Enterprise System

EULA – End User License Agreement

FRI – Food Research Institute

FTP – File Transfer Protocol

GUI – Graphical User Interface

HTML – Hypertext Markup Language

IIS – Internet Information Server

IMAP – Internet Message Access Protocol

IT – Information Technology

JSP – Java Server Pages

LDAP – Lightweight Directory Access Protocol

LTS – Long Term Support

NDA – Non-Disclosure Agreement

OS – Operating System

RAID – Redundant Array of Independent Disks

RDBMS – Relational Database Management System

RM – Record Management

SME – Small and Medium Size Enterprises

SQL – Structured Query Language

WCM – Web Content Management

WebDAV – Web-based Distributed Authoring and Versioning

XML – Extensible Markup Language

ABSTRACT

Today's enterprises create an increasing volume of unstructured content that includes documents, e-mail messages, videos, images, instant messages, Web pages, and other digital assets. The steadily decreasing costs for storage space and the general availability of large scale storage systems, enable enterprises to store all their relevant business data. However, this content often exists in a state of unmanaged chaos with little or no routines regarding capturing, storing, sharing, retrieving, archiving, etc. which prevents an enterprise from properly using these valuable assets for better collaboration, knowledge sharing, improved customer and vendor communications, compliance, and increased process efficiency. Hence enterprises are facing increasing challenges concerning management of their content.

The Food Research Institute of the Council for Scientific and Industrial Research is a leading food research institution in Ghana where there is an immense amount of unstructured content and information created and received internally and externally every day. However, these content are scattered across staff personal computers and exits in an unmanaged state. This project seeks to develop and implement a secure and robust web-based ECM system solution using the Alfresco platform, an open source ECM software that will take control of the unstructured content and enhance the information management routines. The project focuses on managing the unstructured content with an enterprise-wide approach.

The end result was a successfully rolled out ECM system that easily controls the volumes of information for a comprehensive member view, streamline business processes, and capture documents electronically. The system provided a platform for management and staff of the Institute to store, manage and share the organizational documents (project documents, forms, lab reports, memo, training materials, PowerPoint presentations, spreadsheets, web pages, photos, videos, documentary, digital assets, email messages, notices, research findings, research proposals, research reports etc) in a centralized repository. This allows users from different divisions to access their content, latest revision documents anywhere anytime and on any device.

Table of Contents

1.0 CHAPTER ONE: INTRODUCTION

1.1 Introduction to the CSIR-Food Research Institute

The Food Research Institute (FRI) is one of the thirteen (13) Research Institutions of the Council for Scientific and Industrial Research, (CSIR) which operates as a Science and Technology Research Development Organisation. The Food Research Institute was established in October 1963, incorporated by L I No. 438 of 19th March 1965 and became an institute of CSIR in October, 1968 by NLC Decree 293.

CSIR-FRI is mandated to conduct market-oriented applied research, provide technical services and products to the food industry and assist in poverty alleviation through the creation of opportunities for income generation, thus contributing to food security and foreign exchange earnings. CSIR-Food Research Institute is the leading S&T Institute in the transformation of the food processing industry in Ghana.

The core research interests and programs of the FRI include:

I. Root and Tuber Products Programme

II. Meet, Fish, Poultry and Dairy Products Programme

III. Cereal and Grain Legume Products Programme

IV. Fruit and Vegetable Products Programme

1.1.2 Vision

The Institute`s vision is to play a key role in the transformation of the food processing industry and to be internationally competitive with particular reference to product safety, quality and preservation.

1.1.3 Mission

The Institute`s mission focusses on providing scientific and technological support to the growth of the food and agricultural sectors of national economy in line with corporate priorisation and national objectives. Primarily, the Food Research Institute`s mission is to conduct market-oriented applied research and provide technical services and products profitably to the private sector and other stakeholders.

1.1.4 Core Mandate

- The CSIR-FRI conduct applied research into problems of:
- Food processing and preservation
- Food safety and storage

- Food marketing, distribution and utilisation
- National food and nutritional security in support of the food industry
- Advice Government on its food policy
- To assist in poverty alleviation through creation of opportunities for generating and increasing income within the micro, small, medium and large-scale food industries
- In support of the food and agricultural sectors of the national economy

1.1.5 Core Values

CSIR-Food Research Institute believes and ascribes to the following values:

- Professionalism
- Team work
- Innovativeness
- Competitiveness
- Quality Delivery

1.2 Background, Purpose and Motivation of the Project

Enterprises today are facing a rapidly increasing amount of digital information and content to be handled (Aleksy & Schwind, 2006). As this abundance of available data grows, managing the information becomes more difficult, which can lead to information overload. As early as in 1970, Alvin Toffler coined the term information overload in his book *Future Shock* (Toffler, 1970 cited in vom Brocke et al., 2010). Today's digital information age is characterized by steadily decreasing storage space prices and the general availability of large scale storage systems, enabling enterprises to store all their relevant business data (Aleksy & Schwind, 2006). In turn this has led to a tremendously increasing information flood, thus fulfilling Toffler's forecasts (vom Brocke et al., 2010).

Most organizations today generate information at such a rate that the challenge is putting this information in a format and in a place where it can be found again, when needed (O'Callaghan & Smits, 2005). Within enterprises, unstructured information currently accounts for the majority of a company's overall data (Rogalski, 2006). The Gartner Group, cited in O'Callaghan & Smits (2005), estimates that most of the data captured in organizations today is unstructured (75%-80%) and not in such a format that it can be found when needed. The quantities of unstructured content such as documents, images, e-mail, Web content, audio and video are all growing at an astonishing

rate, and Rogalski (2006) states that this explosion of unstructured data is one of the biggest challenges facing businesses today. At the same time companies are asking their knowledge workers to do more and more in less and less time (Glick-Smith, 2004).

Lewis (2009) says that information is the most valuable enterprise asset next to people and argues that one of the best ways for an enterprise to increase its competitive advantage is through leveraging its information assets. However, he further states that information also can be the most problematic asset. "Organizations are buried in digital content, leaving people scrambling to find the right information when they need it" (Gottlieb, 2005**)**. It has been estimated that information workers spend up to 30% of their working day searching for data and approximately 15–25% on non-productive, primarily information-related tasks (Burnett et al., 2006 cited in vom Brocke et al., 2010).

E-mail remains a major contributor to this information overload, as people struggle to keep up with the rate of incoming messages (Meyer, 2005). As well as filtering out unsolicited messages (spam), users also have to contend with the growing use of e-mail attachments in the form of lengthy reports, presentations and media files (Meyer, 2005). It is not unusual to hear of people responding to and managing 150 e-mails daily.

Many organizations are also struggling with the increasing amount of functional systems and applications, with each application creating its own separate location to file and store information, and where each networked system represents a boundary between pieces of information (Meyer, 2005). Sometimes information even is stored on local disks (Bandorf et al., 2004). Hence the consequence is that users are forced to navigate through complex folder structures to file or retrieve their information (Meyer, 2005), that the data cannot be found at all, that data is kept redundantly in several copies, that changes are not managed properly, etc (Bandorf et al. 2004). Meyer (2005) suggests that the situation is even worse for businesses trying to manage information across work groups or spread over geographic locations, where the folder structures become more and more complex. This makes it practically impossible to provide users with uniform access to complete and accurate information in real-time (Meyer, 2005). As stated by Gottlieb (2005), "Information is power, but it is useless when scattered across employee hard drives and e-mail in-boxes across the enterprise."

At the same time, within today's fast paced global competitive business environment, organizations are striving to achieve the maximum efficiency out of the available resources (Usman et al., 2009). This has forced organizations to improve the supporting activities for their core business. Nowadays, particularly large-scale enterprises are becoming more and more aware of this information overload and are confronted with the challenge of efficiently handling it (vom Brocke & Simons, 2008), together with internal and external government pressures to meet compliance requirements (Blair, 2004).

In order to meet these factors, organizations around the world are moving towards Enterprise Content Management (ECM) solutions (vom Brocke & Simons, 2008). ECM is regarded as a new and emerging class of information systems (IS), that integrates and extends several areas related to Information Management (IM) (Tyrvainen et al, 2006), Knowledge Management (KM) and

Information Technology (Munkvold et al., 2006), and is increasingly being perceived as the state of the art for storing and retrieving unstructured content and documents in practice (Aleksy & Schwind, 2006; Dilnutt, 2006). The central ECM business problem is the management of the entire content lifecycle across the enterprise from beginning to end (Alsup & Strong, 2004).

Documents are the life blood of organizations. Every event, every transaction, every development, many notices or announcements, many results, official and unofficial documents can be needed for analysis, for review, for legal issues or simply for the record or the purpose it was expected to accomplish. This creates a need for an efficient solution to manage these documents and thus Enterprise Content Management Systems (ECMS or simply ECM) and Enterprise Content Management systems (ECM) came into existence.

Documents, emails, spreadsheets and presentations are created faster than we can count them. Documents, images, web pages, and other types of content are known as unstructured content. As the amount of content continues to grow within organizations, challenges with its creation, management, and distribution continue to grow as well. Enterprise Content Management (ECM) provides organizations with a platform to house unstructured content and deliver it in the proper format to multiple enterprise applications. With this technology, you can efficiently build content-rich business applications, reuse content, and integrate hundreds of content services with other business applications. ECM helps to decrease costs, automate processes, reduce resource bottlenecks, share content effectively, minimize the number of lost documents, and better manage risk.

The Food Research Institute of the Council for Scientific and Industrial Research is a leading food research institution in Ghana where there is an immense amount of unstructured content and information created and received internally and externally every day. However, theses content are scattered across staff hard drives and are not managed properly. In order to take advantage of the benefits that ECM solution seeks to offer, this report describes the processes leading to the successful development and implementation of a secure and robust web-based ECM system solution using the Alfresco platform, an open source ECM software that will take control over the unstructured content and enhance the information management routines. The project focuses on managing the unstructured content with an enterprise-wide approach.

1.2.1 Problem Statement

The continuous growth of information is a major challenge facing organizations and if not properly managed could have an adverse effect that impairs the ability of IT to meet the strategic needs of organizations (Clark, 2008). Information, including structured data (stored in databases) and unstructured content (stored in file systems, content management systems, email servers and more) is growing and this could generate negative effects, storage costs, draining IT budgets, costs paralleling the explosion in growth and employee productivity limited by time wasted searching for information that is inaccessible and is recreated. The continuous rise in the demand for ECMS applications within organizations is warranted by the fact that whiles other content management systems such as Business Process Management (BPM) and ERP seeks to combine all the

traditional business applications, making them exchange information among themselves, ECMS seeks to go further by creating knowledgebase of the enterprise's business and content resources (Allen, 2007). ECMS accepts both structured and unstructured data repository that has a uniform structure.

ECMS aims to combine the functionalities of the Web, content management and to systematically incorporate not only traditional publishing activities but e-mail, financial records, human resource, documents, etc., for an entire organization (Clark, 2008).

The Gartner Group estimates that an average document gets copied 19 times, companies spend $20 in labour to file a document, $120 in labour to find a misfiled document, and $220 in labour to reproduce a lost document and that a typical organization of 1,000 people wastes over $11 million per year through manual document handling and management, (Smith and McKeen 2003). In a bid to curtail these deficiencies and meet the business challenges in a very competitive market, organizations are adopting the Enterprise Content Management Systems as a strategic investment with the intention of improving the supporting activities of their business processes (Rouse, 2005). Kaplan (2002a) claims that ECMS is a big business and that as at 2001, it had about $3.5 billion in sales and this was expected to double by 2006. Research by Association for Information and Image Management (AIIM, 2008), also shows that the "network overall spending on licenses in all ECMS applications is set to increase considerably over the next 12 months compared to the last 12 months, and spending on consultancy services is likely to increase slightly overall". It is not a hidden fact that most organizations are overwhelmed by the number of paper documents, data, reports, Web Pages, and digital assets which have literally grown tremendously in recent years causing considerable information overload (Kaplan, 2002b).

Tools, processes and skills needed to deal with each type of enterprise information is increasingly growing and as such distinctions among various forms of data are becoming obvious and irrelevant. In response to these problems, ECMS seek to manage organizations information assets (Kaplan, 2002b).

1.2.2 Project Objectives

The main objective of the project is to develop and implement a web-based Enterprise Content Management System (ECM) for the CSIR-Food Research Institute. The specific objectives included the following:

- To identify current bottlenecks with document management and specify user functional requirements for proper document life cycle.
- To evaluate and select the right tool for the ECM solution.
- To install and configure the development environment and the necessary software
- To develop and implement the ECM application with the selected software

- To test the usability and functionality of the ECM application.
- To deploy the system developed to a production server.

1.2.3 Outline of the Project

Chapter One introduces CSIR-Food Research Institute, its vision, mission, mandate and core values. It also presents the background, purpose, motivation and problem statement as well as the objectives, duration and outline of this report.

Chapter Two introduces the concept of ECM by giving a brief historic overview of the phenomenon and discusses different ECM definitions.

The third chapter deals with the different components of ECM such as the enterprise wide perspective, content and tagging of content with metadata, and management of content.

In the fourth chapter, open source ECM vs. Commercial ECM softwares are discussed. The chapter also looked at the selection of the appropriate software for the system. It presents the justifications for choosing Alfresco as the preferred software for the implementation of this project.

Chapter Five "Project Implementation", describes the processes and methods used in the development and implementation of the ECM Solution. It describes the installation and configuration of the development environment and other softwares. Testing and deployment as well as the description and features of the developed system are all presented in this chapter.

The sixth chapter summarizes the project with conclusions and recommendations.

1.2.4 Duration of the Project

This project was carried out between the periods of January 2015 through to October 2015.

1.2.5 Initial Situation

Before the implementation of the ECM solution, documents were kept on staff individual PC's, sending of documents were done via emails (leading to bandwidth wastage), transfer of documents via pen drives, external hard drives, and on network location as shared folders. There was no centralized repository for storing, managing and sharing of documents. This resulted in difficulties in managing documents leading to duplication of documents, confusion between multiple document versions, virus infected documents from pen drives, data loss, etc. Figure 1 below shows the initial situation.

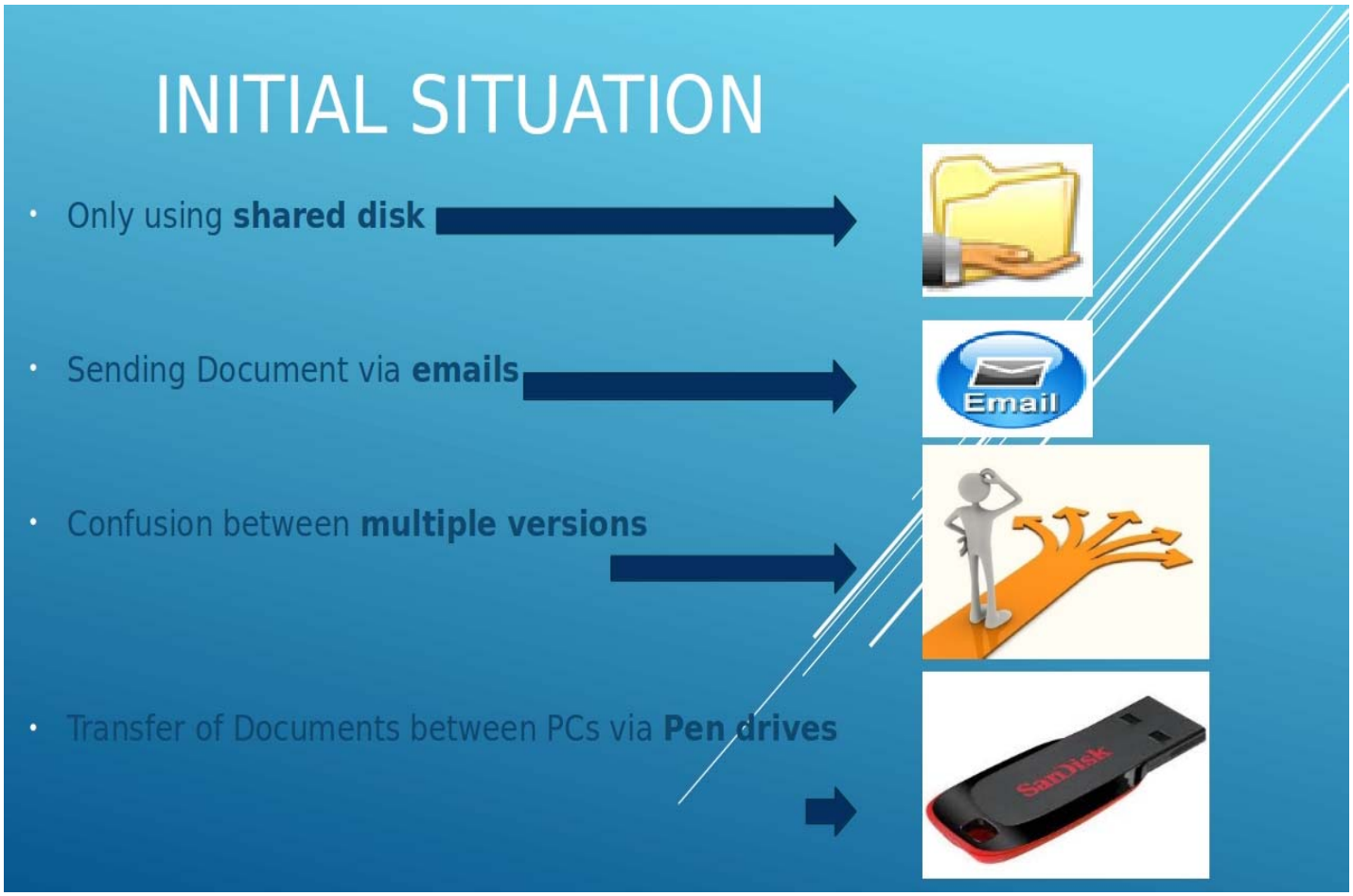


Figure 1: Initial Situation

2.0 CHAPTER TWO: LITERATURE REVIEW

2.1 Overview of Enterprise Content Management

This chapter examines the phenomenon of Enterprise Content Management (ECM). To get a better understanding of the concept, the next section gives a brief historic overview of ECM, before defining the concept in more detail.

2.2 Brief Historic Overview of ECM

From the early days of the IS discipline, IM and a variety of concepts for controlling an organization's digital information assets has been investigated (vom Brocke et al., 2010). Various concepts for controlling an organization's digital information assets, such as Electronic Document Management (EDM) or Records Management first released during the late 1980's have been investigated (Kemp, 2007). Later on, the adoption of the Internet in the 1990's resulted in uncontrolled growth of information assets in Web sites, Intranets, and Extranets, which gave rise to development of concepts like Web Content Management (WCM), as there was a need for managing corporate Web contents (McKeever, 2003). The Web moved from small informally designed Web sites with just 1,000 Web pages on the Web in 1992, into large, rapidly changing sites, where the need for strong management tools became greater. Towards the end of the 1990's however, organizations began to demand EDM products with an integrated approach to managing documents, Web content, and digital assets that could address more than one business need (Wilkoff et al., 2001), and in 2000/2001 software companies responded to this by starting to produce ECM systems (Kemp, 2007).

According to Päivärinta & Munkvold (2005) the concept of ECM has evolved during the past 20 years, and that ECM originates from the development of WCM. The scope of this practice, at its early stages, was just a small fraction of what we know today as ECM (Usman et al., 2009).

2.3 What is ECM

The concept of ECM is a new and emerging field in both the IS academia (Tyrväinen et al., 2006) and practice (Dilnutt, 2006). Today, ECM is being used in organizations and industries, often where there is strong competition, rapid product innovation and changing consumer behavior (Kemp, 2007). To maintain a competitive advantage (Porter, 1998) in such an environment there is a great need for IT and organizations to adapt and innovate. Effective capturing, managing, storing, preserving and delivering of information, content and documents can help facilitate this, such as ECM intend to (Kemp, 2007).

However, despite its widespread use, there is very little consensus on the meaning of the term ECM (Gottlieb, 2005), and there is still a lack of understanding and unified definitions of the concept (vom Brocke et al., 2010). According to Smith & McKeen (2003) there exists no single, unified perspective on ECM and there is still a considerable confusion about the meaning of the concept, where managers, academics, and vendors are all trying to understand and define it.

There are various closely related and well-researched concepts to ECM such as Electronic Document Management (EDM), Content Management (CM) and Web Content Management (WCM) (vom Brocke et al., 2010). Hence Päivärinta & Munkvold (2005) discuss "whether ECM actually represents anything new compared to the established constructs of IM" (p. 1). They conclude that ECM differs from the existing solutions in that it aims to combine these (amongst others) previously separated ISs, and that ECM goes beyond their individual and collective scopes. Munkvold et al. (2006) also acknowledge that a majority of issues associated with ECM initiatives can be traced back to these established research areas when studied separately, and that the concept of ECM integrates these issues in a new manner.

2.4 Defining ECM

The term ECM has been given various definitions and meanings since its origin in 2000/2001 (Gottlieb, 2005), thus defining the concept is not an easy task.

Smith & McKeen (2003, p. 648) define ECM as, "the strategies, tools, processes and skills an organization needs to manage all its information assets (regardless of type) over their lifecycle".

Blair (2004, p. 65) regards ECM as a concept that covers a wide range of technologies:

"ECM is the technologies, tools, and methods used to capture, manage, store, preserve, and deliver content across an enterprise. ECM is also concerned with information that would not normally be classified, retained, and managed as a record. ECM focuses on unstructured information, that is, the free-form content that exists outside the confines of databases or systems with fixed routines and pathways, such as e-mail, word processing documents, digital images and Portable Document Format files."

Moore cited in Iverson & Burkart (2007, p.407) explains ECM as an umbrella covering different technologies:

"We use enterprise content management as an overarching term that describes a number of different technologies that up until recently have been seen as discrete markets. It includes document management, Web content management, records management, document imaging and digital asset management, among other things. ECM encompasses all of the unstructured content in an organization."

ECM could be the technical solutions to publish and share content like Microsoft SharePoint, EZ Publish, Joomla or other publication tools. But ECM is much more than technology, and it seems as the technology perspective of ECM has received less attention in the latest definitions, whilst strategy and methods have received more consideration. Mescan (2004) acknowledge this perception by calling ECM "a strategy rather than a solution".

The latest and more recently updated definition by The Association for Information and Image Management (AIIM) puts equal emphasis on strategies, methods and tools:

"Enterprise Content Management (ECM) is the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes.

ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists." (AIIM, 2008).

Stalters (2007) also gives minor importance to technology, and emphasize ECM as:

"a management practice that provides for governance of an information management environment toward the goal of improving compliance, information reuse and sharing, and operational performance. ECM is a structured approach employing methods, policies, metrics, management practices and software tools to manage the lifecycle of information and to continuously optimize an organization's collections of information and information management processes."

Technology is an important enabler for ECM, but ECM is above all organizational (Munkvold et al. 2006). Tyrväinen et al. (2006) also have a wide perspective on what ECM is, and argue that technology is only one part of ECM and the challenges inherent. Development of new content management solutions in organizations requires new strategies for how content is to be handled and also changes in business processes and work practices (Salminen et al., 2006 cited in Tyrväinen et al., 2006). "Computer science and software engineering may well produce various technological innovations for content management, but they tend to neglect the content aspect and the organizational context of ECM" (Tyrväinen et al., 2006, p. 628). Duhon (2009) argues that technology, including ECM, will not solve any business problem, and that you got to understand the business first, then apply technology to it. He further states that "too often, there is confusion between ECM as strategy and ECM as technology. It is both, but strategy first; technology second". Munkvold et al. (2006) argue that "the rationale of ECM resides in the global collaboration needs of an organization's employees, customers, and partners through digital information content".

The focus of this project is how ECM System can support unstructured information, like e-mail and text documents. Such files include a large portion of many enterprises' total amount of information, and have a tendency to stay where they are created, e.g. on the individual local PC, if there are no routines to handle this kind of content (unstructured content will be addressed more detailed later). Content is often created by authors working in isolation from other authors within the organization (Rockley, 2003). Walls are created among and within content areas, which leads to content being created, and recreated, and recreated, often with changes or differences at each iteration. Aagre (2008) adds that sharing of information is a sensitive area, because sharing information means one has to give up on something. That is why it is important that management clearly expresses the benefits with an ECM project. For this reason, Aagre (2008) argues that ECM deals with corporate/company policy and management as well as with technology.

ECM is very popular among major corporations, and ECM vendors target large organizations with complex needs (Iverson & Burkart, 2007). de Carvalho (2007) argues that small and medium enterprises (SME), education and research institutions, and government bodies have the same need for unstructured information management, and Alsup & Strong (2004) call to attention that content management problems exists in every organization, regardless of size, number of locations or business complexity. However, cost is a significant barrier to ECM adoption in SMEs and nonprofit organizations (Iverson & Burkart, 2007), and most of the time, they cannot afford the high acquisition and customization costs (de Carvalho, 2007). Although it is possible to use smaller

content management systems for e.g. Intranets and Extranets as well as implement other knowledge management strategies (Iverson & Burkart, 2007). An Open source ECM solution could also be an alternative (de Carvalho, 2007). Iverson & Burkart (2007) also argue that when there is a scarcity of resources, the decision making regarding content and knowledge management is even more important.

Iverson & Burkart (2007) state that any organization considering ECM or a smaller content management system should carefully consider the impact that the system will have on the organization. To assist the decision making, Iverson & Burkart (2007) present a model for evaluating ECM and other similar systems. This content management model provides a framework for analyzing some of the impact that content management systems such as ECM may have on the organization. The questions raised by the model give decision makers who may not be familiar with the technology a way to ask important questions about how the costs of implementing ECM or other content technology will compare to the advantages in their particular case.

According to Alsup & Strong (2004), with the advent of electronic documents, records management (RM) has become a dysfunctional business discipline. Most organizations today are confused about RM roles and boundaries, but they are clear that they are not managing electronic records in a way that is compliant with either their internal policies, if they exist, or the legal requirements that affect their businesses. However, practitioner, specialist and master designations from different course programs regarding ECM, ERM, etc. are beginning to be used to differentiate staff in job postings, resumes, and proposals (Duhon, 2009). Duhon (2009) argues that while companies may not be looking for an "ECM specialist" by name, they are looking for persons with the skills that such a specialist possesses. He states that IT professionals need to have a combination of information management, business processes, and industry experience to understand the user perspective.

Nordheim and Päivärinta (2006) acknowledge that an ECM system can be regarded as a type of Enterprise System (ES) due to its coverage and complexity and that it fulfills the characteristics of ES in general, like Enterprise Recourse Planning (ERP) systems and Customer Relationship Management (CRM) systems that have been given significant attention recent years.

An ECM system usually requires little or no software to be installed on a personal computer, because ECM software leverages Internet technology to deliver services to people. To access such software therefore usually requires only a Web browser, a username, and a password (Jenkins et al., 2006). An ECM solution is often accessed through portals (Mack et al., 2001 cited in Nordheim and Päivärinta, 2006) where information resources from multiple sources and applications are combined, providing navigational aids to cover the information resources of the organization (Nordheim & Päivärinta, 2006).

2.5 Benefits of Enterprise Content Management Systems

It has been argued that ECMS can be beneficial where organizations can create a structured platform that can manage the information from the business processes, functions and the extended enterprise (Smith and McKeen, 2003). Porter (1985), states that it is imperative for organizations

to engage in value creation activities that create competitiveness that is beneficial to the overall well-being of the organization.

Experts are of the view that the ECMS as an information system, should result in outcomes that are beneficial to the organization and if this is not the case, then it should not be implemented. The Oxford English Dictionary defines benefit as "to do good to, to be of advantage or profit to; to improve, help forward" (Oxford English Dictionary Online, 2014). This definition of benefit will be used in this paper, i.e. the word benefit refers to the advantage, profit or an improvement that occurs from the implementation of an ECMS in an organization.

According to Luan and Serban, 2002; Asprey and Middleton, 2003, from the perspective of the ECMS vendor community, they are of the view that potential ECMS benefits can be the following; compliance, efficiency, consistency, customer service and consolidation. Also AIIM (2008) is of the view that potential benefits could also include; return on investment, cost reduction, user satisfaction, amongst others.

2.5.1 Change of Organizational Culture

The enterprise system is a tour de force that impacts the organizational culture. It enables the streamlining of management structures by creating flatter, more flexible, and democratic organizations. With the enterprise system in place, the control of information and standardization of processes which are more consistent with hierarchical, command and control organizations with uniform cultures can be centralized (Davenport, 1998). Junco et al (2005, p.343) acknowledges that the implementation of a content management system leads to a positive change in the corporate culture of the organization with more attention been paid to the implementation of the system by everyone

2.5.2 User Satisfaction

Delone and Mclean(2003) state that higher system quality is expected to lead to higher user satisfaction and use, leading to positive impacts on individual productivity, resulting in organizational productivity improvements. Shang and Seddon (2002) are also of the view that an efficient enterprise leads to increase work efficiency and users who are satisfied with services rendered by the systems.

2.5.3 Improved Efficiency

Chiu and Hung (2005, p.1) are of the opinion that improved business efficiency and accelerated exception handling is a benefit that the organization derives from the implementation of the enterprise content management system in the finance sector. The ECMS also allows for "efficiency, effectiveness, and flexibility of knowledge work and business processes, including reuse of previously created content, metadata, templates, and navigation aids (Paivarinta and Munkvold 2005, p.2).

2.5.4 Improved Content Management

According to Kohler-Kruner (2007) as cited in Kemp (2008), strategic ECMS allows organizations to

"index all user generated content including content residing in existing, disparate business critical or legacy applications, which when coupled with rich search capabilities allow users to find the content they want and need". Accurate and precise delivery of content to the right people at the precise time, it is needed on the particular device, is a benefit that can be derived from the strategic usage of ECMS (Huff and Dirking, 2010).

According to Kemp (2008), in managing content, the following benefits can be derived from the implementation of the ECMS;

- Using ECMS as a strategy has the capability to reduce content duplication thus preventing existing content from being reproduced.
- ECMS can be used to create backups of critical content in an organization thus ensuring no information is ever lost.
- Strategic ECMS can make sure that content is retained for future use

Also the ECMS can lead to "improve the efficiency and effectiveness of documents in their role as a primary mechanism for storing and communicating concepts and ideas within and between organizations (and their groups and individuals)" (Sprague 1995, p.33).

2.5.5 Change of Business Processes

Another benefit derived from the implementation of ECMS is an upgrade of the business processes. Sprague (1995) believes that the implementation of ECMS not only leads to automation of the processes that deals with the computerization of paper work systems that are based on documents but also a reengineering of this processes. The implementation of ECMS leads to the reduction of paperwork and more reliance on automated processes (Sprehe, 2005).Smith and Mckeen (2003) also agree with the assessment made by Sprague (1995) that implementation of the ECMS leads to the simplification of forms and work processes.

2.5.6 Risk Mitigation

Risk mitigation involves prioritizing, evaluating and implementing the appropriate risk-reducing controls recommended from the risk assessment process (Stoneburner et al., 2002). AIIM (2008) is of the view that the implementation of ECMS mitigates risk by restricting access to content, both during its creation and management as well as when delivered.

Other ways risk is mitigated are;

- "By preventing the illegal distribution of rights-managed content by restricting access to content down to the sentence level as well as granting/restricting permissions for forwarding and accessing content.

- Ensuring the identity of a document sender, and the authenticity of the message been sent.
- Using a public and private key pair held by a trusted third party to transact business over the public Internet".

2.5.7 Compliance

"Enterprises must adopt good records management as part of their essential infrastructure for the basic reason that keeping good records protects the enterprise from harmful damage and ensures regulatory compliance" (Sprehe 2005, p.297). A key to successful ECMS compliance strategy is the ability of the organization to integrate and comply with regulations that could be viewed as an opportunity to improve common business processes and not just an ongoing cost to the business. It is not a hidden secret that there can be high costs associated with the compliance initiative for both technology and employees. One of the strategies to help limit the risks and cost will be the adoption of proactive ECMS strategies that can be developed within key areas, such as records management and business process management, (Duhon,2005). This must also ensure that the best practices are followed and that content is properly captured, stored, managed, and disposed of at the appropriate and legal time in its lifecycle. The implementation of a successful ECMS strategy enables the development of a compliance initiative properly increase the chances of tapping into many areas of expertise, particularly legal, IT, and records management, all in support of the overall business objectives of the organization (AIIM, 2008).

2.5.8 Cost Leadership

According to Porter generic strategies, gaining a competitive advantage through cost leadership strategy requires an organization to reduce the costs of production (labour and capital), undergo innovative procedures while still maintaining industry standards and prices (Hedman and Kalling, 2002).

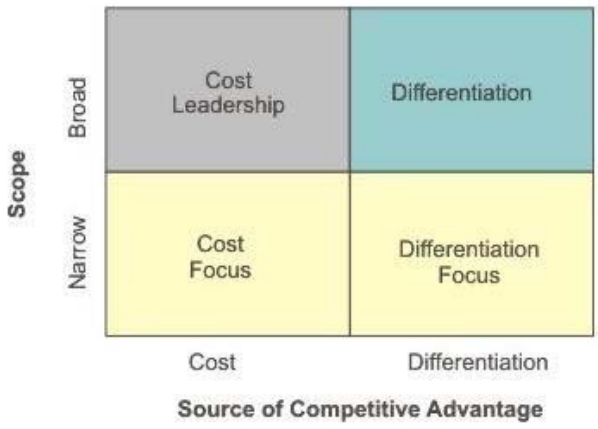


Figure 2: The generic strategies by Porter, Source: Hedman and Kalling (2002)

According to Shang and Seddon (2002), the enterprise system builds cost leadership by building a lean structure with streamlined processes; reaching business economies of scale in operation and by shared services.

2.5.9 Collaboration

Collaboration as explained by Duhon (2005) is the art of working together and utilizing resources in harmony. The deployment of the ECMS (a set of technologies such as instant messaging, whiteboards, online meetings, email, etc) allows work to take place wherever needed. Collaborations allows individuals with complementary, or overlapping areas of expertise to create better outcomes faster than before so that business units and teams can work together anytime, whether in adjoining offices or a world apart (Duhon ,2005). The ECMS addresses operational objectives like saving time, streamlining processes, cutting cost, and improving time to market. Paivarinta and Munkvold (2005, p.2) states that ECMS improves "internal and external collaboration, involving knowledge creation and sharing through digital content in and among enterprises with commonly enacted practices".

2.5.10 New or Value-Added Products or Services

The deployment of the ECMS creates new or value-added services or products or capabilities involving digital content for their customers; "That is, ECM development as such builds competence and technological platforms in the enterprise, on which it becomes quicker to develop and maintain targeted content management applications for emerging purposes"(Paivarinta and Munkvold 2005, p.3).

ECMS has also resulted in modern web-based applications which integrate the previously separated issues of structured databases and dynamic application interfaces, semi-structured documents, and unstructured file management (Paivarinta and Munkvold, 2005).

2.5.11 Improved Governance

According to Terra and Gordon (2003), "an ECMS infrastructure is valueless if the content is inaccurate, outdated or irrelevant for improving users ability to perform". Metadata capture should ideally occur after content creation by combining an automatic, i.e., (author, name, data and more) including manual processes such as (keywords, categories). Dynamism in those that develop diverse taxonomies and maintenance of the information structure requires shared responsibilities and sophisticated tools and more importantly human effort.

It has also been discovered that automatically generated taxonomies poorly performs compared to those managed by professional editors or what is also termed librarians (Terra and Gordon, 2003). The upkeep of information architecture requires committed personnel with non-experts rendered the help to understand a specific function on how content is organized and linked to other content. Another important point to note is that information architects, domain experts and business analysts with the interest implementing a content management solution can transform the expertise of its in-house information technology (IT) to specialists (Terra and Gordon, 2003). In response to these needs, Päivärinta (2005) stated that ECMS encourages efficiency, effectiveness and the

required flexibility of knowledge including the reuse of previously created content, metadata, templates and navigation aids.

2.5.12 Change in work organization

Hagberg et al. (1995) defines work organization as the manner in which work is structured, distributed, processed and supervised. It basically deals with issues such as work scheduling (work-rest patterns), job design (complexity of tasks, skills required, worker control), the inter-personal aspects (relationship with managers and colleagues), career concerns (job security and opportunities for advancement), management style, and organization characteristics (corporate culture and level of corporate communications), (Carayon and Smith 2000, p.649). The implementation of the ECMS in the organization makes it easier for the issues about the work organization to be handled in a more transparent and efficient manner. Readily accessible makes it easier for management, employees to plan work and the creation of new job descriptions in the work place.

2.5.13 Cost Reduction

Understanding the cost of potential losses such as business processes and customer interactions allows an organization to see ECM systems implementation as an investment that add value and often can be measured. A way to measure the success of the ECMS implementation is to measure upfront your success based on expectations and deliver to that (Duhon 2005). “Most companies are taking a bottom-up approach to ECM strategy at present (i.e., one that focuses on delivering immediate benefits) because cost reduction is a top priority” (Smith and Mckeen 2005, p.650). ECMS tools can assist an organization to be more efficient and drive down the cost of doing business. These technologies offer value to an organization by more efficiently coordinating information for its repeat retrieval, use, and ultimately disposition. In addition, as these tools are used by more organizations, it becomes an apparent part of how to work (AIIM 2008). Päivärinta (2005) also states that the implementation of ECMS results in direct cost savings in information processing operations and facilities. Sprehe (2005) is also of the view that a fully functional ECMS, can begin to realize positive benefits and real cost savings to the information management of essential business functions.

2.5.14 Return on Investment

For instance, measuring the revenue based on improved information in the Call Centre can be done as well as measuring the cost benefits of improvements in the process for a loan application or claim in the case of a bank. Whiles figuring a direct Return on Investment (ROI) can be a complex task, it is not impossible to see the impacts of the upgraded process efficiency on the businesses (Shang and Seddon, 2002).

Table 1: A Summary of Benefits from Literatures reviewed

	Benefit	**Description**	**Sources**
1	Change of Organizational Culture	Streamlined Management Structures and Centralized Control	Davenport(1998); Junco et al(2005)
2	User Satisfaction	Increased employee morale and services provided by employees	Shang and Seddon (2002)
3	Improved Efficiency	Improved business efficiency, Ability to reuse previously created data.	Chiu and Hung (2005); Paivarinta and Munkvold (2005).
4	Improved Content Management	Reduction in content duplication, Creation of backups, Retention of content.	Kemp (2008); Huff and Dirking (2010); Sprague (19959.
5	Change of Business Processes	Upgrade in business process, Simplification of work processes	Smith and Mckeen (2003); Sprague (1995); Sprehe (2005)
6	Risk Mitigation	Implementing the appropriate risk reducing measures	Stoneburner et al. (2002) ; AIIM (2008)
7	Compliance	Adherence to rules and regulations	AIIM(2008);Duhon(2005); Sphrehe (2005) ;
8	Cost Leadership	Competitive Advantage	Hedman and Kalling (2002); Shang and Seddon (2002).
9	Collaboration	Improved knowledge creation and sharing	Duhon(2005);Paivarinta and Munkvold(2005)
10	New or Value Added Products/Services	Creation of new or value added services	Paivarinta and Munkvold (2005)
11	Improved Governance	Implementation of content management by in-house specialists	Paivarinta (2005);Terra and Gordon (2003)
12	Change in work organization	Improved work scheduling	Carayon and Smith(200); Hagberg et al.(1995)

13	Cost Reduction	Reduction in the cost of transacting business	AIIM(2008);Duhon(2005); Smith and Mckeen(2005); Paivarinta (2005); Sphrehe(2005)
14	Return on Investment	Returns on the initial investment made by the business	Shang and Seddon(2002)

Source: The Authors´ illustration

3.0 CHAPTER THREE: COMPONENTS OF ECM

Enterprise Content Management is abbreviated with the acronym ECM or sometimes ECMS where the S stands for system. In this chapter the words Enterprise, Content, and Management will be discussed more in detail. Metadata will also be discussed as this underpins the content aspect in the ECM context.

3.1 Enterprise and the Enterprise Problem

The term enterprise refers to an organization of individuals working together to achieve a common goal (Ahmed & Umrysh, 2001). Enterprises generally have some common needs, such as information management, information sharing and processing, asset management and tracking, resource management, customer or client management, and so on. The term enterprise software is used to collectively refer to all software involved in supporting these common elements of an enterprise (Ahmed & Umrysh, 2001).

ECM intends to address the needs of an entire organization rather than just the business processes of a single department (Alsup & Strong, 2004). The word "enterprise" may bring to mind large organizations operating in multiple locations with complex business processes and large numbers of content-generating applications and technologies. But as mentioned above, content management problems exists in every organization regardless of size, number of locations or business complexity.

The following figure by Alsup & Strong (2004) illustrates the typical enterprise interfaces with contractors and business partners on the left side, as well as other parties, including governments, customers and vendors on the right side. In each of these interfaces in today's modern enterprises, there is a usually a high degree of content movement and processes that are not organized or managed as well as it needs to be (Alsup & Strong, 2004).

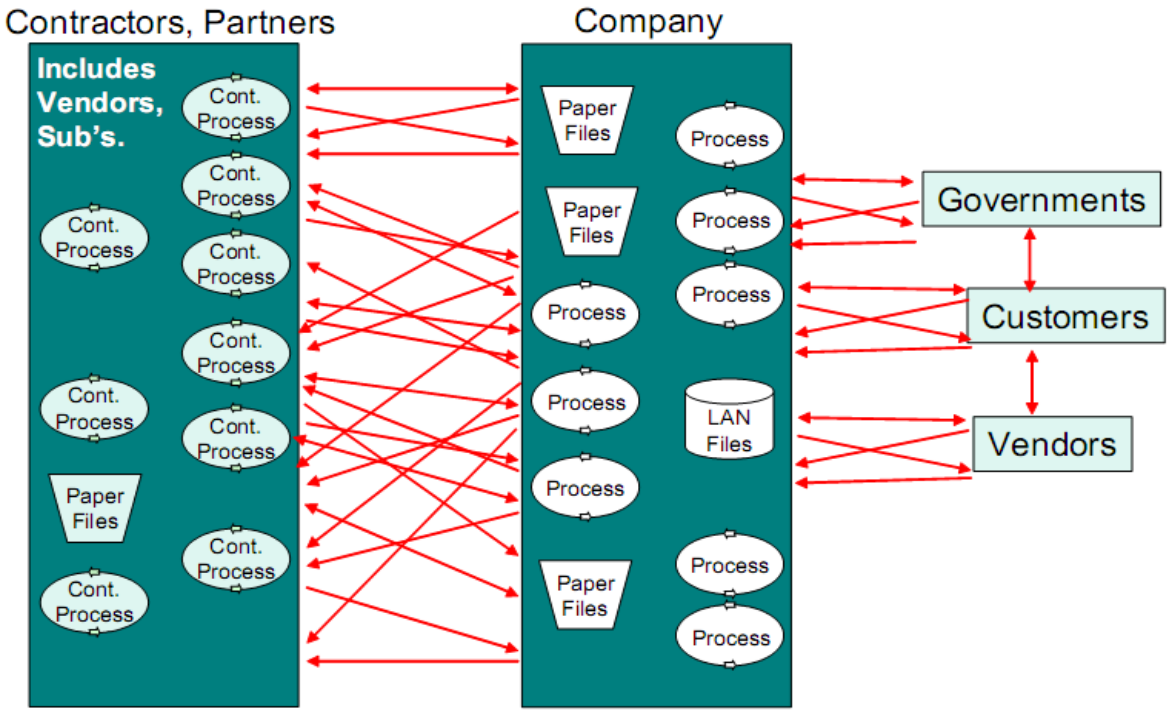


Figure 3: Typical enterprise interfaces, Alsup & Strong (2004)

Alsup & Strong (2004) explain this further:

- "Most large organizations are composed of multiple, integrated processes, many of which almost entirely depend on documents and forms.
- Documents are created and reviewed internally.
- They are received from contractors, reviewed, and approved or rejected.
- They are delivered to customers and governments based on contractual or regulatory requirements.
- There are complex document management processes related to the tracking and control of the documents on a project.
- In addition, the content being exchanged between organizations is in a variety of media, including paper, e-mail, faxes and electronic documents, which are not managed in an integrated manner." (Alsup & Strong, 2004, p. 4).

Alsup & Strong (2004) says that as a result of ineffective content management, opportunities for process improvement, risk reduction and competitive advantage are lost. Usman et al. (2009, p. 84) argues that if these complexities and challenges are not addressed in a systemic and planned manner, there is a great risk of organizational chaos and inability to meet the strategic business goals and objectives. They also go as far as saying that it threatens organizations to "seize to exist and ultimately vanishing from the face of the business world".

Clearly, as Alsup & Strong (2004) point out, there exists a potential business case for the consistent use of ECM technologies across the enterprise to ensure access and protection of critical information assets.

3.2 Content

Everyone within an enterprise is responsible for producing or consuming some sort of content (Gottlieb, 2005), and there is as previously stated usually a high degree of content movement within an enterprise (Alsup & Strong, 2004). The term "content" will be addressed further, with main focus on unstructured content.

3.2.1 What is Content

According to Tyrväinen et al. (2006) the word "content" has a number of meanings, but essentially it always refers to something contained in an entity. When comparing the word with the terms "data" or "information", they state that content clearly is associated with a container. "We talk about the content of a document, content of a Web site, or content of the Internet, among others. Content is often opposed to some other aspect of the container, for example, structure or form or representation. In an XML document, for example, we can separate content, structure, and one or more external presentations." (Tyrväinen et al., 2006, p. 628). Kampffmeyer (2004) notes that the meaning of content in the ECM scope is not unambiguous. When looking at the literature two or three types of content frequently appear. Some authors make a clear cut between structured and unstructured content, while some put weakly or semi-structured content between these two categories.

3.2.2 Structured, Unstructured and Weakly/Semi-Structured Content

Structured content is information that is standardized in layout and size, e.g. a data set where each column represents a variable and each row represents a member of the data set. This type of content is usually stored and managed in databases (Kampffmeyer, 2004). Other structures as diagram, trees and grids are also usual representations of structured content (Willenborg, 2000). This type of information is often associated with ERP systems, such as financial information or employee records. The standardized structure of the information brings possibilities for automatic analysis, classification and re-use of content (ibid), i.e. a computer can with ease interpret the information because of the standardized layout, and therefore this type of content is generally easy to manage since its format makes it easy for a system to structure it, and select relevant parts of it upon a request from the user (Kampffmeyer, 2004). Structured content is out of the focus for this report and will not be getting more attention.

A lot of information in an enterprise is contained in weakly or semi-structured content documents, such as e-mails, text documents or HTML pages (Aleksy & Schwind, 2006).This type of content is often the first type of information that comes to mind when thinking about content. This is mainly word processor files and such content can contain both layout and metadata but these are not standardized, as is the case with structured content. This can be thought of in terms of that this type of content is lexically interpretable for a system, i.e. understand the meaning of the words.

But it is relatively hard for the system to understand the semantic meaning of the information, i.e. understand the meaning of the words in that particular context. This makes weakly structured content harder to classify and also harder for the user to retrieve in comparison to structured content.

Within unstructured content resides content that is hard for a software system to even lexically interpret, and classification is therefore even harder. Examples of this type of content are images, videos, sound recordings and scanned document (Kampffmeyer, 2004). They are often stored in different ways and created individually and manually rather than automatically. Even though difficult for a system to understand, interpret and classify, it is often very powerful information for users and therefore important for a business to manage.

There is not total consensus about which of these types of content an ECM system should handle. Weakly/semi-structured and unstructured content are in literature always regarded as ECM content. But when addressing the structured content some argue, as addressed in some of the ECM definitions in chapter one, that this is not a part of ECM. As mentioned earlier, the focus of this report is how ECM can support unstructured information, and to simplify the reading, unstructured content and weakly/semi-structured content will be used under the term unstructured content further in this report. Structured content will not be the focus of this work.

3.3 Metadata

When addressing content in the ECM context it is natural to introduce metadata that recently has emerged as an important concept for those who are developing search and retrieval strategies for information (Rockley, 2003).

Traditionally, metadata has been defined as "data about data" or as "information about information" (NISO, 2004). More recent literature expresses a dissatisfaction with definitions like these, and they really do not tell much about what metadata is, its purpose and for what it can be used (Chisholm, 2008). Chisholm (2008) defines the term as follows: "metadata is the data that describes any aspect of an enterprise's information assets and enables the organization to use and manage these assets". NISO (2004) describes metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource". Others say that the most important applications for metadata are to describe information and thereby make it searchable, so that it can be retrieved and combined in meaningful ways for users (McNay, 2002). Rockley (2003) says that this could be regarded as labeling, cataloguing and describing information that allows the content elements to be properly processed and searched by a computer. Near (2010) argues that metadata is critical to any ECM system, and that metadata is "how you file your content, how you find your content, it defines the processes that apply to your content, and it is how you ensure that your content is managed in a compliant way."

Rockley (2003) also states that metadata can be used to describe the behavior, processes, rules and structure of the data, not just descriptive information, and that these elements are important when developing a metadata strategy for content search and retrieval which is part of ECM, because they determine not only what the content is, but who uses it, how it will be used, how it will be delivered,

and when. The next section explains the usefulness of metadata to the three activities regarding content stored in a data system.

3.3.1 Metadata for Reuse, Retrieval and Tracking

Rockley (2003) says that metadata for reuse can be particularly useful in a content strategy, eliminating content redundancies. In this case, metadata is applied to each content element. Thereafter, authors can search for elements before beginning to write, to see if they already exist somewhere in another document stored in the ECM system. Metadata for reuse could include: content type, where the content should appear, creation date, content owner, keywords and links to where content is already used. Some metadata for reuse is applied automatically, based upon the document definition, (e.g., type of content), while other metadata is added by the author (e.g., keywords) (Rockley, 2003).

Metadata for retrieval enables content to be retrieved through searching. Rockley (2003) says that metadata for retrieval can include much of the same metadata one defines for reuse, but is usually much more extensive. It can include metadata such as: title, author, date (creation, completion, and modification), keywords, responsible party, security status, and tracking (e.g., status). Metadata for retrieval enables users to specifically define which content elements they want to view (Rockley, 2003).

Further, Rockley (2003) says that metadata for tracking is particularly useful when you are implementing workflow as part of a content strategy. By assigning status metadata to each content element, one can determine which elements that are active, control what can be done to an element, and who can do it. Generally, status changes based on the metadata, are controlled through workflow automation, not by end users (Rockley, 2003). Status metadata can include: draft (under development by the author), draft for review, reviewed, approved, final, and submitted.

3.3.2 Defining Metadata and Metadata Categories

Rockley (2003) states that properly defining and categorizing the types of metadata that an organization wants to capture about the organization's information is extremely important to the success of the metadata strategy. "Improperly identified metadata, or missed categories of metadata, can cause problems ranging from misfiled and therefore inaccessible content, to more serious problems such as those encountered by the National Aeronautics and Space Administration's (NASA's) 1999 Mars Climate Orbiter mission, in which misidentified metadata resulted in the loss of the spacecraft, at a cost of $300 million!" (Rockley, 2003, p. 13).

Rockley (2003) points out that first of all, it should be determined if the metadata is being defined for retrieval, reuse, or tracking. Then one has to understand the end business results that the business is trying to achieve, in order to build the metadata backwards to achieve that result (Rockley, 2003). By properly defining metadata that a business needs, Rockley (2003) states that this helps to make sure that the right information is delivered to the right person, for the right reason, at the right time.

3.3.3 Metadata in Practice

After nearly a decade of consulting customers in the ECM market, Near (2010) reports that organizations experience both joy and pain of managing metadata. When large organizations embark on an ECM project he opines that metadata will be the point where the various parts of the organization begin to reveal how very different they are. And that the reality of an enterprise is that every stakeholder group has different needs, that they all use content in different ways, and that they all have some unique piece of information that they absolutely need to track. This shows that the use of metadata is complex (Near, 2010).

Enforcing standard metadata classifications gives users a common language to describe and find content (Wilkoff et al., 2001). However, Wilkoff et al. (2001) say that building a standard taxonomy that spans all enterprise content is a daunting task. They suggest that firms should take a practical approach by focusing first on the content that matters most. As a starting point, content administrators should identify the top-three customer and end user goals and tag the content that facilitates those scenarios.

Banerjee (2000, p.217) discusses many of the challenges that were encountered during a project designed to explore "using metadata as well as other tools to reduce the need to manually create records for electronic resources". The results suggest that metadata must be entered consistently if they are to be useful. This involves what sort of metadata that belongs in the various fields. To achieve consistent entry of metadata, Banerjee (2000) states that people have to agree upon what type of information that belongs in the metadata and what they hope to accomplish by putting it there. Attaining this agreement is extremely difficult as a practical matter according to Banerjee (2000). Different user communities have diverse interests, and metadata typically is defined in terms that reflect the interests of a particular user community (Vellucci, 1997 cited in Banerjee, 2000).

Further Banerjee (2000) reports that adding consistent metadata requires significant time and effort. Although some metadata such as date of creation may be generated automatically by the system in use, only humans can identify many important relationships in data. Banerjee (2000) also reports that there was a great deal of variability in how users filled out the submission form with metadata values. Because of this inconsistent use of metadata, some felt that the value of the fields was questionable. Another problem that was identified was that people supplied metadata for different reasons. Milstead (1993) cited in Banerjee (2000) informs that some people may be more interested in achieving their own financial or emotional objectives than in helping other users to find what they need.

Banerjee (2000) also report that it was difficult to find local expertise who could dedicate time to support the project, and with the specialized skills necessary to evaluate, implement, and maintain systems that exploit metadata. Finally, Banerjee (2000, p.223) notes that "metadata is a tool, not a solution to problems. Just as high tech woodworking and machine tools are potentially useless (or even harmful) in the hands of unskilled people, people need to understand what metadata does and develop certain skills to make use of it".

3.4 Management of Content

Once the terms enterprise, content and metadata is addressed, it is time to consider how to handle this content within the concept of ECM. Based on the ECM definitions provided for in chapter one, this part is normally divided into five distinctive tasks: capture, manage, store, preserve, and deliver. Sometimes other terms and activities are included but the meaning is in essence the same.

Chieu et al. (2008) asserts that content management essentially deals with human workflow that involves the lifecycle management and exchange of business documents among organizations and users. They further explain that in such document exchange workflow, a business document typically starts from an initial "draft" created by a submitter. It then goes through numerous rounds of reviews, changes and validation before it gets the approval and signatures and is sent to its final dealing parties. The dealing parties on the other hand may have to go through a similar process to review and approve the business document. The number of intermediate steps during the document lifecycles depends on the complexity of the business process among the involved parties. Chieu et al. (2008) states that without a content management system to handle this lifecycle, the various workflow steps are typically carried via traditional fax or e-mails, thus imposing delays and inefficiencies with the manual processes.

There are many models of information or content lifecycle management. One model by Kampffmeyer (2006) is illustrated below. These phases will briefly be addressed next, based on Kampffmeyer (2006):-

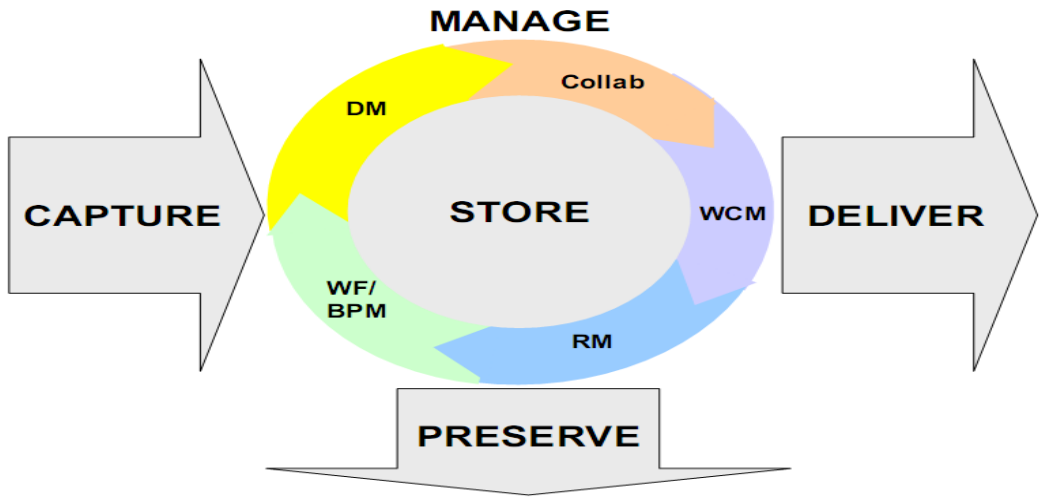


Figure 4: The 5-component-model of ECM, Kampffmeyer (2006)

Capture

This part refers to the process where new content is collected, generated, or created within the enterprise. This content has to be identified and classified with metadata in some way into the ECM system that will house and manage it. Content items may be unstructured such as addressed earlier.

Manage

The Manage components are for the management, processing, and use of information. This incorporates database for administration and retrieval, and access authorization systems for protection of information. Kampffmeyer (2006) notes that the goal of an ECM system is to provide these two components just once as services for all "Manage" solutions such as Document Management, Collaboration, Web Content Management, Records Management and Workflow / Business Process Management. To link the various "Manage" components, an ECM system should have standardized interfaces and secure transaction processes.

Store

"Store" components are used for the temporary storage of information which it is not required or desired to archive.

Preserve

The "Preserve" components of ECM handle the long-term, safe storage and backup information. In some cases, content such as business records may need to be preserved for long periods of time in a trustworthy and accurate manner.

Deliver

This involves providing access and presentation of content from the location it is stored or preserved in a timely and secure manner, to the systems and people who have access to it. This is supported by the metadata associated with the content.

4.0 CHAPTER FOUR: ECM SOFTWARE

This chapter introduces the popular ECM software's that are available and gives an overview of Alfresco, the open-source ECM software used for the project. A brief introduction to open source versus commercial ECM solutions are also discussed. The chapter also explains the concept of open-source and proprietary software.

4.1 Open-Source Software

Open source software is typically developed, marketed and distributed by a loosely organized community of individuals. The term "open source" means that the source code itself, the instructions that cause the application to do what it does, can be easily viewed, modified or downloaded by anyone with technical expertise.

Open source software is sometimes called Free software or FOSS (Free and Open Source Software), because of four tenets of freedom that are a core part of the philosophy of the open source movement. First, you are free to run these software packages for any purpose--you generally don't pay anything to acquire them. Second, the source code is free--you can see the code and understand how it works. Third, you are free to copy and redistribute the package to anyone you want, and finally, you are free to modify the software however you like, and to release those modifications.

4.2 Commercial / Proprietary Software

Commercial Software refers to any software that is designed for sale to serve a commercial need. Commercial software is usually proprietary software.

Proprietary software or **closed-source software** is software, where the developers or distributors reserve all freedoms and rights.

Among the freedoms and rights that proprietary software deprives (to end-users), are:

- the freedom to analyze the software, and to change it (often deprived through intentional non-availability of source code, or through Non-disclosure agreements (NDA))
- the freedom to share the software (often deprived through copy prohibition via EULA (End User License Agreement) or NDA)
- the freedom to run the software for any purpose (often deprived through user-restrictions via EULA).

In contrast to proprietary software, free software, is software that grants a user *all* these freedoms, on reception of the software.

Proprietary software is licensed under legal right of the copyright holder, with the intent that the licensee is given the right to use the software *only* under certain conditions, and restricted from other uses, such as modification, sharing, studying, redistribution, or reverse engineering. Usually the source code of proprietary software is not made available.

4.3 Proprietary Enterprise Content Management Software

4.3.1 Documentum

EMC Documentum Platform: Critical enterprise content management access and control of all information assets. The Documentum platform offers an enterprise content management system with the essential capabilities to control all information content.
Documentum Standard provides a robust, scalable, fault-tolerant, and extensible architecture to manage and control all information content.
Control repository access and improve compliance through fine-grained authentication, authorization, and auditing
Support diverse environments using open standards such as Content Management Interoperability Service (CMIS) within a service-oriented architecture.

4.3.2 SharePoint

SharePoint is a web application platform offering a set of tools that can be used to provide intranet portals, document and file management, collaboration, and social networks. The SharePoint platform also offers many more process and workflow capabilities.

SharePoint is a web application platform in the Microsoft Office server suite. Launched in 2001, SharePoint combines various functions which are traditionally separate applications: intranet, extranet, content management, document management, personal cloud, enterprise social networking, enterprise search, business intelligence, workflow management, web content management, and an enterprise application store. SharePoint servers have traditionally been deployed for internal use in mid-size businesses and large departments alongside Microsoft Exchange, Skype for Business, and Office Web Application Server; but Microsoft's 'Office 365' software as a service offering (which includes a version of SharePoint) has led to increased usage of SharePoint in smaller organizations.

While Office 365 provides SharePoint as a service, installing SharePoint "on-premise" typically requires multiple virtual machines, at least two separate physical servers, and is a somewhat significant installation and configuration effort. The software is based on an n-tier service oriented architecture. Enterprise application software (for example, email servers, ERP, BI and CRM products) often either requires or integrates with elements of SharePoint. As an application platform, SharePoint provides central management, governance, and security controls. The SharePoint platform manages Internet Information Services (IIS) via farm-based management tooling.

4.4 Open Source Enterprise Content Management Software

4.4.1 The Nuxeo Platform

The Nuxeo Platform is a highly customizable and extensible set of content management components, services and APIs that has been designed for content-centric application development. It is based on modern Java EE technologies, highly modular thanks to its OSGi plugin architecture, and provides most of the services expected for ECM

Designed by developers for developers, the Nuxeo Platform is a highly customizable and extensible content management platform for building business applications. The Nuxeo Platform can be used to develop business solutions such as document management, digital asset management, case management, platform for SAAS applications, advanced knowledge base and application development.

4.4.2 OpenKM

OpenKM is a web base document management application that uses standards and Open Source technologies. OpenKM provides full document management capabilities including version control and file history, metadata, scanning, workflow, search, and more. It also allows the social activities around content to be used to connect people to other people, information to information, and people to information; helping to manage, more efficiently, the collective intelligence of the human resources of the company.

OpenKM integrates all essential document management, collaboration and advanced search functionality into one easy to use solution.

4.5 The Alfresco Platform

Alfresco is an open source Enterprise Content Management (ECM) system that manages all the content within an enterprise and provides the services and controls that manage this content.

At the core of the Alfresco system is a repository supported by a server that persists content, metadata, associations, and full text indexes. Programming interfaces support multiple languages and protocols upon which developers can create custom applications and solutions. Out-of-the-box applications provide standard solutions such as document management, and web content management.

As an entirely Java application, the Alfresco system runs on virtually any system that can run Java Enterprise Edition. At the core is the spring platform, providing the ability to modularize functionality, such as versioning, security, and rules. Alfresco uses scripting to simplify adding new functionality and developing new programming interfaces. This portion of the architecture is known as web scripts and can be used for both data and presentation services. The lightweight architecture is easy to download, install, and deploy.

Figure 5: Alfresco Architecture (Alfresco, 2015)

Alfresco Client Applications

Alfresco client applications provide a means of accessing the Alfresco content repository. They provide an interface that allows the user to create, upload, and manage content. Alfresco provides Share, which is a web-based client application.

The main Alfresco web client application is Share. The user interface is built entirely with the Alfresco web script technology, which can be used to extend the application. Share provides content management capabilities with simple user interfaces, tools to search and browse the repository, content such as thumbnails and associated metadata, previews using Flash renditions of content, and a set of collaboration tools such as wikis, discussions, and blogs. Alfresco Share is organized as a set of sites that can be used as a meeting place for collaboration. Alfresco Share is a web-based application that can be run on a different server to the server that runs the Alfresco repository, thus providing opportunities to increase scale and performance.

Content Applications

Alfresco can be used for building most ECM applications. Aside from the major applications such as document, image, records, digital asset, and web content management, there are a number of specific applications and use cases that add value to the enterprise.

The following are typical ECM applications that can use Alfresco applications as the foundation. Applying programming models lets you extend these applications or you can build your own applications using Alfresco.

- **Document management** manages and shares office documents, and incorporates business processes. It can be industry or role-specific. The Alfresco Share is a good foundation for building document management applications.

- **Records management** controls important information for retention over time. You would use records management over document management in regulated or compliant environments, such as in managing governmental information or personnel records, or where information is audited. Alfresco is certified to the U.S. Government 5015.2 records standard and is useful for controlling retention and review periods, providing specialized security, and determining whether the records are archived or destroyed after a specified period of time.

- **Shared drive replacement** is a more basic form of document management in the enterprise with a content repository that provides easy access points to content. Shared drives are simple to use because users do not need to be trained and all applications work with them. Because Alfresco supports the protocol used by shared drives, Common Internet File System (CIFS), the repository appears to be a shared drive. With rules, actions, and extensions, you can build complete document management applications that are transparent to users while getting the content under control and enabling it to be searched.

- **Enterprise portals and intranets** communicate with employees with news and developments in the enterprise. While part of enterprise portals focus on reporting and analyzing data, many are devoted to content and documents. Although folder hierarchies are an easy way to organize information for a portal, classifications and metadata are often a better way to target information in the portal to end users. Thus, there are elements of document management and business process for delivering into the portal; however, the presentation of lists of content and navigation through classifications require programming portlets using web scripts or Java. Portlets provided as part of the Alfresco platform can supplement this development with standardized navigation, search, and content presentation.

- **Web content management** manages websites, the content that goes into websites (such as HTML and images), and the processes of building, testing, and deploying websites and content. While Alfresco can be used for simple websites, it is frequently used for creating websites that are web applications, particularly those developed using Java. Some examples of these websites publish a lot of information from multiple sources and integrate e-commerce and back office systems. Surf is a good platform for creating these types of web applications and websites.

- **Knowledge management** captures knowledge from employees or customers and provides it in a form that others can use. Content tends to be the best and most reusable container of knowledge in sharing that knowledge with others.

- **Information publishing** encompasses real-time publishing of content from different sources to the website and the deployment of that content to the web farm for Internet access. This can be digital assets such as articles, written internally or syndicated from other sources, or photos. Media companies use Alfresco to combine this content and publish it to their websites. This straight-through publishing of information requires both strong content control and performance to aggregate and push out the content.

- **Case management** handles information related to a case, such as an insurance claim, an investigation, or personnel processing. Alfresco's document management capabilities, folder structure, classification schemes, and workflow is well suited to managing cases and distributing work in handling cases. Alfresco incorporates the Activiti workflow engine and can handle sophisticated workflows and queue management. Alfresco has a content-oriented task model that aggregates all the resources required to perform specific tasks within the case handling process.

At the core of the Alfresco system is a repository supported by a server that persists content, metadata, associations, and full text indexes. Programming interfaces support multiple languages and protocols upon which developers can create custom applications and solutions. Out-of-the-box applications provide standard solutions such as document management, and web content management.

Alfresco architecture

As a Java application, the Alfresco system runs on virtually any system that can run Java Enterprise Edition. At the core is the Spring platform, providing the ability to modularize functionality, such as versioning, security, and rules. Alfresco uses scripting to simplify adding new functionality and developing new programming interfaces. This portion of the architecture is known as *web scripts* and can be used for both data and presentation services. The lightweight architecture is easy to download, install, and deploy.

There are many ways to deploy Alfresco, however most deployments follow a general pattern. Ultimately, Alfresco is used to implement ECM solutions, such as Document Management (DM), Web Content Management (WCM), and Digital Asset Management (DAM). There can also be elements of collaboration and search across these solutions.

The solutions are typically split between clients and server, where clients offer users a user interface to the solution and the server provides content management services and storage. Solutions commonly offer multiple clients against a shared server, where each client is tailored for the environment in which it is used.

Clients

Alfresco offers a web-based client called Alfresco Share, built entirely with the Alfresco web script technology. Share provides content management capabilities with simple user interfaces, tools to search and browse the repository, content such as thumbnails and associated metadata, previews, and a set of collaboration tools such as wikis and discussions. Alfresco Share is organized as a set of sites that can be used as a meeting place for collaboration. Alfresco Share is a web-based application that can be run on a different server to the server that runs the Alfresco repository, thus providing opportunities to increase scale and performance.

Alfresco Share can be deployed to its own tier separate from the Alfresco content application server. Alfresco Share focuses on the collaboration aspects of content management and

streamlining the user experience. Alfresco Share is implemented using Alfresco Surf and can be customized without JSF knowledge.

Clients also exist for portals (by using JSR-168 portlets), mobile platforms, Microsoft Office, and the desktop. In addition, using the folder drive of the operating system, users can share documents through a network drive. Using JLAN technology, Alfresco can look and act just like a folder drive. JLAN is the only Java server-side implementation of the CIFS protocol, letting users interact with Alfresco as they do any other normal file drive except the content is now stored and managed in the Alfresco content application server.

Server

The Alfresco content application server comprises a content repository and value-added services for building ECM solutions.

The Alfresco content application server provides the following categories of services built upon the content repository:

- Content services (transformation, tagging, metadata extraction)
- Control services (workflow, records management, change sets)
- Collaboration services (social graph, activities, wiki)

Clients communicate with the Alfresco content application server and its services through numerous supported protocols. HTTP and SOAP offer programmatic access while CIFS, FTP, WebDAV, IMAP, and Microsoft SharePoint protocols offer application access. The Alfresco installer provides an out-of-the-box prepackaged deployment where the Alfresco content application server and Alfresco Share are deployed as distinct web applications inside Apache Tomcat.

Guiding Design Principles

The Alfresco architecture supports the requirements of Enterprise Content Management (ECM) applications, such as Document Management (DM), Web Content Management (WCM), Records Management (RM), Digital Asset Management (DAM), and Search.

Alfresco Web tier and Surf

Alfresco provides ECM capabilities as data services, user interfaces, and user applications. The user interface capabilities are provided by applications and application components using Alfresco's web tier, Surf, originally developed as a faster way to develop content applications using scripting and REST architecture.

Development of web scripts allows for the creation of a REST-based API. Web scripts can be executed without compilation, and provide a quick and easy way to extend and enhance Alfresco's standard services.

The web script infrastructure accommodates Java beans as easily as JavaScript. Web scripts add little overhead but provide a great deal of flexibility and development productivity. Web scripts in the web tier let you quickly build user interface components with Surf or simple HTML and deploy them as Alfresco Share components, portlets, or other web platforms such as Google Gadgets.

Alfresco content application server

At the heart of the Alfresco system is the content application server, which manages and maintains the content repository. The server's primary responsibility is to provide services for use in building ECM solutions. All the applications of the Alfresco ECM suite are built upon and executed by the content application server.

The Alfresco content application server exposes a set of remote public interfaces for allowing a client to communicate with it. The remote public interfaces are the only part of the server visible to the client. There are two types:

- **Remote APIs** - for interacting with services of the server programmatically
- **Protocol bindings** - for mapping services for use by a protocol-compliant client

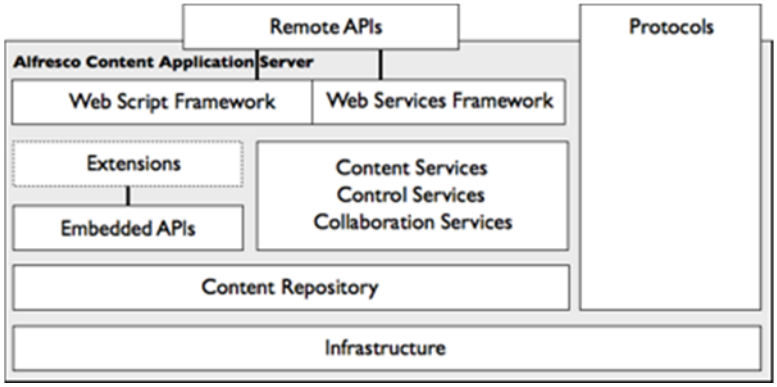


Figure 6: Alfresco content application server

Internally, the server comprises several layers. The foundation includes infrastructure concerns, such as configuration, authentication, permissions, and transactions that cut across all capabilities. Infrastructure also shields the server from being tied to any specific environmental implementation, such as transaction managers or caching mechanisms.

The content repository is built on this infrastructure, which itself is the building block for content, control, and collaboration services. Each capability of the content repository and content services is individually bundled as a module with its own in-process interface and implementation. Modules are bound together by the infrastructure through their interfaces.

You can deploy extensions to the content application server to extend or override its capabilities. Their implementation might use the in-process interfaces offered by the content repository and content services.

Alfresco content repository

The content repository is comparable to a database, except that it holds more than data. The binary streams of content are stored in the repository and the associated full-text indexes are maintained by SOLR indexes.

The actual binary streams of the content are stored in files managed in the repository, although these files are for internal use only and do not reflect what you might see through the shared drive interfaces. The repository also holds the associations among content items, classifications, and the folder/file structure. The folder/file structure is maintained in the database and is not reflected in the internal file storage structure.

The Alfresco content repository implements services including:

- Definition of content structure (modeling)
- Creation, modification, and deletion of content, associated metadata, and relationships
- Query of content
- Access control on content (permissions)
- Versioning of content
- Content renditions
- Locking
- Events
- Audits
- Import/Export
- Multilingual
- Rules/Actions

The Alfresco content repository implements and exposes these services through an Alfresco API, CMIS protocol bindings, and the JSR-170 Java API. The storage engine of the repository stores and retrieves content, metadata, and relationships, and operates on the following constructs:

- **Nodes** - provide metadata and structure to content. A node can support properties, such as author, and relate to other nodes such as folder hierarchy and annotations. Parent to child relationships are treated specially.
- **Content** - the content to record, such as a Microsoft Word document or an XML fragment.

Content models are registered with the content repository to constrain the structure of nodes and the relationships between them, and to constrain property values.

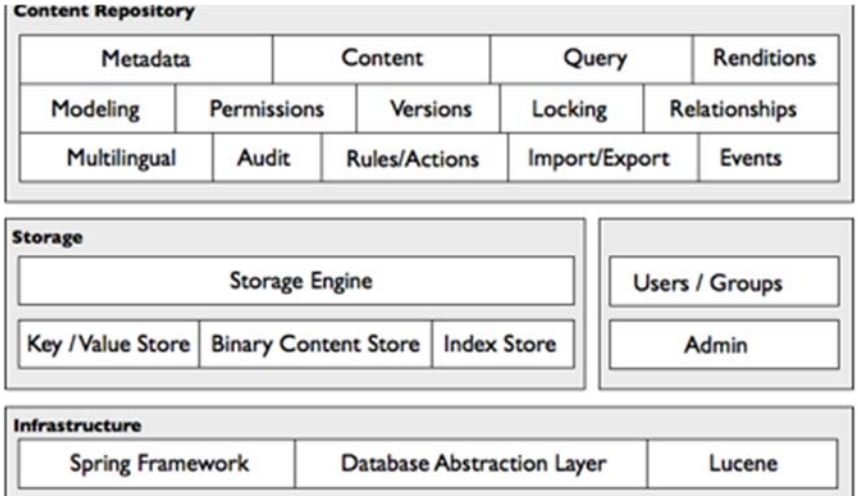


Figure 7: Alfresco content repository

The storage engine also exposes query capabilities provided by a custom query engine built on Apache Lucene that supports the following search constructs:

- Metadata filtering
- Path matching
- Full text search
- Any combination of these search constructs

The query engine and storage engines are hooked into the transaction and permission support of the infrastructure, offering consistent views and permission access. Several query languages are exposed, including native Lucene, XPath, Alfresco FTS (Full Text Search), and CMIS Query Language (with embedded Alfresco FTS).

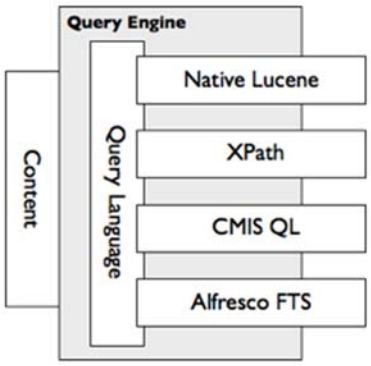


Figure 8: Alfresco query engine

By default, Alfresco stores nodes in an RDBMS while content is stored in the file system. Using a database provides transaction support, scaling, and administration capabilities. Alfresco uses a database abstraction layer for interacting with the database, which isolates the storage engine from variations in SQL dialect. This eases the database porting effort, allowing the certification of Alfresco against all the prominent RDBMS implementations. The file system stores content to allow for very large content, random access, streaming, and options for different storage devices. Updates to content are always translated to append operations in the file system, allowing for transaction consistency between database and file system.

You can bundle and deploy the Alfresco content repository independently or as part of a greater bundle, such as the Alfresco content application server.

Alfresco content services

Services address the core use cases for content management applications including the logical organization of content, file management, version control, and security. Services also support the control of content through workflow and process management, and social and collaborative applications.

Alfresco exposes services at various levels including:

- Java
- Scripting
- REST
- Web services
- Client interfaces, such as Alfresco Share

Some services are considered internal; others are public. For example, the Java level services are internal. The majority of these are accessible through other public interfaces including the public APIs, client applications, and CMIS. Alfresco services are divided into two main categories; content application services and content repository services.

- **Content repository services** Content repository services are the fundamental services for working with content in Alfresco. Content repository services are written in Java.

- **Content application services** Content application services extend the repository services, providing extended capabilities for rich content and collaborative applications.

Alfresco programming models

A number of programming models are available for building an application using the Alfresco content application server.

- The simplest model for non-programmers is to use out-of-the-box components of the Alfresco Share application and the Rules and Actions model, a set of conditions and actions to take on content based on those conditions. You can define rules and actions using a wizard and perform actions such as converting content, moving content, or executing a simple JavaScript snippet.

- Web scripts let you perform more sophisticated processing without complex programming. The Alfresco Content Management Interoperability Services (CMIS) implementation was built using web scripts. By using JavaScript to build these data services, it is easy to create new services in Alfresco. To build new user interfaces or extensions to Alfresco Share, you can also use web scripts by using a web templating language like FreeMarker. Most of Alfresco Share was built using web scripts.

- To use Java to build applications or extend Alfresco Share, you can use the many tools associated with Java that were used to build the Alfresco system. Surf, the web runtime framework, lets you extend Alfresco Share and build web applications. Because Alfresco Share was built using Surf, you can build your own extensions as a combination of Java programming and web scripts, or with Java alone. You can also use Java to access or even replace whole pieces of Alfresco, content application server, or Alfresco Share by using the Spring platform. You can use the source code as an example for rewriting pieces and using Spring beans and configuration to extend or replace functionality in Alfresco.

- To write applications that use Alfresco but are portable to other ECM systems, you can use Content Management Interoperability Services (CMIS), the OASIS standard for accessing content repositories.

APIs

To access and extend Alfresco out-of-the-box services, the Alfresco content application server exposes two flavors of API, each designed for a specific type of client.

The two main categories of API are embedded and remote APIs.

- **Embedded APIs:** The Embedded API is used for developing extensions to the Alfresco content application server. Extensions deployed into the server often depend on existing services provided by the server. Therefore, developers of extensions use the Embedded API to gain access to those services.

- **Remote APIs:** The Remote API is primarily used to build ECM solutions against the Alfresco content application server.

Content modeling

Content modeling is a fundamental building block of the Alfresco content repository that provides a foundation for structuring and working with content. Content modeling specifies how nodes stored in the content repository are constrained, imposing a formal structure on nodes that an ECM application can understand and enforce. Nodes can represent anything stored in the repository, such as folders, documents, XML fragments, renditions, collaboration sites, and people. Each node has a unique ID and is a container for any number of named properties, where property values can be of any data type, single or multi-valued.

Nodes are related to each other through relationships. A parent/child relationship represents a hierarchy of nodes where child nodes cannot outlive their parent. You can also create arbitrary

relationships between nodes and define different types of nodes and relationships. A content model defines how a node in the content repository is constrained. Each model defines one or more types, where a type enumerates the properties and relationships that a node of that type can support. Often, concepts that cross multiple types of node must be modeled, which the Alfresco content repository supports through aspects. Although a node can only be of a single type, you can apply any number of aspects to a node. An aspect can encapsulate both data and process, providing a flexible tool for modeling content.

Content modeling puts the following constraints on the data structure:

- A node must be of a given kind.
- A node must carry an enumerated set of properties.
- A property must be of a given data type.
- A value must be within a defined set of values.
- A node must be related to other nodes in a particular way.

These constraints allow the definition (or modeling) of entities within the ECM domain. For example, many ECM applications are built around the notion of folders and documents. It is content modeling that adds meaning to the node data structure.

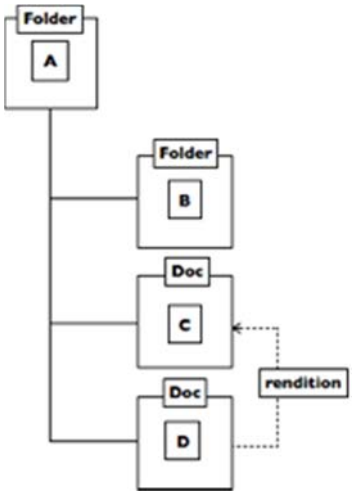


Figure 9: Alfresco node data structure

The Alfresco content repository provides services for reading, querying, and maintaining nodes. Events are fired on changes, allowing for processes to be triggered. In particular, the content repository provides the following capabilities based on events:

- Policies: event handlers registered for specific kinds of node events for either all nodes or nodes of a specific type
- Rules: declarative definition of processes based on addition, update, or removal of nodes (for example, the equivalent of email rules)

Models also define kinds of relationships, property data types, and value constraints. A special data type called content allows a property to hold arbitrary length binary data. Alfresco comes prepackaged with several content models. You can define new models for specific ECM use cases from scratch or by inheriting definitions from existing models.

- Content metadata model: Content modeling is about metadata; data describing data. Content modeling in Alfresco refers to a four-layer metadata model.

- Deploying a content model: A content model is defined in its entirety as a single XML document that must comply with the content metamodel XSD schema provided by the Alfresco content repository. Each model contains a set of related and coherent definitions and is deployed as a unit.

- Content metamodel: Defining a content model is expressed in terms of the content metamodel.

- Out-of-the-box-model: The content repository comprises several content models out of the box for specifying the core content types of an ECM system. They are expressed in terms of the content metamodel and provide a set of samples on which to base custom content models.

- Localizing models in Alfresco share: Every type, aspect, property, association, constraint, and data type defined in a model has a title and description. Both of these values are provided in the model XML file but only one language is supported, which is the language of the values specified in the XML file.

- Alfresco content models and CMIS: CMIS defines a data model, which encapsulates the core concepts found in most content repositories. Alfresco provides an implementation of the CMIS bindings and maps the Alfresco content metamodel to the CMIS domain model. This allows content models defined in Alfresco to be exposed and manipulated by using CMIS.

Protocols

The Alfresco content application server supports many folder and document-based protocols to access and manage content held within the content repository using familiar client tools.

All the protocol bindings expose folders and documents held in the Alfresco content repository. This means a client tool accessing the repository using the protocol can navigate through folders, examine properties, and read content. Most protocols also permit updates, allowing a client tool to modify the folder structure, create and update documents, and write content. Some protocols also allow interaction with capabilities such as version histories, search, and tasks.

Internally, the protocol bindings interact with the content repository services, which encapsulate the behavior of working with folders and files. This ensures a consistent view and update approach across all client tools interacting with the Alfresco content application server.

An Alfresco subsystem for file servers allows configuration and lifecycle management for each of the protocols either through property files or JMX.

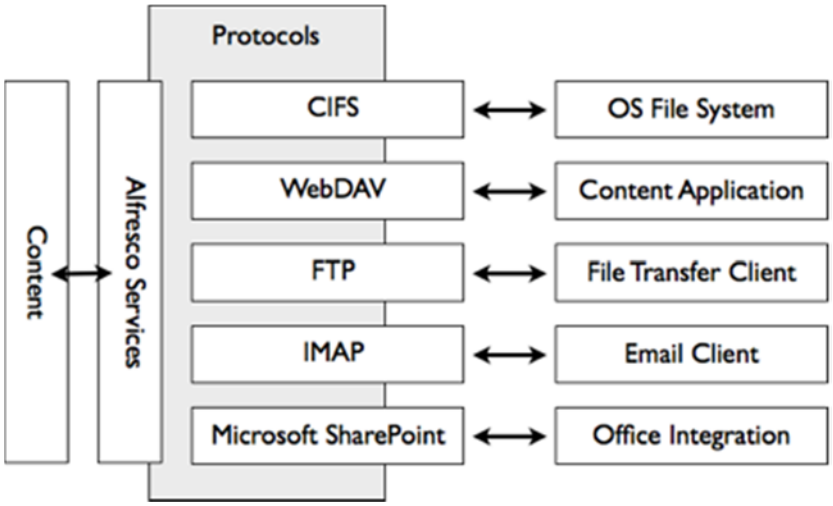


Figure 10: Alfresco supported protocols

Supported protocols include:

CIFS (Common Internet File System)

CIFS allows the projection of Alfresco as a native shared file drive. Any client that can read and write to file drives can read and write to Alfresco, allowing the commonly used shared file drive to be replaced with an ECM system without users even knowing.

WebDAV (Web-based Distributed Authoring and Versioning)

WebDAV provides a set of extensions to HTTP for managing files collaboratively on web servers. It has strong support for authoring scenarios such as locking, metadata, and versioning. Many content production tools, such as the Microsoft Office suite, support WebDAV. Additionally, there are tools for mounting a WebDAV server as a network drive.

FTP (File Transfer Protocol)

FTP is a standard network protocol for exchanging and manipulating files over a network. This protocol is particularly useful for bulk loading folders and files into the Alfresco content repository.

IMAP (Internet Message Access Protocol)

IMAP is a prevalent standard for allowing email access on a remote mail server. Alfresco presents itself as a mail server, allowing clients such as Microsoft Outlook, AppleMail, and Thunderbird to connect to and interact with folders and files held within the Alfresco content repository. IMAP supports three modes of operation:

1. Archive: allows email storage in the Alfresco content repository by using drag/drop and copy/paste from the IMAP client

2. Virtual: folders and files held in the Alfresco content repository are exposed as emails within the IMAP client with the ability to view metadata and trigger actions using links embedded in the email body

3. Mixed: a combination of both archive and virtual

Microsoft SharePoint Protocols

Alfresco Office Services support Microsoft SharePoint Protocols. This allows Alfresco to act as a SharePoint server, creating tight integration with the Microsoft Office suite. A user who is familiar with the Microsoft task pane can view and act upon documents held within the Alfresco content repository. Collaborative features of Microsoft SharePoint are mapped to Alfresco Share site capabilities.

Modularity

The Alfresco system is modular. Every moving part is encapsulated as a service, where each service provides an external face in a formally defined interface and has one or more black-box implementations.

The system is designed this way to allow for:

- Pick and mix of services for building an ECM solution
- Reimplementation of individual services
- Multiple implementations of a service, where the appropriate implementation is chosen based on the context within which the solution is executed
- A pattern for extending Alfresco (at design and runtime)
- Easier testing of services

To support this approach, Alfresco used the Spring framework for its factory, Dependency Injection, and Aspect-Oriented Programming (AOP) capabilities. Services are bound together through their interfaces and configured using Spring's declarative Dependency Injection.

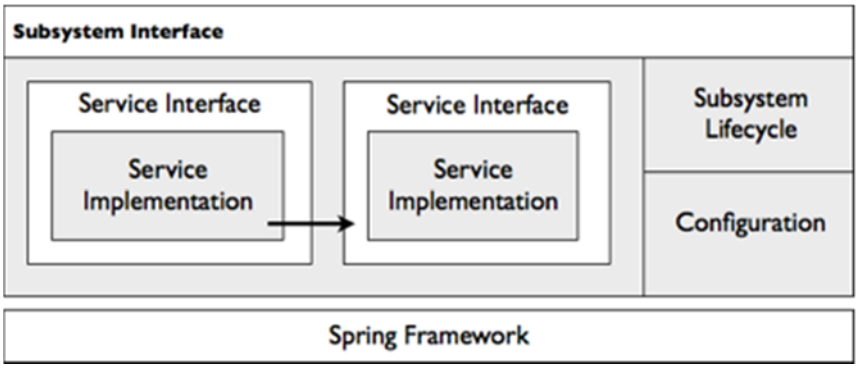


Figure 11: Spring Framework

A service interface is defined as a Java interface. For services that form the internal embedded API for extensions, cross-cutting concerns such as transaction demarcation, access control, auditing, logging, and multi-tenancy are plugged in through Spring AOP behind the service interface. This means that service implementations are not polluted with these concerns. It also means the cross-cutting concerns can be configured independently or even switched off across the server if, for example, performance is the top-most requirement and the feature is not necessary.

Multiple services are aggregated into an Alfresco subsystem where a subsystem represents a complete coherent capability of the Alfresco server, such as authentication, transformation, and protocols. As a unit, subsystems have their own lifecycle where they can be shut down and restarted while the Alfresco server is running. This is useful to disable aspects of the server, or reconfigure parts of it, such as how LDAP synchronization is mapped. Each subsystem supports its own administration interface that is accessible through property files or JMX.

Alfresco web application framework

Alfresco Share and all new Alfresco web applications are built on Alfresco Surf. This web application framework provides the typical features of this kind of framework and supports web content management needs.

At the heart of Alfresco Surf is a site assembly framework that bundles a full site construction object model and toolkit for building websites and applications.

Its features include:

- A Site Dispatcher to create pages easily, link them to the overall navigation of a website, and build pages in a way that promotes reusability.
- Templates for defining a page layout once and then reusing it across a large set of pages. You can develop pages using FreeMarker, JSP, HTML, or Java.

- A UI Library containing reusable UI components comprising back-end application logic and front-end presentation code that can be bound into regions (or slots) within a page or template.
- Pages that you can render in multiple formats, such as print, PDF, or mobile device.
- AJAX support for integration with the Yahoo! User Interface (YUI) library.
- Forms in a rich forms engine for rendering and collecting data.

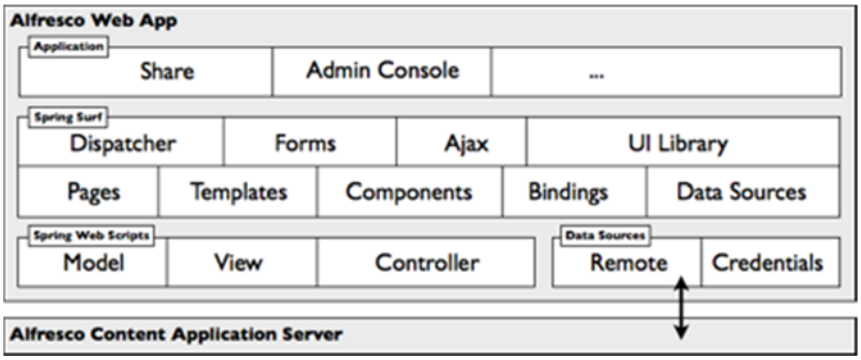
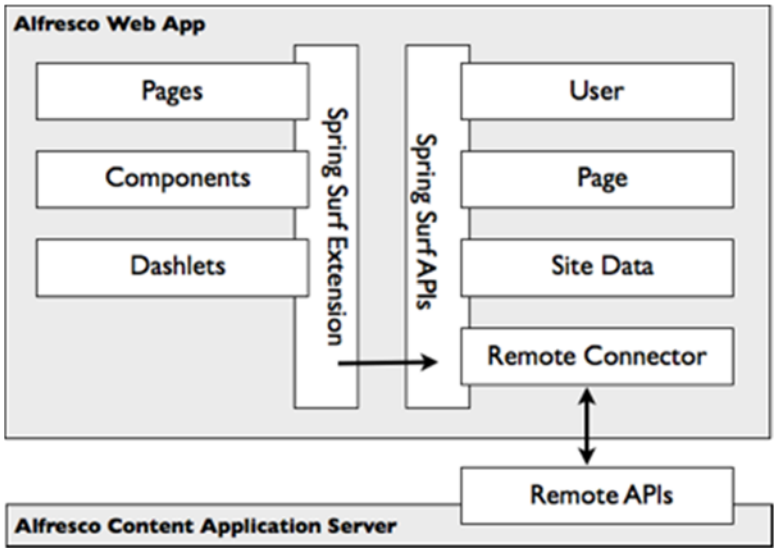


Figure 12: Alfresco web application framework

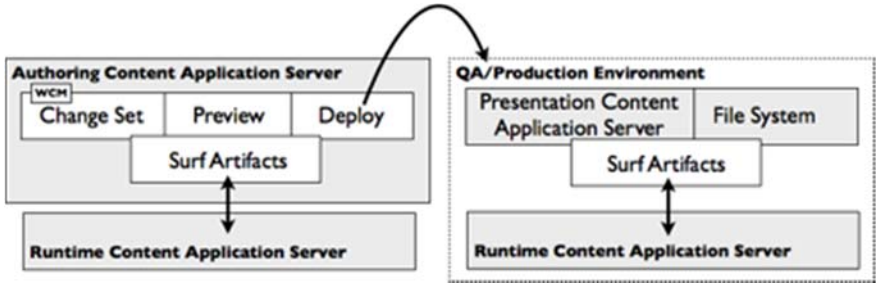
Alfresco Surf embeds Spring web scripts, allowing developers to use the same techniques that were used when building Alfresco content application server RESTful APIs. Often, an Alfresco Surf website requires access to and management of content held within the application content server, such as to support user-generated content, dynamic site artifacts, personalized presentation, and tagging. To support this, Alfresco Surf provides the following integration services:

- Remote: encapsulates any number of data sources with out-of-the-box support for the Alfresco content application server
- Credentials: manages user authentication with out-of-the-box support for the Alfresco content application server

Alfresco Surf works with Alfresco web content management and provides virtualized content retrieval, preview, and test support for user sandboxes and web projects. Applications built with Alfresco Surf can be deployed from Alfresco web project spaces to production servers. To facilitate this, Alfresco Surf uses a lightweight XML-driven model to represent all site artifacts, such as pages, templates, themes, and chrome. This means that Alfresco services, such as change sets, preview, and deployment, can manage Alfresco Surf sites. In addition, an embedded API supports programmatic control of the same artifacts (see the following figure).



As Alfresco Surf sites are XML and file-based, they are easily managed in Alfresco WCM, as shown in the following figure:



It offers features such as:

- Safe editing of all Alfresco Surf artifacts, including the ability to snapshot a site and roll it backward in time
- Review and Approve workflow of Alfresco Surf site changes
- Preview of site changes
- Deployment of site changes to test or production servers

With the CMIS client API, Alfresco Surf provides an open stack for implementing web-based, content-enabled applications. Alfresco 5.0 includes the new UI framework built on Alfresco Surf, **Aikau**. Aikau provides a modern, higher-level approach to developing custom UI applications, and features a simplified method for creating pages and widgets. New pages with standard widgets can be created through JSON code, and then extended as required using JavaScript.

Deployment options

You can deploy Alfresco in many different forms and topologies. Because its infrastructure foundation protects Alfresco from the environment within which it executes, you can choose components such as operating system, database, application server, web browser, and authentication system. Alfresco is designed to scale down as well as up.

Embedded Alfresco

An embedded Alfresco is contained directly within a host where the host communicates with Alfresco through its embedded API, meaning the host and Alfresco reside in the same process. Typical hosts include content-rich client applications that require content-oriented storage, retrieval, and services, but can also include hosts such as test harnesses and samples. A client can choose to embed the Alfresco web application framework or content application server, or both, treating Alfresco as a third-party library. In any case, the client can pick and mix the services of Alfresco to embed, allowing very small-footprint versions of Alfresco. The host is responsible for the startup and shutdown of Alfresco.

Alfresco content application server

An Alfresco content application server is a stand-alone server capable of servicing requests over remote protocols. A single server can support any number of different applications and clients where new applications can be arbitrarily added. Clients communicate with Alfresco through its Remote API and protocol bindings, although you can configure a server to omit or prohibit specific access points. This type of deployment takes advantage of an application server where Alfresco is bundled as a web application. Application server features, such as transaction management and resource pooling, are injected into the Alfresco infrastructure foundation, allowing Alfresco to take advantage of them.

For example, you can embed the Alfresco content application server inside Apache Tomcat for the lightest-weight deployment, as well as inside Java Enterprise Edition compliant application servers from JBoss, Oracle, or IBM to take advantage of advanced capabilities such as distributed transactions.

Multi-tenancy

Multi-tenancy allows a single Alfresco content application server (clustered or not) to support multiple tenants, where a tenant such as a customer, company, or organization believes they are the only user of the server as they connect to a logical partition. Physically, all tenants share the same infrastructure, such as deployed nodes in a cluster and content, repository storage. However, data maintained by one tenant cannot be read or manipulated by another tenant. A deployment of this type eases administration and reduces the cost associated with maintaining many different applications and user bases, in particular when upgrading core services or performing backups, as this only needs to be done once for all tenants.

Alfresco provides administration tools for managing tenants, including the creation of tenants at runtime. In conjunction with clustering, multi-tenancy provides an ideal deployment option for the Cloud.

4.6 Software Selection

After carefully comparing all the possible ECM software available, the project Team selected Alfresco. Alfresco community 5d Edition was chosen for this project mainly because the community edition is free and open source and has been used to implement a similar project in a very large organization.

5.0 CHAPTER FIVE: PROJECT IMPLEMENTATION

This chapter describes the steps taken towards the development and implementation of the ECM system.

5.1 System Requirements

The ECM solution was expected to meet the following functional requirements:

- Provide global access to information
- Provide better collaboration
- Support and improve work processes among work groups and teams supported by new IT tools
- Proper handling of emails
- Better information sharing between the Institute and its customers, employees and business partners
- Improve tracking of documents and in this way make it easier to retrieve information and ensure sharing and reuse of information
- Establish traceability of as well as easy, correct and secure access of information through the information life-cycle and with regards to legal requirements.
- Limit the duplication of data by classifying information with metadata, enabling the new search engine to find all relevant information regarding a project, process, discipline, organization etc.

5.2 Technical Requirements of Alfresco

5.2.1 Software Requirements

Alfresco can be installed on Linux, Windows or Mac Operating Systems.

The table below list the required software for Alfresco installation.

Table 2: Alfresco software requirements

Component	**Recommendation**
Java SE Development Kit (JDK)	The Sun Microsystems JDK 6 or higher is required. The JAVA_HOME environment variable must be set to the location of the JDK installation.
Application server	Alfresco runs within an application server. Alfresco Enterprise/Community Edition runs within Tomcat but can be installed on other application servers.
Database	Alfresco comes preconfigured with the PostgreSQL database. If you intend to use Alfresco in a production environment, you can use one of the supported databases. Eg MySQL
LibreOffice or OpenOffice.org	Alfresco uses OpenOffice or LibreOffice for transforming documents from one format to another, for example, a text file to a PDF file. If you do not install OpenOffice or LibreOffice, you will not have access to the transformation functionality. Use the latest (stable) version of OpenOffice.org or LibreOfficer.
ImageMagic	Alfresco uses ImageMagick to manipulate images for previewing.
Flash Player	Alfresco Share requires Flash Player Version 10.x to upload multiple files and view Flash previews. If you do not install Flash, you see the upload screen for single files. Use the latest (stable) version of Flash Player for your platform.
FFmpeg	Libraries and programs for handling multimedia data.
SWF Tools	Alfresco Share uses the pdf2swf utility for previewing PDF files. If you do not install SWF

	Tools, you will not see PDF previews, but image previews will still be available.

5.2.2 Hardware Requirements

Alfresco requires the following minimum hardware requirements for production use.

Table 3: Alfresco hardware requirements

Hardware	**Minimum Recommendation**
Processor	1x Processor
RAM	2GB
Hard Drive	100GB

5.2.3 Hardware for the Project

Although one could run Alfresco on almost any modern PC, we were guided by the following recommendations: server-grade components with no less than 4GB RAM, 1 dual-core CPU and 40GB HDD. The hardware details of our Alfresco servers are shown in Table 4.

For our main server we used an existing entry-level Dell PowerEdge T310, and we re-purposed a previously acquired Dell PowerEdge T110 meant for ECOAGRIS Project as a backup server.

Table 4: Hardware Specifications for the Project

Component	**Main Server**	**Backup Server**
System	Dell PowerEdge T310	Dell PowerEdge T110
Processor(s)	Intel® Xeon® CPU X3430@ 2.4GHz	Intel® Xeon® CPU E31270@ 3.4GHz
RAM	12GB	8GB
RAID Controller	PERC S300	PERC S300
Hard Drive (s)	2TB	1TB
Total Available Storage	1.5TB	750GB

Operating System	Ubuntu Server 14.04 (LTS)	Ubuntu Server 14.04 (LTS)

After putting together the hardware components, the next step was to setup the RAID controller for which we selected RAID 6 for its high drive-fault tolerance (two simultaneous drive failures) and read/write performance, giving us 1.5 terabytes (TB) and 720 gigabytes (GB) available space for the main server and backup server respectively. We surveyed our target users and estimated that this would provide approximately two years of storage space.

5.3 Installation of Alfresco and other software

The operating system selected for the project was Ubuntu Linux OS for production use, specifically Ubuntu server 14.04 LTS. Ubuntu is the most popular Linux OS, and the Long Term Support (LTS) version allows for stability and longevity. Linux source code is freely distributed. Tens and thousands of programmers have reviewed the source code to improve performance, eliminate bugs, and strengthen security. No other operating system has ever undergone this level of review. The key advantages of Linux are listed as follows:

- The best technical support available
- No vendor lock-in
- Runs on a wide range of hardware
- Exceptionally stable
- Supports many tools and applications you need
- Interoperates with many other types of computer systems
- Low total cost of ownership

Ubuntu was successfully installed and configured on the two servers (the main server for Alfresco and the backup server).

5.3.1 Installation of the dependencies

Alfresco requires certain dependency software to be installed. The following software were installed on the Alfresco server.

- Java SE Development Kit (JDK) 1.7u67
- Apache Tomcat 7.0.55
- PostgreSQL 9.3.5
- ImageMagick

- FFMPeg
- SWFTools
- LibreOffice

After installation and configuration of all the dependencies, test was done to determine the functionality of Apache Tomcat. Apache Tomcat was running smoothly. Figure 13 below shows Tomcat page.

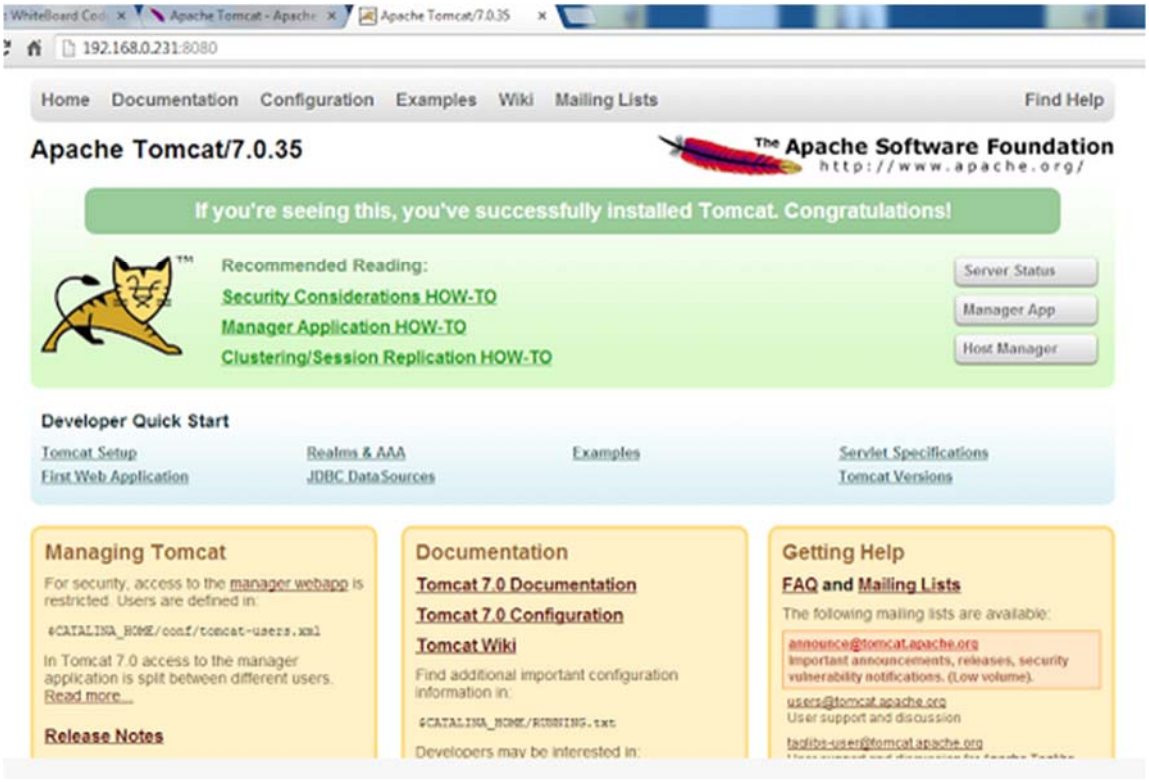


Figure 13: Apache Tomcat page

5.3.2 Installation of Alfresco

After all the required software were installed and configured, we were now ready to install Alfresco on the main server. As at the time of writing this report, the latest community version of Alfresco was Alfresco version 5.0.d. The community version was therefore downloaded from sourceforge website. The downloaded package was named 'alfresco-community-5.0.d.zip', and unzipped into a temporary folder with the name 'Alfresco-community-5.0.a'.

Alfresco was successfully installed on the production server with the help of an official online guide. Once Alfresco and its components were installed, the configuration of the various components was needed. We used a common Alfresco community configuration guide to obtain the basics working and tested the system functionality. Alfresco was therefore up and running and could initially be accessed at http://192.168.1.7:8080/share. Figure 14 below shows the initial login page of Alfresco.

Figure 14: Alfresco Community 5.0.d Login Page

Upon providing basic functionality, we moved to the network and security configuration. The first step was to purchase and install a Secure Sockets Layer/Transport Layer Security (SSL/TLS) certificate with Tomcat. Next, we started to patch up the server's network security by closing all the ports, except 22 for Secure Shell (SSH), 443 for Hypertext Transfer Protocol Secure (HTTPS), and 5432 for PostgreSQL. Using Internet Protocol (IP) tables, we restricted port 22 to only internal IP address and 5432 to the loopback (127.0.0.1).

Finally, a batch script was written to autostart Alfresco service upon server reboot and shutdown the service on a weekly basis, and copy the database/data directory from the main server to the backup server using rsync.

5.4 Advanced Configuration and Customization

5.4.1 Apache Tomcat configuration

Since much of the software components run in Java Virtual Machine (JVM), and because of the anticipated system demand, there was the need to do some tweaking to increase tomcat memory. Therefore, Java virtual memory was increased from the default size to 2 gigabytes (GB) to give enough memory to tomcat to handle high system demand.

Apache tomcat default port of 8080 was changed to port 8070. This was due to the fact that port 8080 was already in use by another program.

```
#
JAVA_OPTS="-XX:MaxPermSize=160m -XX:NewSize=256m -Xms512m "
JAVA_OPTS="$JAVA_OPTS -Xmx2048m -Xss512K "

# OS specific support.  $var _must_ be set to either true or
cygwin=false
darwin=false
os400=false
case "`uname`" in
CYGWIN*) cygwin=true;;
Darwin*) darwin=true;;
```

Figure 15: Java memory increment

5.4.2 Email server configuration

Alfresco requires Simple Mail Transfer Protocol (SMTP) server to send out email notifications. To this end, we setup a SMTP server. Google SMTP server was our preferred email server mainly because it's free, secure and reliable. The email server was therefore setup to send out notifications.

5.4.3 Customizing the login page

We also did a customization of the login page to reflect the Institutes branding. With the knowledge in CSS, HTML, JavaScript and Photoshop, the login page was successfully customized. Figure 16 shows the customize login page.

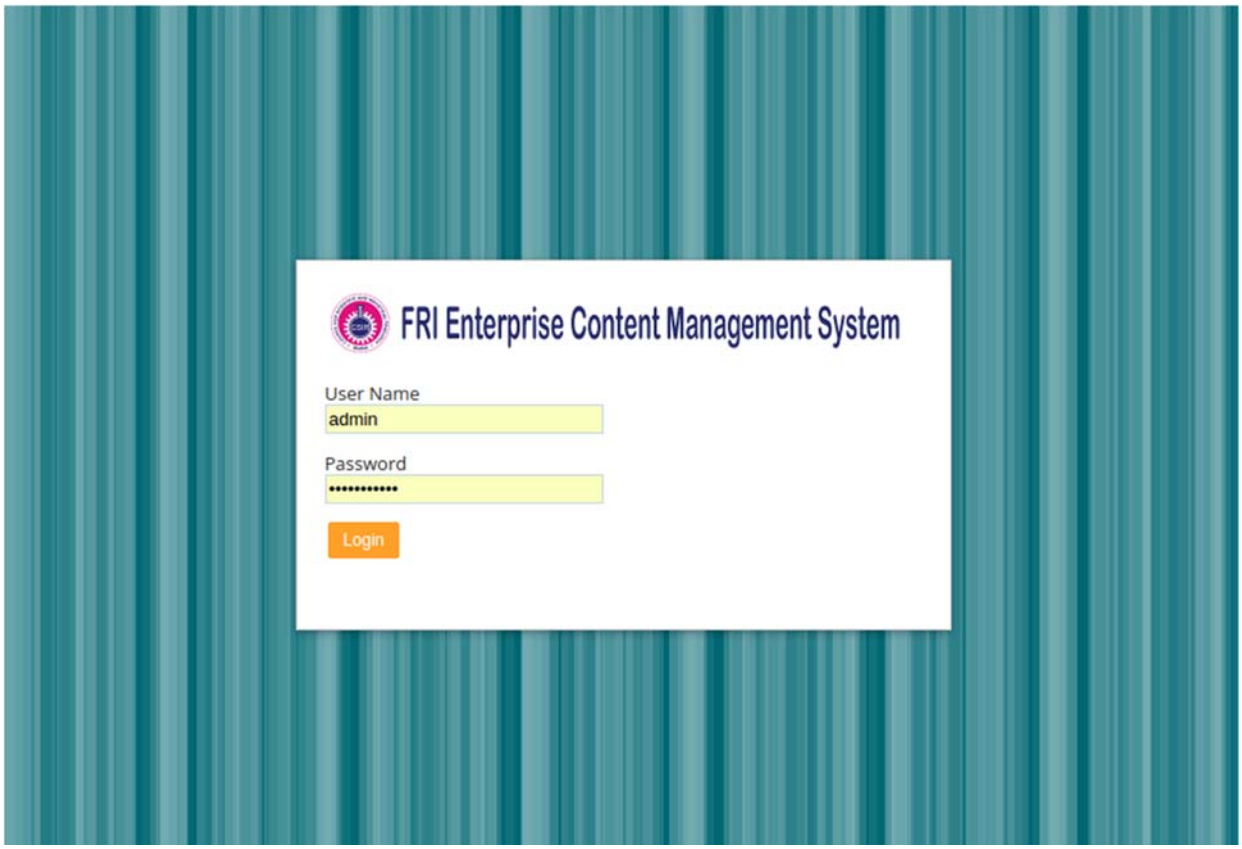


Figure 16: FRI enterprise content management system login page

5.4.4 Creating a custom theme

Organizations that uses Alfresco as an enterprise content management systems often want to match their own corporate branding within the Alfresco web app. The team decided to create a custom theme to reflect our corporate branding. We started by adding our logo. Adding a custom logo to the alfresco system was easy as one needed to navigate to the tools-⟶ options section of the Admin console. However, changing the colors and styles throughout the application was very tedious.

In order to get a better understanding of the level of effort involved in a custom theme creation, the team decided to create one for our institute based on the green theme that comes with Alfresco. We called our theme *fritheme* and used colors based on our corporate logo. We also changed the default favicon to use our customized version.

5.4.5 Customizing the Dashboard

Various customizations were done to the site dashboard including hiding the create site and help menu from non-admin users, permanently removing the welcome dashboard and customizing user invitation messages. Figure 17 below shows a screenshot of the site dashboard.

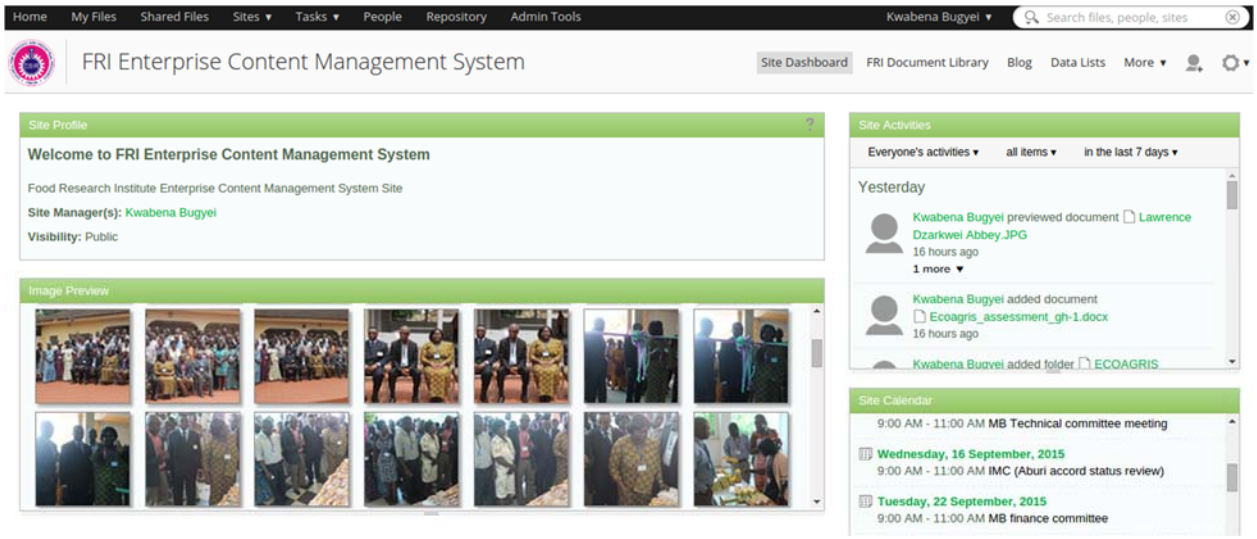


Figure 17: FRI Enterprise Content Management System Site Dashboard

5.5 Description of the Developed System

5.5.1 Creation of workspaces and user accounts

We have integrated the ECM system into the daily operations of the Institute. Each staff member has his or her own user account which gives access to one's own personal content, along with other aggregated content from other workspaces. The system uses Share –Alfresco's graphical user interface (GUI) as a web client and is viewed via standard web browser by accessing the server address at http://dms.foodresearchgh.org:8070/share.

Upon providing valid credentials (username and password), the user is directed to his/her personalized dashboard (Figure 18).

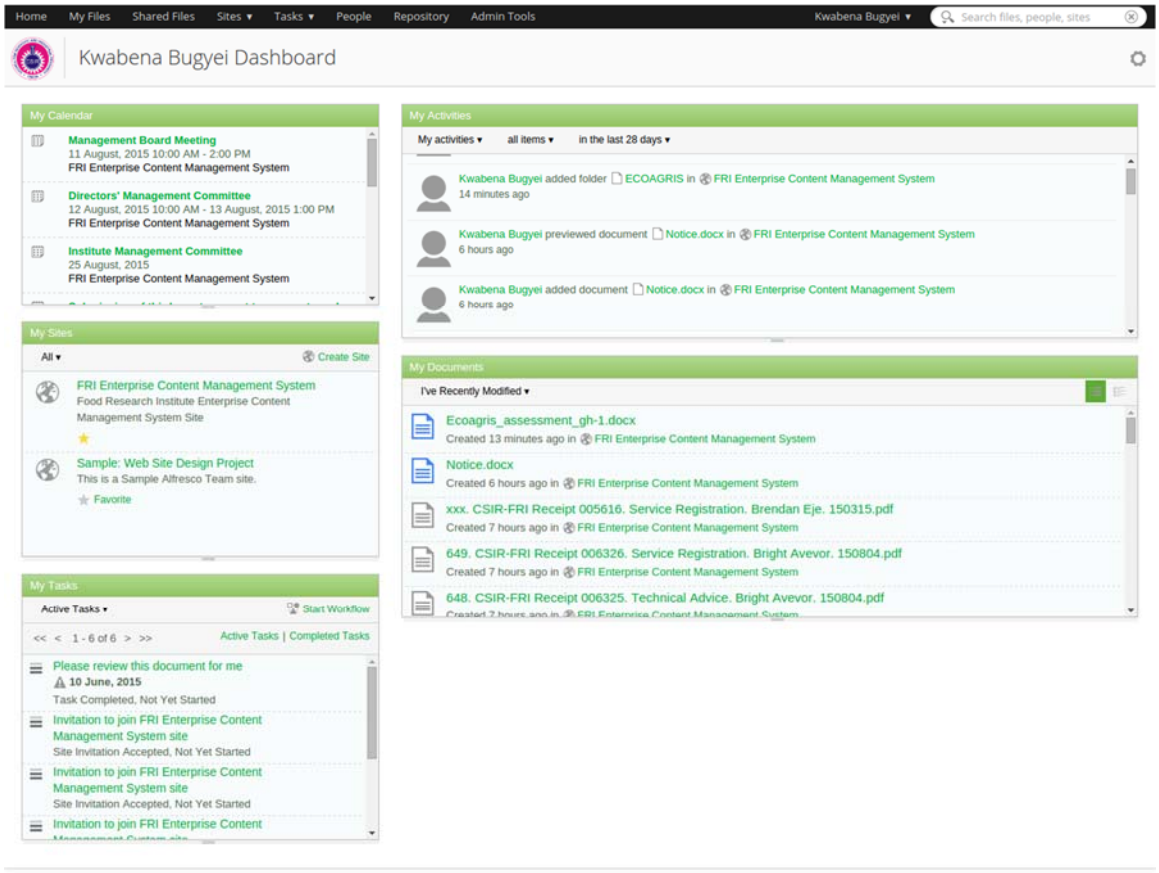


Figure 18: FRI-ECM personalized dashboard

The top-left portion of the dashboard contains the basic navigation menu: 'My Dashboard' links to the user's dashboard page; 'Sites' displays links to the user's favorite sites; 'People' links to the user directory; 'Repository' links to the user's personal files and directories; 'Tasks' displays links to tasks that has been assigned to user; 'Shared files' displays list of shared files and folders. The upper-left portion of the dashboard contains the user's name, links to user options (by hovering over the user's name), search textbox and help button. The small rectangular windows are 'widgets' which are small applications for the user based on individual requirements.

The team used the default authentication system that comes with Alfresco to create and assign roles to users of the system. Each user is a member of one or more Sites, which are essentially custom webpages that allow groups of users to manage content specific to a common goal or activity. Figure 19 below shows some users or members of the FRI Enterprise Content Management System Site.

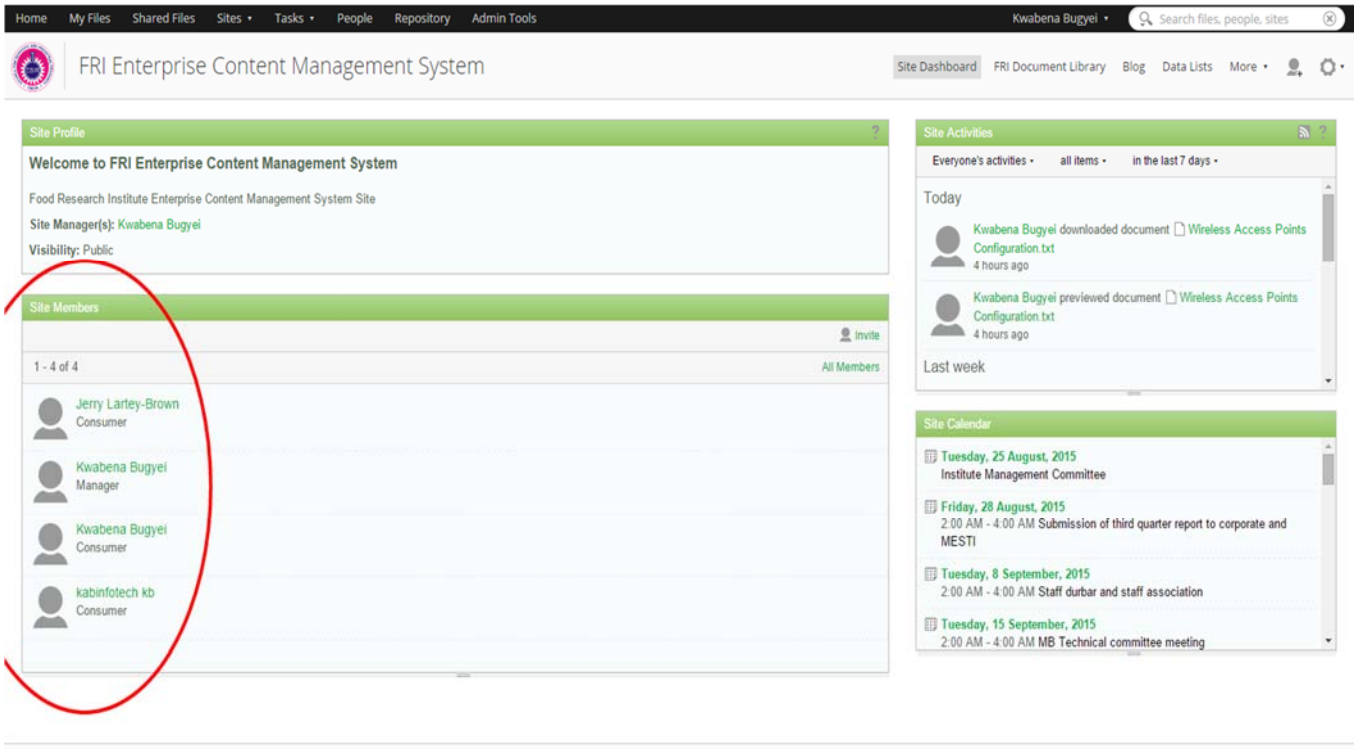


Figure 19: FRI-ECM site homepage showing site members

Each site contains a suite of content management tools available to its members, including: Dashboard, Wiki, Blog, Document Library, Calendar, Links, Discussions, Data Lists and Members. The Site's dashboard serves the same purpose as the user's dashboard but specifically aggregates the Site's content updates. The wiki, blog and discussion tools serve a similar function as collaborative mediums for the Site's members. The document library (which was renamed to FRI Document Library) is a data and information repository specific to the Site (FRI-ECMS). The calendar tool provides basic scheduling functionality. The links tool allows the members to share and store relevant web links (both internal and external), while the members tool provides a Site-level user directory. Finally, the data lists tool integrates lists (i.e., simple database) with workflow management (e.g., user tasks, document approval). Thus, each Site essentially serves as a custom ECM system for each group of users. Sites and their members are generally aligned to projects or organizational charts. For example, the Commercialization and Information Division can create a site comprising all the division's active members and contents that are relevant to the entire division.

Within the FRI Document Library, various folders were created to represent a workspace for each of the divisions of the Institute. These divisional folders will serve as smart workspace within which the divisional members will work and collaborate on documents relating to their division. It will provide a one-stop shop for any documents within the division. Figure 20 shows some workspaces created.

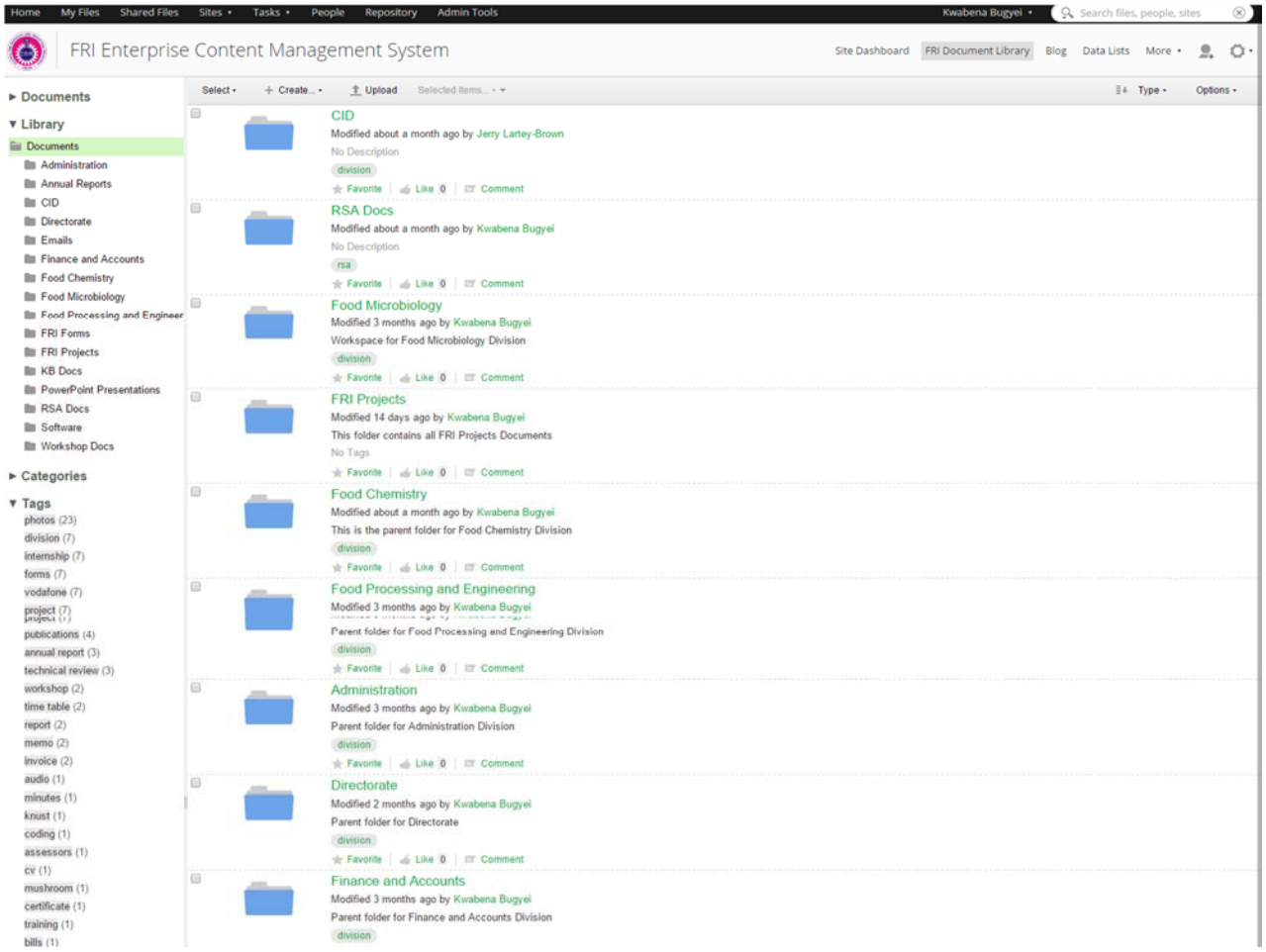


Figure 20: Workspaces/smart folders

5.5.2 Features of the System Developed

This section highlights a few of the key features of the system developed; however, this is by no means exhaustive.

5.5.2.1 Data and Information Management

The most critical and fundamental function of the system is the data and information management (unstructured and semi-structured documents), represented by the 'Repository' feature for users and the 'FRI Document Library' for Sites (hitherto referred to as 'repository' for both groups). Specifically, the repository allows users to upload, download, search, organize, update, comment, delete, move, or copy nearly any file or folder from nearly any computer or mobile device from anywhere in the world. The functionality is similar in nature to commercial solutions such as DropBox or iCloud, however, it offers unmatched benefits.

The repository can be accessed using several protocols: hypertext transfer protocol (HTTP), Web Distributed Authoring and Versioning (WebDAV), file transfer protocol (FTP), secure shell FTP (SFTP), content management interoperability service (CMIS) and more. The most popular file access options are the first three: HTTP (shown in Figures 18 & 19 above) can be accessed via

most web browsers and provides a pleasant GUI along with powerful management tools and search functionality. All data and information transferred via HTTP can be secured by adding a TLS/SSL certificate to the ECMS server which encrypts communication between the system and its users. This may be more familiar to users as HTTPS (the same security used by all retail/banking websites). WebDAV may be relatively lesser known; however, it is actually an extension of the HTTP protocol and enables the same basic functions of FTP along with collaborative functions such as 'locking' and 'version control'. As an extension of HTTP, WebDAV can utilize the same TLS/SSL security certificate and functionality as used by Alfresco's web service.

Finally, FTP and SFTP are supported. However, the former is universally deemed insecure, and the latter is more difficult to setup and offer only basic functions (i.e., file upload, download, copy, paste).

The ECMS is compatible with almost every system and computer configuration, thus allowing seamless management of all the Institute's data and information. While there are numerous direct benefits to this approach, such as backup and security, the ability to search through all this content directly is one of the most critical functions enabled by this system. After uploading any file to the system, an 'indexing' function actively searches and extracts all text from the file, including inside compressed files (e.g., zip or tar). Indexing allows users to search for internal content just as though they were using a commercial search engine such as Google or Yahoo. Furthermore, because the system developed integrates user management, the user can only search and view content for which they have security access, which is a very powerful feature to deal with sensitive content.

5.5.2.2 User Management

The system provides enterprise-class user management features that greatly enhance its value and usability. Administrators of the system can configure specific user security rules, or use the default settings. Every user starts with the same basic security access to his/her personal content. New users are generally invited directly to a specific Site (via the 'invite' feature on the Site dashboard), and are given a username, password and specific access level: Manager, Collaborator, Contributor, or Consumer. These Site access levels control all aspects of Site and associated content. For example, only Managers can invite new users, and Consumers cannot delete or modify content (i.e., read-only). Users can give other users access to a file, folder or their entire private repository. The user management features are greatly valuable as they simplify inherently complicated functions of most IT systems.

5.5.4.3 Knowledge Management

The system has 'knowledge management' tools, which are necessary for users to collaborate, document and share their knowledge. Tools such as the Wiki, Blog and Calendar functions are quite useful for these purposes. This feature may be less familiar but nonetheless quite powerful for managing an organization's knowledge.

5.5.4.4 Workflow Management

Workflow management (WFM) is already built directly into the system, which allows content to be 'processed' by specific timing and combinations of users and tasks. For example, the 'Review & Approve' workflow allows a user to assign content (e.g., a group of files/folders) to another user for review and approval. After initiating the workflow (directly from the respective content page), the recipient user is sent a notice detailing their specific tasks along with the associated content to be reviewed. The user downloads the content, reviews and/or changes the content, uploads the changed content, and submits the completed task. Every step in the workflow is documented, and its status can be easily monitored by both the workflow initiator and recipients.

5.5.4.5 Accessing the System via Mobile Devices (Android and iOS)

There is an open-source application released by Alfresco for accessing Site content via mobile devices (Android and iOS). The mobile application allows users to access their content on the ECMS server from literally anywhere in the world. Figure 21 and 22 are screenshots of the application (Samsung galaxy smart phone) showing documents in the system.

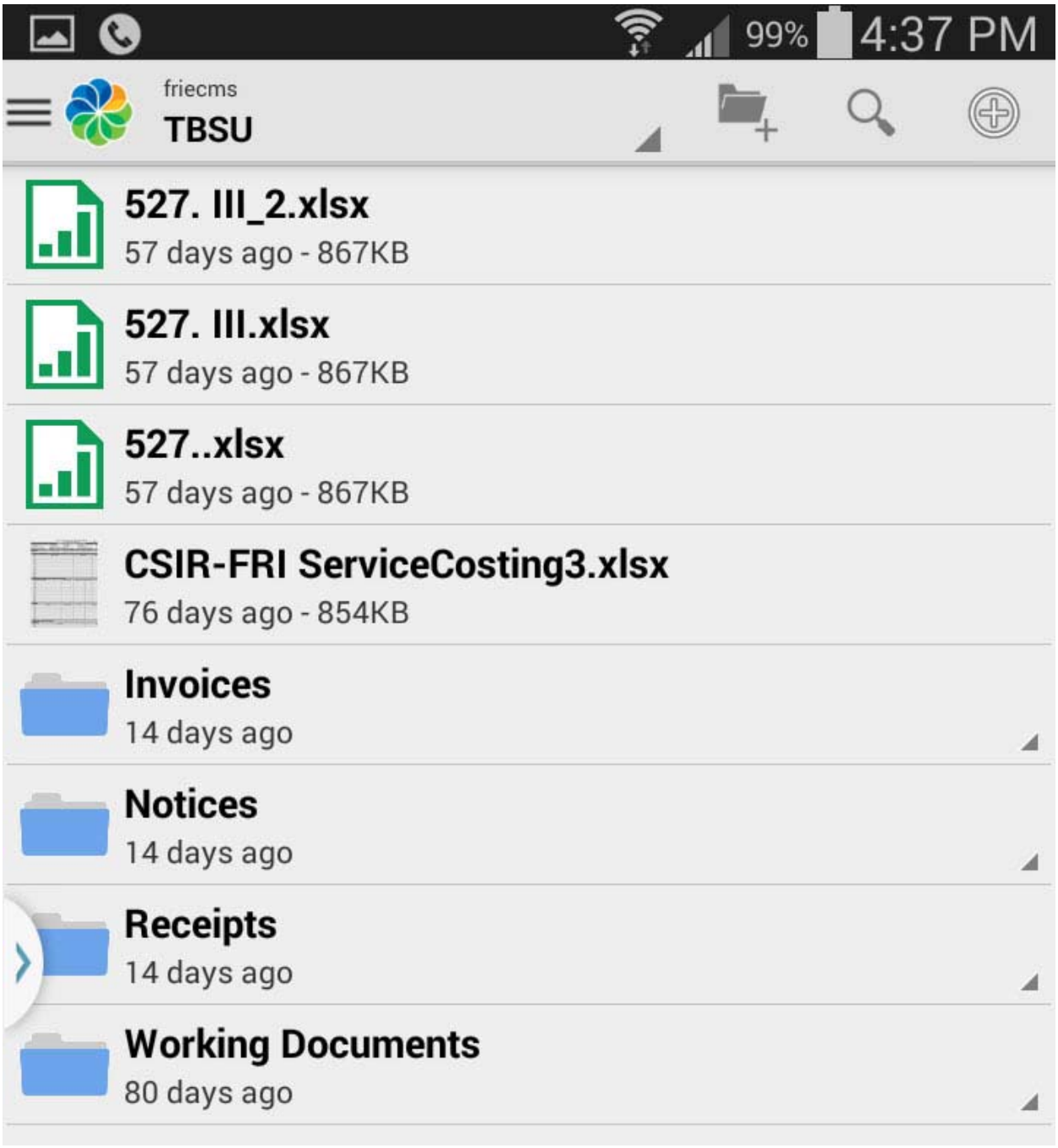


Figure 21: Accessing FRI-ECMS on Samsung galaxy smart phone

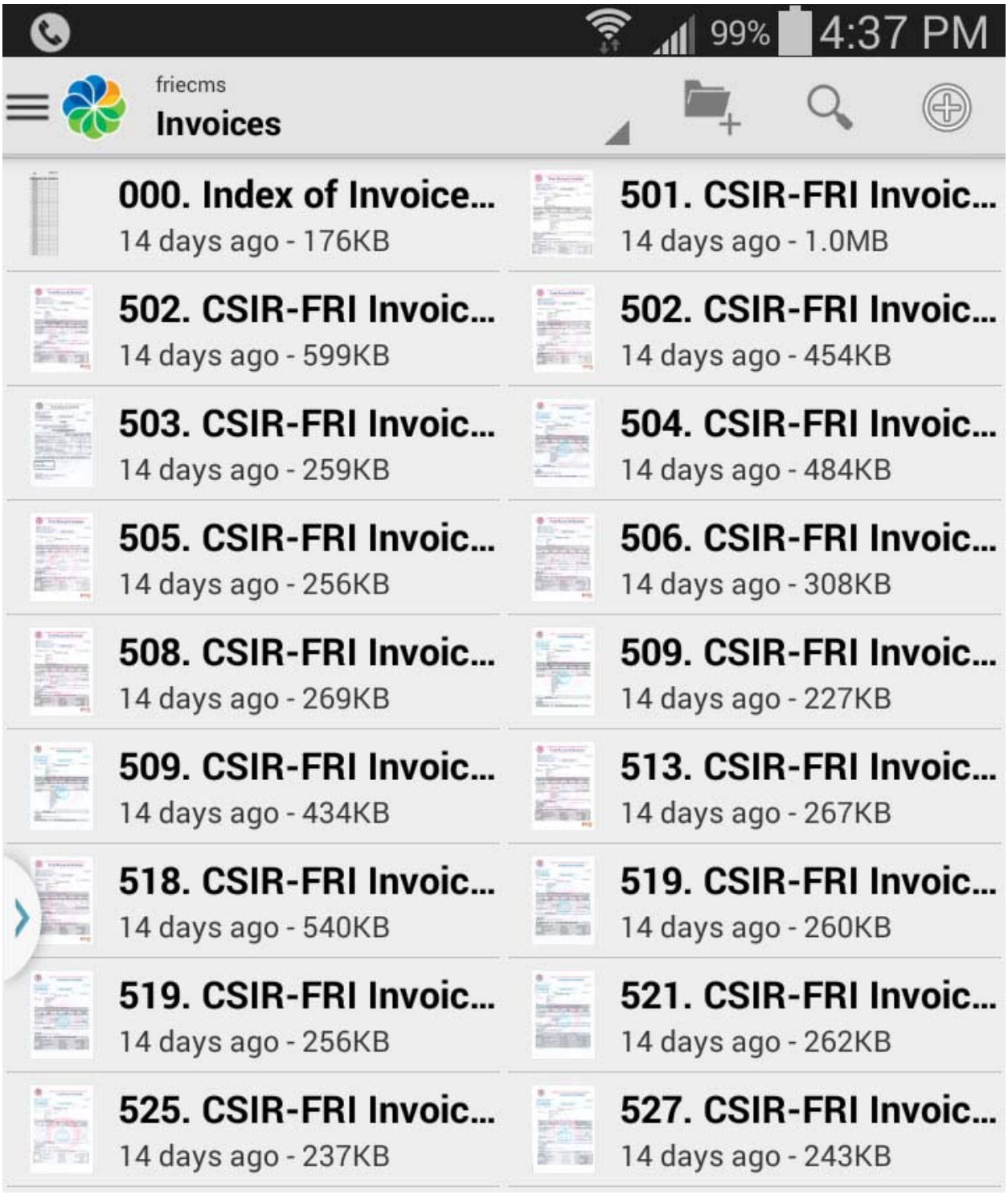


Figure 22: Screen capture of documents on FRI-ECMS server using Samsung galaxy mobile

5.6 Testing and Deployment

The developed system was thoroughly tested for performance and stability. After the test run, the system was ready to be deployed onto a production server. A new instance of a virtual machine was created on the Institutes newly developed private cloud Infrastructure as a Service (IaaS) powered by OpenStack, an open source cloud platform. The virtual machine created runs Ubuntu 14.04 server LTS OS, PostgreSQL database, Apache tomcat application server and other required software needed for the smooth running of the system. The system is therefore ready to be used by the Institute.

5.7 Screenshots of Developed System (FRI ECM System)

Please refer to the Appendix

6.0 CHAPTER SIX: CONCLUSION AND RECOMMENDATION

6.1 Conclusion

Our Institute's siren call for enterprise content management system occurred in 2011, when we lost a decent portion of our Institutes documents due to a backup hard drive failure. Our initial impetus was to design and implement a simple data management system (e.g., network file storage) for our documents and files. However, the search for an appropriate solution led us to the nascent world of open-source ECM. Alfresco was therefore chosen to implement enterprise content management for our Institute due to its maturity and enormous benefits. In the end, the project was successfully implemented.

The outcome of our efforts in deploying ECM system has been a resounding success, and we have been steadily building upon the system to enhance functionality and connectivity. It should come as no surprise that the ECM system implemented have dramatically improved our Institute's operations and collaboration as well as bringing a central control to all unstructured and semi-structured information across the Institute.

6.2 Recommendations

6.2.1 Upgrade ECMS server in terms of memory and storage space

The system developed consumes a lot of computer resources, therefore management should consider an upgrade of the underlying physical server's storage space and memory in order to accommodate large volumes of data.

6.2.2 Purchase of additional server purposely for backups.

For the purposes of ensuring a very good disaster recovery systems and business continuity, a dedicated server should be acquired to serve as a backup system for the ECMS server.

6.2.3 The system developed should be replicated in all CSIR Institutes

It is the authors' believe that replicating this system in all sister Institutes of CSIR will go a long way to ensure proper data and information management to help inform management decisions.

6.3 Future Work

6.3.1 Integrating the ECM system with a Portal

In embarking on this project, we discovered that the functionality required to manage our internal content differed significantly from our external content, and therefore the system developed should be integrated with a portal system to have a dual-layer integrated system: Portal system for external content and the ECM system for internal content.

6.3.2 More sites should be created to target specific groups

More sites should be created from the ECM to target specific groups within the Institute. That is to say customized site should be created specifically for groups of people to collaborate on a project for example.

7.0 REFERENCES

Aagre, A. (2008*). Seks steg til ECM*. Computerworld Norway, IDG Magazines Norge.

AIIM (2008) ECM Primer: *Understand the core ECM technologies and how they affect your organization's processes.* Available at http://www.inligo.us/PDF/ECM%20Primer.pdf, Accessed on 20 August 2014.

AIIM (2008) *What is ECM*. AIIM Web site: Accessible at: http://www.aiim.org/What-is-ECM-Enterprise-Content-Management.aspx, Access on February, 08. 2015.

Aleksy, M., & Schwind, M. (2006) *Using generic services for document life cycle management*. In: Proceedings of the 10th IEEE enterprise distributed object computing conference workshops (EDOCW), Hong Kong, pp. 60–65.

Aleksy, M., & Schwind, M. (2006) *Using generic services for document life cycle management*. In: Proceedings of the 10th IEEE enterprise distributed object computing conference workshops (EDOCW), Hong Kong, pp. 60–65.

Alfresco Forum.(2015). Available at http://forums.alfresco.com/en/, Accessed on 8th March 2015.

Alfresco Help (2015). Available: at http://www.alfresco.com/help/34/community/all/, Accessed on December 2014.

Alfresco JVM Tuning. (2015). Available: at http://wiki.alfresco.com/wiki/JVM_Tuning, Accessed on February, 20, 2015.

Allen D., 2007. Cost/benefit analysis for implementing ECM, BPM systems: determining the ROI for a significant investment, such as adopting an ECM. *Information Management Journal*

Alsup, M., & Strong, K. (2004). *Enterprise Content Management: Charting the Path.* AIIM E-Doc magazine, 18(3), pp.16-27.

Alsup, M., & Strong, K. (2004). *Enterprise Content Management: Charting the Path.* AIIM E-Doc magazine, 18(3), pp.16-27.

Apache Tomcat 6.0: SSL Configuration HOW-TO. Available at: http://tomcat.apache.org/tomcat-6.0-doc/ssl-howto.html, Accessed on April 20th 2013.

Asprey, L., and Middleton, M., 2003. *Integrative Document and Content Management*, Idea Group Publishing, Hershey, Pennsylvania

Banerjee, K. (2000). Challenges *of Using Metadata in a Library Setting: the Collection and Management of Electronic Links (CAMEL) Project at Oregon State University.* Library Collections, Acquisitions, & Technical Services, 24, pp. 217-227.

BitKinex. (2010, Dec. 15, 2012). FTP, SFTP and WebDAV

Blair, B. T. (2004). *An Enterprise Content Management Primer.* Information Management Journal, 38(5), pp. 64-66.

Carayon, P., & Smith, M. J. (2000). Work organization and ergonomics. *Applied Ergonomics*, 31(6), pp.649–662.

Chieu, T. C., Zeng, L., & Mohindra, A. (2008). *An Extensible Enterprise Content Management System with Service Component Architecture.* IBM T. J. Watson Research Center, IEEE, pp. 1131-1137.

Chiu, D. K. W., & Hung, P. C. K. (2005*).* Privacy and Access Control Issues in Financial Enterprise Content Management, *Proceedings of the 38th Hawaii International Conference on System Sciences*. Waikoloa, HI, USA, January 3-6.

Clark, D., 2008. Content Management and the Separation of Presentation and Content. *Technical Communication Quarterly*, 17 (1), 35-60. Availab[le at: http://www.bitkinex.com/](http://www.bitkinex.com/) Accessed on 25th June 2014.

D. Caruana, Professional Alfresco. Indianapolis, IN: Wiley, 2010.
D. Freshwater, G. Sherwood, and V. Drury, "International research collaboration," Journal of Research in Nursing, vol. 11, pp. 295-303, July 1, 2006 2006.

Davenport, T. H. (1998). Putting the Enterprise into the Enterprise System. *Harvard Business Review*, 76(4), 121–131.

De Carvalho, R. A. (2007). *An Enterprise Content Management Solution Based on Open Source*. Research and Practical Issues of Enterprise Information System II, Volume 1. pp. 173-183.

Dilnutt, R. (2006). *Surviving the information explosion*. IEE Engineering Management 16(1):39–41.

Dilnutt, R. (2006). *Surviving the information explosion*. IEE Engineering Management 16(1):39–41.

Duhon, B. (2009*). Enterprise Content Management: A Career and Job Title in Transition?* Infonomics, 23(4), 15.

EMC,Documentum.(2015),Availableat:http://www.emc.com/enterprise-content management/documentum-platform.htm, Accessed on 18th August, 2015.

Everything Linux. (1999, Sept. 25, 2013). A Tutorial on Using rsync. Available: http://everythinglinux.org/rsync/ Accessed on 24th September 2013.

Glick-Smith, J. (2004) *Planning: The Key to Successful CMS Implementation* The Rockley Report. Volume 1, Issue 1, pp. 3-6.

Gottlieb, S. (2005), *From enterprise content management to effective content management*. Cutter IT Journal Vol. 18 No. 5, pp. 13-18.

Hedman, J. and Kalling, T. (2002), *IT and Business Models: Concepts and Theories*, Författarna och Liber AB, Sweden.

Huff, M., & Dirking, B., 2010. The Benefits of a Unified Enterprise Content Management Platform. *Redwood Shores: Oracle white paper*.

Iverson, J., & Burkart, P. (2007*). Managing electronic documents and work flows: Enterprise content management at work in nonprofit organizations*. Nonprofit Management & Leadership, 17(4), 403-419.

Junco, N., Bailie, R. A., & Ledet, D. (2005). A Case Study of a Content Management System: Choosing and Implementing a CMS, *Proceedings of the IEEE International Professional Communication Conference*, pp. 343–345.

Kampffmeyer, U. (2004) *Trends in Record, Document and Enterprise Content Management*. White Paper, Project Consult Unternehmensberatung GmbH, Hamburg.

Kampffmeyer, U. (2006) *Enterprise Content Management ECM*. White Paper, Project Consult Unternehmensberatung GmbH, Hamburg.

Kaplan, S., 2002a “Cool Tool”, *CIO Magazine*, January 15

Kaplan, S., 2002b. “Management by any Other Name?” *CIO Magazine,* January 15

Kemp, J. (2007). A *Critical Analysis into the Use of Enterprise Content Management Systems in the IT Industry*. White Paper, AIIM.

Kemp, J., 2008. *A Critical Analysis into the Use of Enterprises Content Management Systems in the IT Industry.* Association for Information and Image Management

Lewis, M. (2009). *Enterprise Content Management 2010 and beyond.* Infonomics11/12 2009. AIIM International, USA.

Luan, J., & Serban, A., 2002. Technologies, Products and Models Supporting Knowledge Management, *New Directions for Institutional Research*, No. 113, pp.85-104

M. Shariff, Alfresco 3 enterprise content management implementation. Birmingham, U.K.: Packt Publishing, 2009.

McKeever, S. (2003) *Understanding Web content management systems: evolution, lifecycle and market.* Industrial Management & Data Systems 103(8/9): pp. 686–692.

McNay, H. E. (2002). *Enterprise Content Management: An Overview.* Professional Communication Conference, IPCC 2002. Proceedings. IEEE International.
Mescan, S. (2004). *Why content management should be part of every organization’s global strategy*. The Information Management Journal 38(4), pp. 54–57

Meyer, R. E. (2005) *Overcoming Information Overload - In the Cyber Soup of Today's Electronic Information, Insurers Need a Real-Time Way to Access Related Facts in E-Mails, Documents, Spreadsheets and More.* Information Management Technology - Best's Review 105, no. 12, pp. 101-103.

Munkvold, B. E., Päivärinta, T., Hodne, A. K., & Stangeland, E. (2006). Contemporary Issues of Enterprise Content management: The Case of Statoil. *Scandinavian Journal of Information Systems, 18*(2), 69-100.

Munkvold, B.E. (1998). *Implementation of information technology for supporting collaboration in distributed organizations Dr.ing. thesis 1998:40*, NTNU, Trondheim, chapter 5, pp. 96-113.

Munkvold, B.E. (1998). *Implementation of information technology for supporting collaboration in distributed organizations Dr.ing. thesis 1998:40*, NTNU, Trondheim, chapter 5, pp. 96-113.

Munkvold, B.E. (2009). *Introduction to qualitative research in IS.* Lecture foils, 05 February 2009, Department of Information Systems, University of Agder, Kristiansand.

Near, P. (2010). *Metadata – The Heart and Soul of ECM.* Peter Near: Available at: *http://www.peternear.com/index.php/2010/02/19/metadata-the-heart-and-soul-of-ecm/* Accessed on April, 18. 2012.

NISO. (2004). *Understanding Metadata*. NISO Press, National Information Standards Organization.

Nordheim, S. (2008). *Corporate User Representatives and the Dialectics of Enterprise Systems: A Quest for Social Actors with Political Skill*. Proceedings of the 41st Hawaii International Conference on System Sciences - 2008

Nordheim, S., & Päivärinta, T. (2006). *Implementing enterprise content management: From evolution through strategy to contradictions out-of-the-box*. European Journal of Information Systems, 15(6), 648-662.

O'Callaghan, R., and Smits, M.A. (2005) Strategy development process for enterprise content management, *Proceedings of the 13th European Conference on Information Systems*, Regensburg, Germany, 1271-128.

OpenKM. (2015), Available at : http://www.openkm.com, Accessed on 4th August 2015,

Paivarinta, T., & Munkvold, B. E. (2005). *Enterprise Content Management: An Integrated Perspective on Information Management.* Paper presented at the Hawaii International Conference on System Sciences, Hawaii.

Porter, M, E. (1998) *Competitive Advantage: Creating and Sustaining Superior Performance.* US: Free Press.

R. Natarajan. (2010, Sept. 25, 2012). 12 Steps to Install and Configure Alfresco on UNIX / Linux. Available at: http://www.thegeekstuff.com/2010/05/alfresco-installation-gu ide-for-linux/

Rockley, A. (2003) *Managing enterprise content: a unified content strategy.* The Rockley Group New Riders Publishing, Berkeley

Rockley, A. (2003) *Managing enterprise content: a unified content strategy.* The Rockley Group New Riders Publishing, Berkeley

Rogalski, S. (2006). The *rising importance of enterprise content management*. DM Review Magazine pp. 36

S. Bhaumik, Alfresco 3 Cookbook. Birmingham, U.K.: Packt Publishing, 2011.

Shang, S., and Seddon, P. B., 2002. Assessing and managing the benefits of enterprise systems: the business manager"s perspective. *Information Systems Journal,* vol. 12,4, pp 271–299.

SharePoint, (2015), Available at: http://www.microsoftvirtualacademy.com/product-training/sharepoint, Accessed on 18th August, 2015,

Shariff, M. Prajapati, J., Bhandari A., Shah, S., Pal, V., Majmudar, P., Avatani, R. R., Choudhary, V. (2013), Alfresco 4 Enterprise Content Management Implementation, 3rd Edition, Packt Publishing Ltd, Birmingham, UK.

Smith, H. A., & McKeen, J. D. (2003). *Developments In Practice: Enterprise Content Management.* Communications of AIS, 2003(11), pp. 647-659.

Smith, H. A., & Mckeen, J. D. (2003). Developments in Practice VIII:Enterprise Content Management. *Communications of the Association for Information Systems* , 11, 647-659.

Sprague, R. H., (1995). Electronic Document Management: Challenges and Opportunities for Information Systems Managers. *MIS Quarterly*, 19(1), 29–49.

Sprehe, J. T., (2005). The positive benefits of electronic records management in the context of enterprise content management. *Government Information Quarterly*, 22(2), pp.297–303.

Stalters, R. (2007) The Better ECM Definition of ECM. Available at: http://betterecm.wordpress.com/category/arma/ *Accessed on February, 08. 2014.*

Stoneburner, G., Alice. G., Alexis, F., 1999. *Risk management guide for information systems*, NIST SP 800-30, US Dept of Commerce.

T. S. Bell, Karen M.; Gilbert, Mark R.; Chin, Kenneth, "Magic Quadrant for Enterprise Content Management," Gartner 2010.

Terra, C.J and Gordon, C (2003), Realizing the promise of corporate portals, Routledge; 1 edition

The Apache Software Foundation. (2015). Available at http://apache.org/ Accessed on January 14, 2015

The Nuxeo Platform, Retrieved on 6th August 2015, Available at: http://www.nuxeo.com Accessed on 2nd February 2014.

Tyrvainen, P., Paivarinta, T., Salminen, A., & Iivari, J. (2006). *Characterizing the evolving research on enterprise content management*. European Journal of Information Systems, *15*(6), pp. 627-634.

Tyrvainen, P., Paivarinta, T., Salminen, A., & Iivari, J. (2006). *Characterizing the evolving research on enterprise content management*. European Journal of Information Systems, *15*(6), pp. 627-634.
Tyrvainen, P., Paivarinta, T., Salminen, A., & Iivari, J. (2006). *Characterizing the evolving research on enterprise content management*. European Journal of Information Systems, *15*(6), pp. 627-634.

Ubuntu. (2015, Sept. 25, 2015). Ubuntu Server Download.

Usman, M., Muzaffar, A. W.,& Rauf, A. (2009). *Enterprise Content Management (ECM): Needs, challenges and recommendations.* 2nd IEEE International Conference on Computer Science and Information Technology (ICCSIT), pp. 283-289.

vom Brocke, J., & Simons, A. (2008). *Towards a Process Model for Digital Content Analysis: The Case of Hilti*. Paper presented at the 21st Bled eConference 2008. eMergence: Merging and Emerging Technologies Processes, and Institutions, Bled, Slovenia, pp. 578-587

vom Brocke, J., Simons, A., & Cleven, A. (2010). *Towards a Business Process-Oriented Approach to Enterprise Content Management: The ECM-Blueprinting Framework.* Information Systems and e-Business Management (ISeB), accepted for publication.

Webopedia,(2015),Availableat:http://www.webopedia.com/TERM/C/commercial_software.html, Accessed on 18th August 2015.

Wikipedia,(2015), Available at: https://en.wikipedia.org/wiki/Proprietary_software, Accessed on 18th August 2015.

Wikipedia,(2015), Available at: https://en.wikipedia.org/wiki/Commercial_software, Accessed on 18th August 2015,

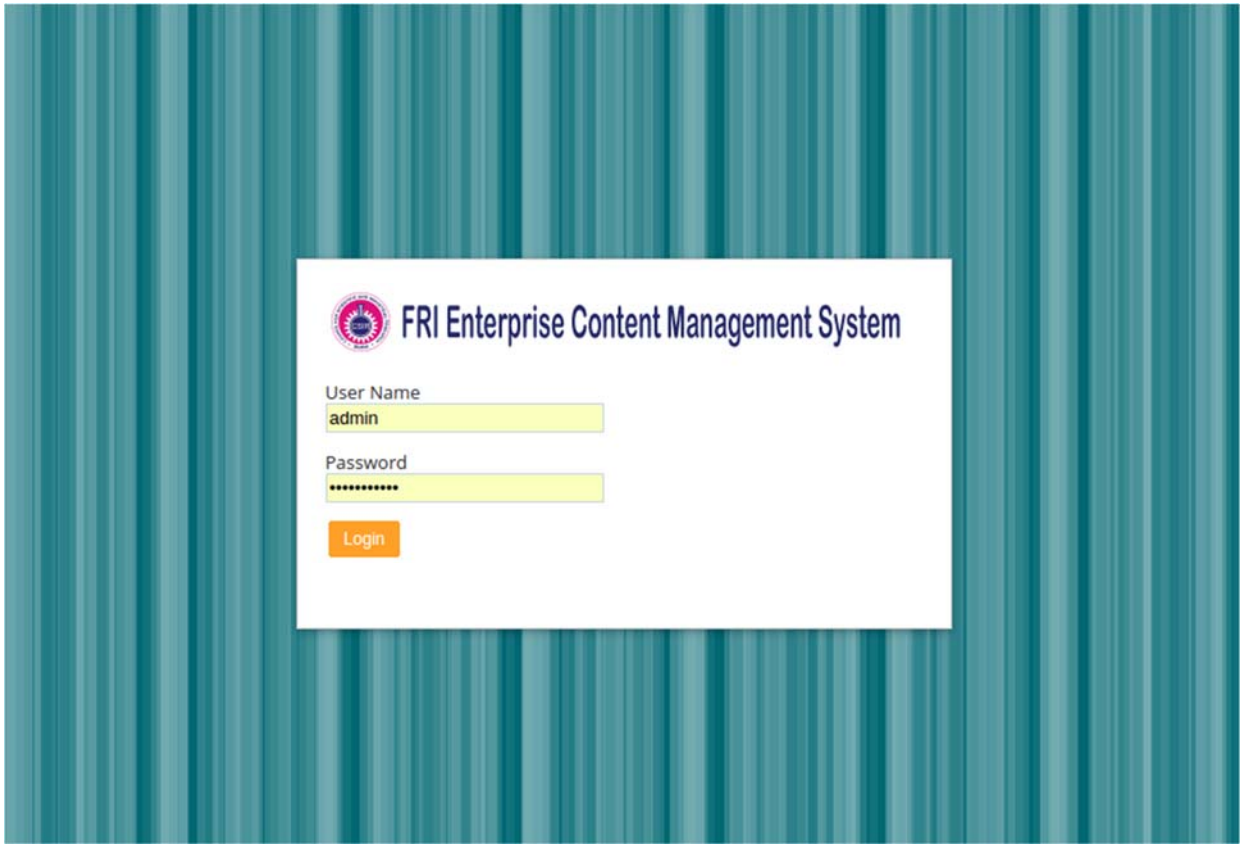
Wikipedia,(2015), Available at: https://en.wikipedia.org/wiki/SharePoint, Accessed on 18th August 2015,

Wilkoff, N., Walker, J., Root, N. & Dalton, .J (2001). *What's next for content management?* Cambridge: Forrester Research Inc.

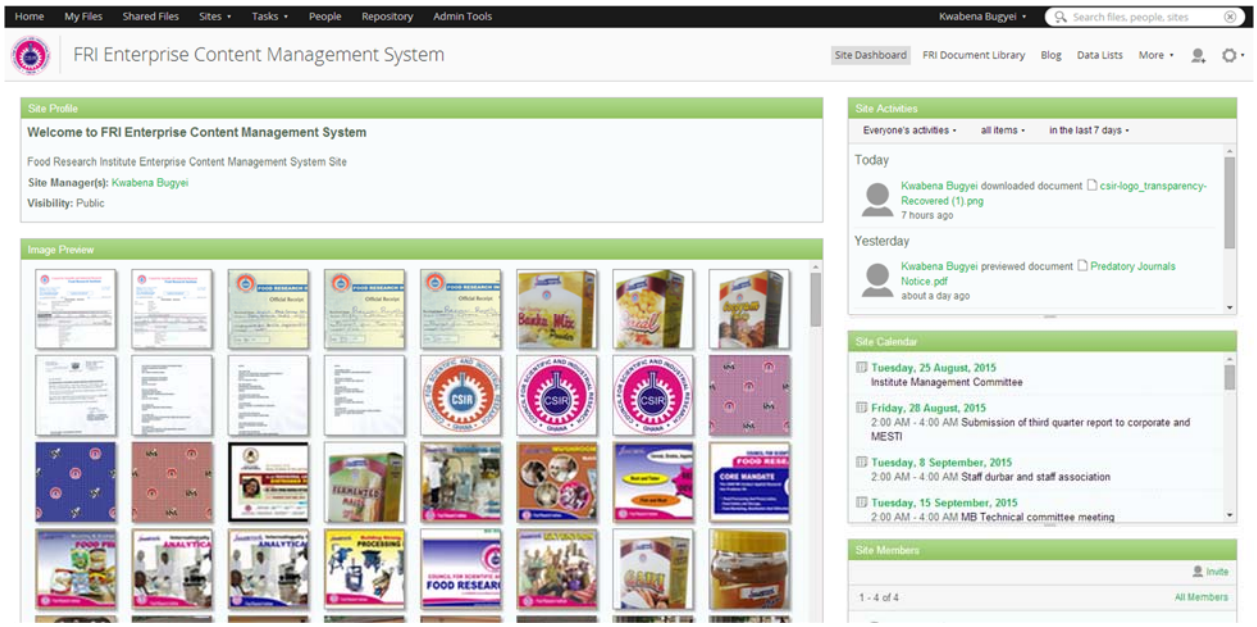
Willenborg, K. (2003). *Structured versus Unstructured Information*. SAP Design Guild Available: http://www.sapdesignguild.org/editions/edition2/sui_content.asp Accessed on March, 7th 2013.

8.0 APPENDIX

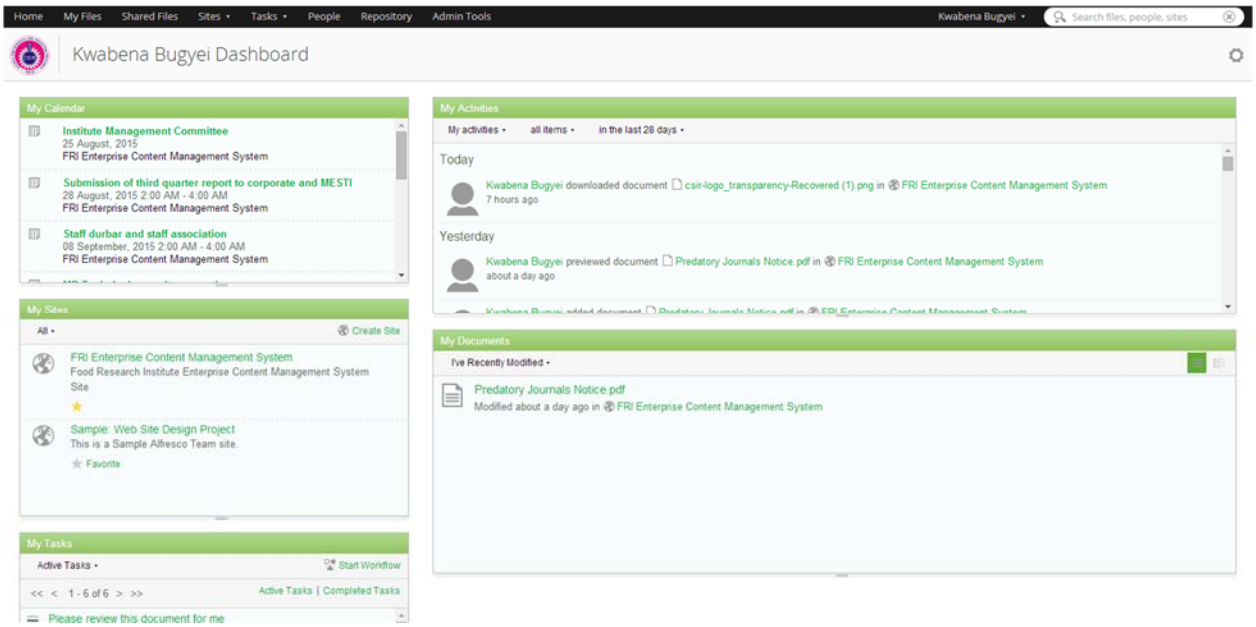
8.1 Login page



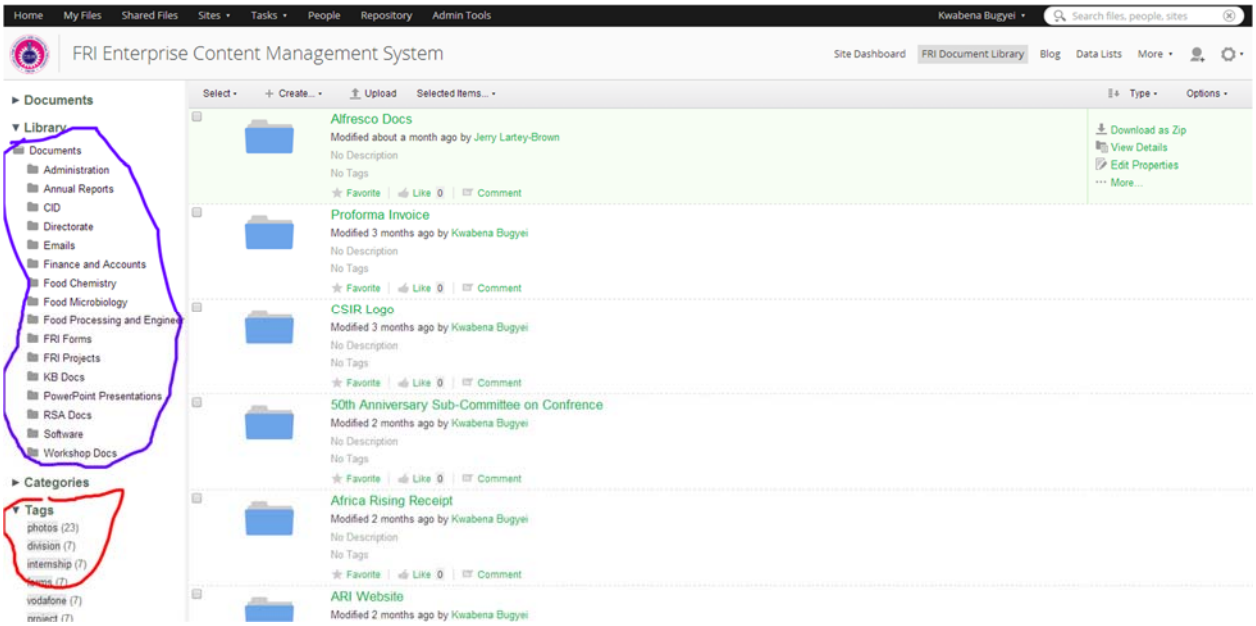
8.2 ECMS site dashboard

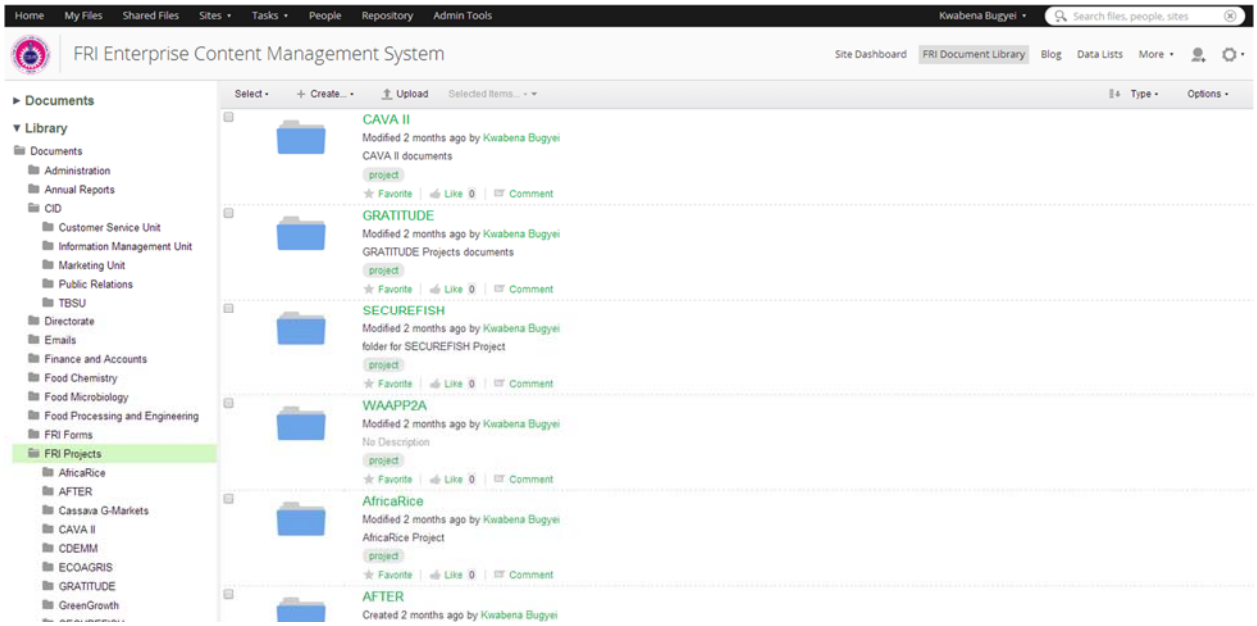


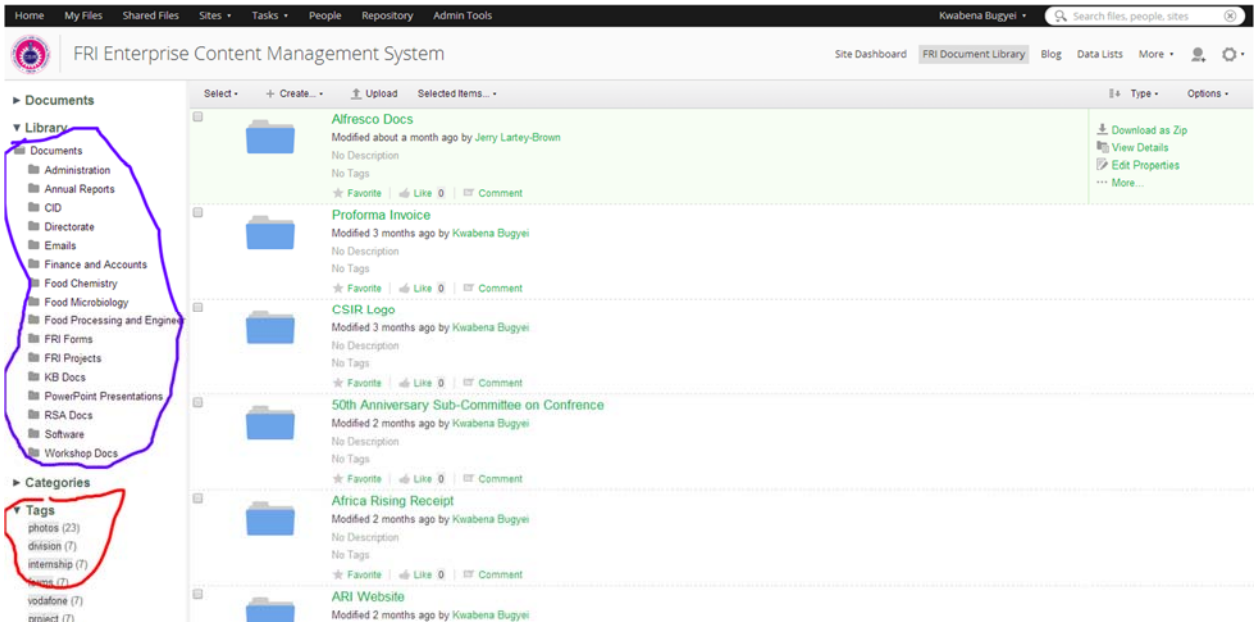
8.3 Personalized (user) dashboard

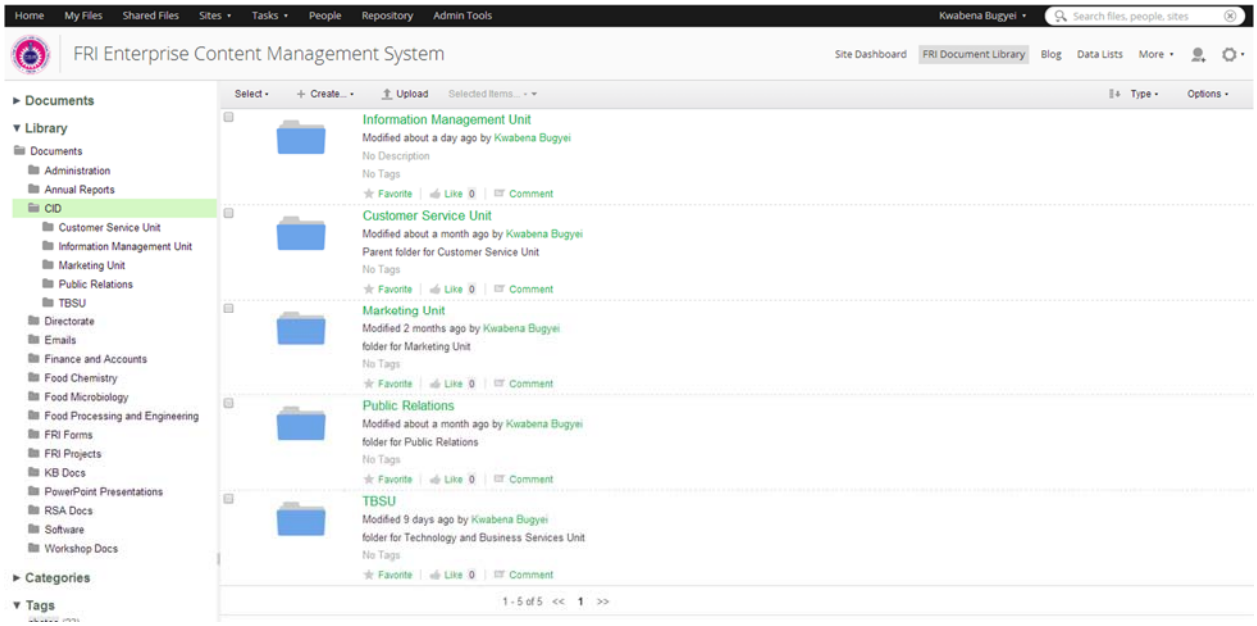


8.4 Document library view

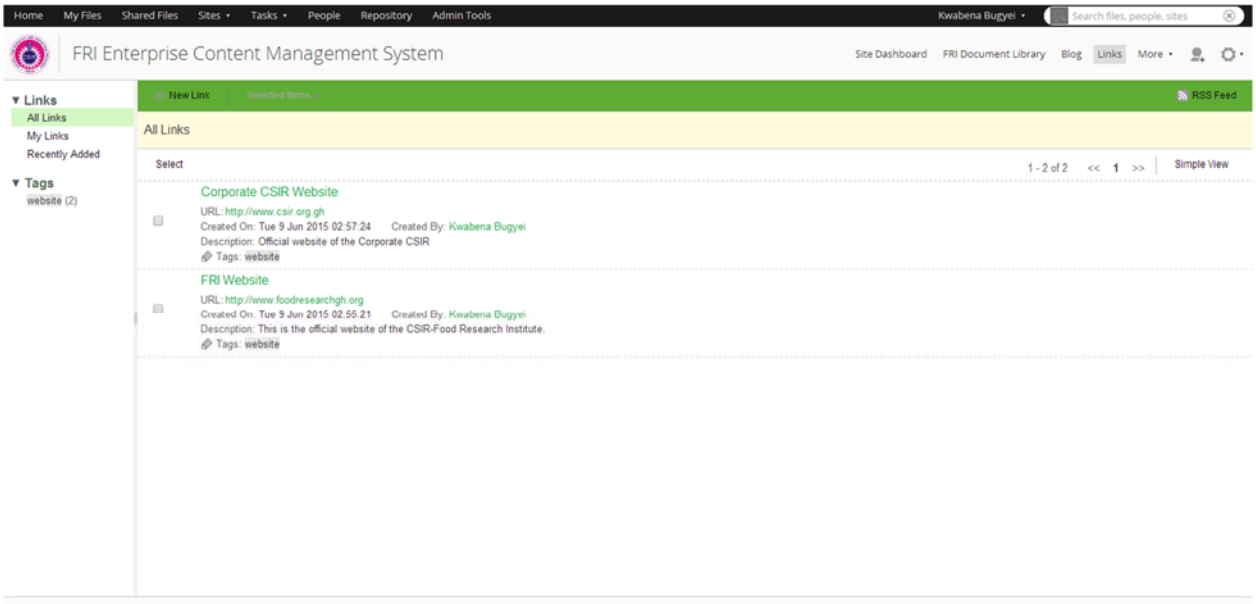



Home
My Files
Shared Files
Sites
Tasks
People
Repository
Admin Tools
Kwabena Bugyei
Search files, people, sites
FRI Enterprise Content Management System
Site Dashboard
FRI Document Library
Blog
Data Lists
More
Documents
Library
Documents
Administration
Annual Reports
CID
Customer Service Unit
Information Management Unit
Marketing Unit
Public Relations
TBSU
Directorate
Emails
Finance and Accounts
Food Chemistry
Food Microbiology
Food Processing and Engineering
FRI Forms
FRI Projects
AfricaRice
AFTER
Cassava G-Markets
CAVA II
CDEMM
ECOAGRIS
GRATITUDE
GreenGrowth
Select
Create...
Upload
Selected Items...
Type
Options
CAVA II
Modified 2 months ago by Kwabena Bugyei
CAVA II documents
project
Favorite
Like 0
Comment
GRATITUDE
Modified 2 months ago by Kwabena Bugyei
GRATITUDE Projects documents
project
Favorite
Like 0
Comment
SECUREFISH
Modified 2 months ago by Kwabena Bugyei
folder for SECUREFISH Project
project
Favorite
Like 0
Comment
WAAPP2A
Modified 2 months ago by Kwabena Bugyei
No Description
project
Favorite
Like 0
Comment
AfricaRice
Modified 2 months ago by Kwabena Bugyei
AfricaRice Project
project
Favorite
Like 0
Comment
AFTER
Created 2 months ago by Kwabena Bugyei

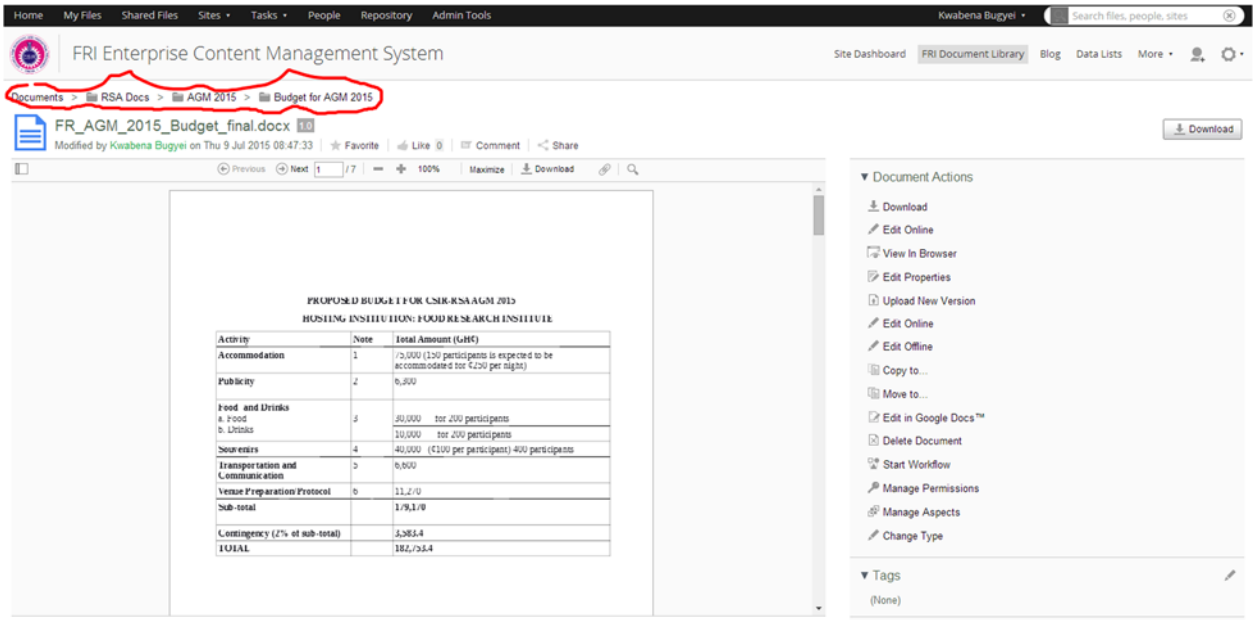

Home
My Files
Shared Files
Sites
Tasks
People
Repository
Admin Tools
Kwabena Bugyei
Search files, people, sites
FRI Enterprise Content Management System
Site Dashboard
FRI Document Library
Blog
Data Lists
More
Documents
Library
Documents
Administration
Annual Reports
CID
Directorate
Emails
Finance and Accounts
Food Chemistry
Food Microbiology
Food Processing and Engineering
FRI Forms
FRI Projects
KB Docs
PowerPoint Presentations
RSA Docs
Software
Workshop Docs
Categories
Tags
photos (23)
division (7)
internship (7)
vodafone (7)
Select
Create...
Upload
Selected Items...
Type
Options
Alfresco Docs
Modified about a month ago by Jerry Lartey-Brown
No Description
No Tags
Favorite
Like 0
Comment
Download as Zip
View Details
Edit Properties
More...
Proforma Invoice
Modified 3 months ago by Kwabena Bugyei
No Description
No Tags
Favorite
Like 0
Comment
CSIR Logo
Modified 3 months ago by Kwabena Bugyei
No Description
No Tags
Favorite
Like 0
Comment
50th Anniversary Sub-Committee on Confrence
Modified 2 months ago by Kwabena Bugyei
No Description
No Tags
Favorite
Like 0
Comment
Africa Rising Receipt
Modified 2 months ago by Kwabena Bugyei
No Description
No Tags
Favorite
Like 0
Comment
ARI Website
Modified 2 months ago by Kwabena Bugyei


Home
My Files
Shared Files
Sites
Tasks
People
Repository
Admin Tools
Kwabena Bugyei
Search files, people, sites
FRI Enterprise Content Management System
Site Dashboard
FRI Document Library
Blog
Data Lists
More
Documents
Library
Documents
Administration
Annual Reports
CID
Customer Service Unit
Information Management Unit
Marketing Unit
Public Relations
TBSU
Directorate
Emails
Finance and Accounts
Food Chemistry
Food Microbiology
Food Processing and Engineering
FRI Forms
FRI Projects
KB Docs
PowerPoint Presentations
RSA Docs
Software
Workshop Docs
Categories
Tags
Select
Create...
Upload
Selected Items...
Type
Options
Information Management Unit
Modified about a day ago by Kwabena Bugyei
No Description
No Tags
Favorite
Like 0
Comment
Customer Service Unit
Modified about a month ago by Kwabena Bugyei
Parent folder for Customer Service Unit
No Tags
Favorite
Like 0
Comment
Marketing Unit
Modified 2 months ago by Kwabena Bugyei
folder for Marketing Unit
No Tags
Favorite
Like 0
Comment
Public Relations
Modified about a month ago by Kwabena Bugyei
folder for Public Relations
No Tags
Favorite
Like 0
Comment
TBSU
Modified 9 days ago by Kwabena Bugyei
folder for Technology and Business Services Unit
No Tags
Favorite
Like 0
Comment
1 - 5 of 5 << 1 >>

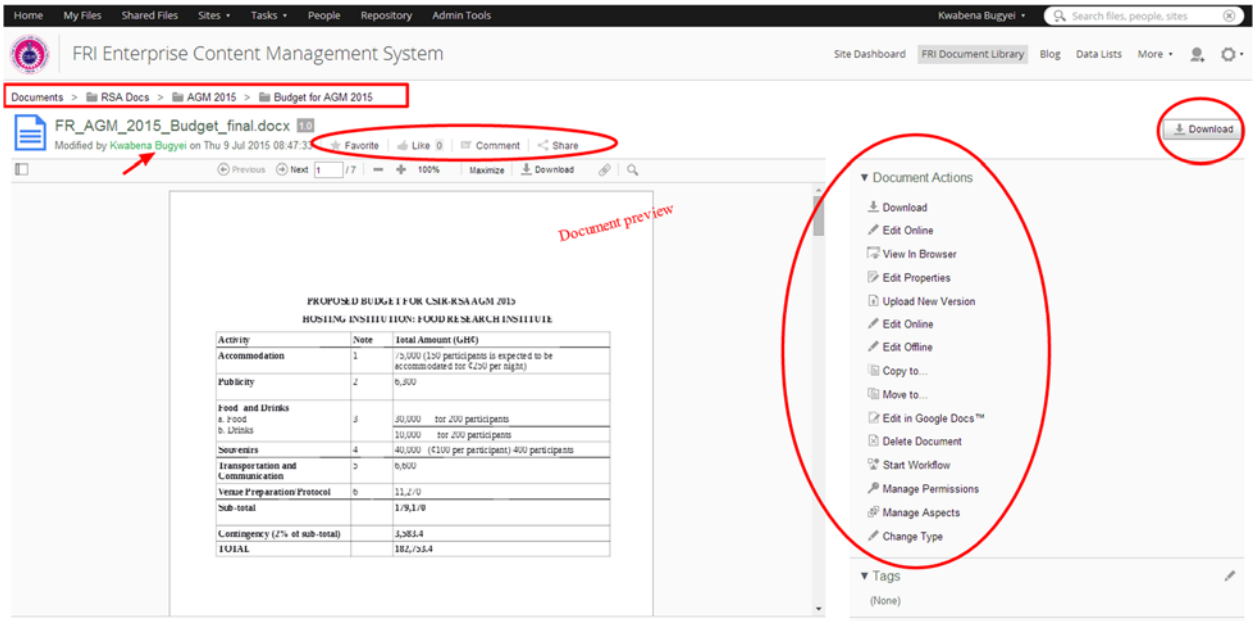
8.5 System links

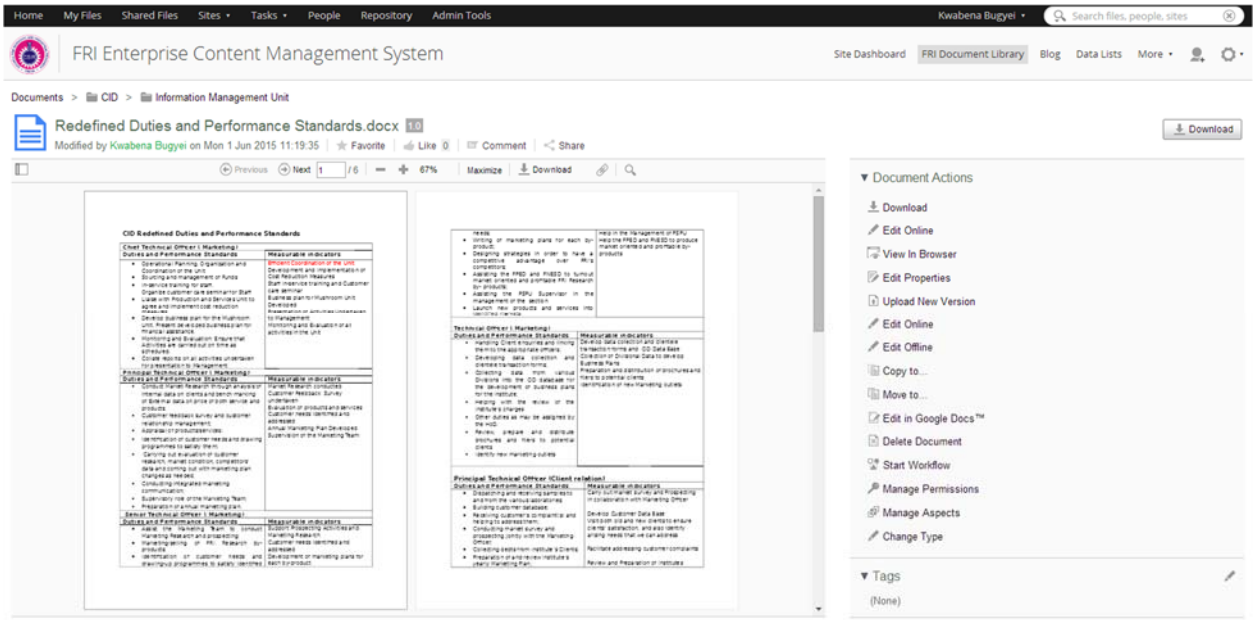


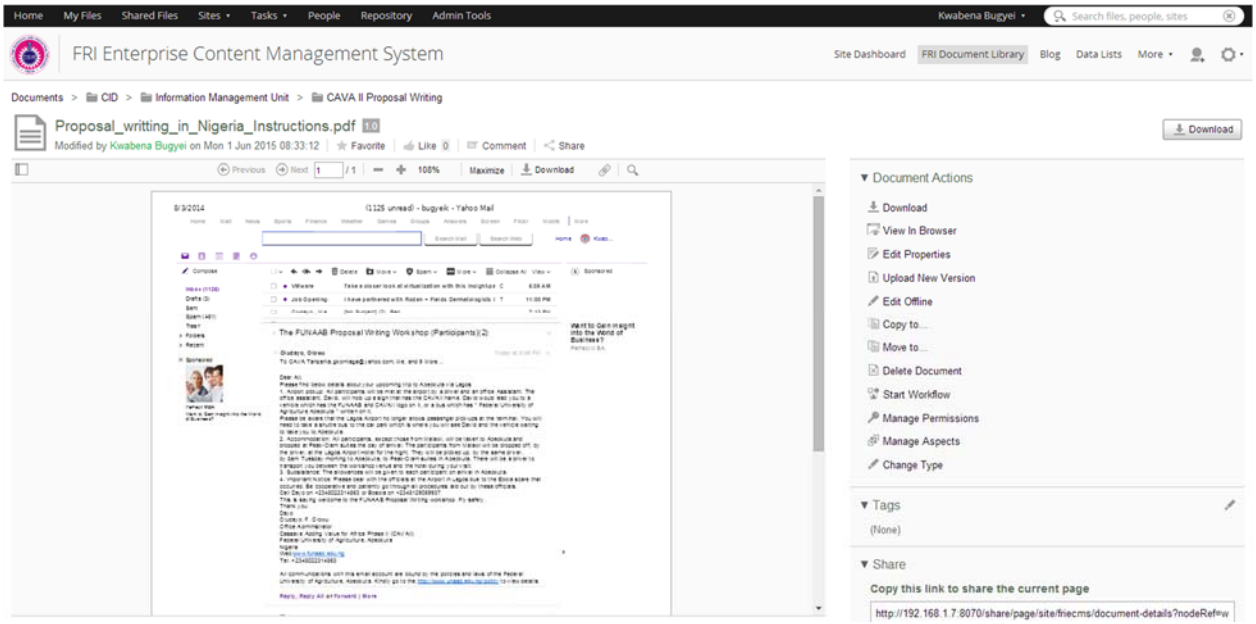
8.6 Document navigation



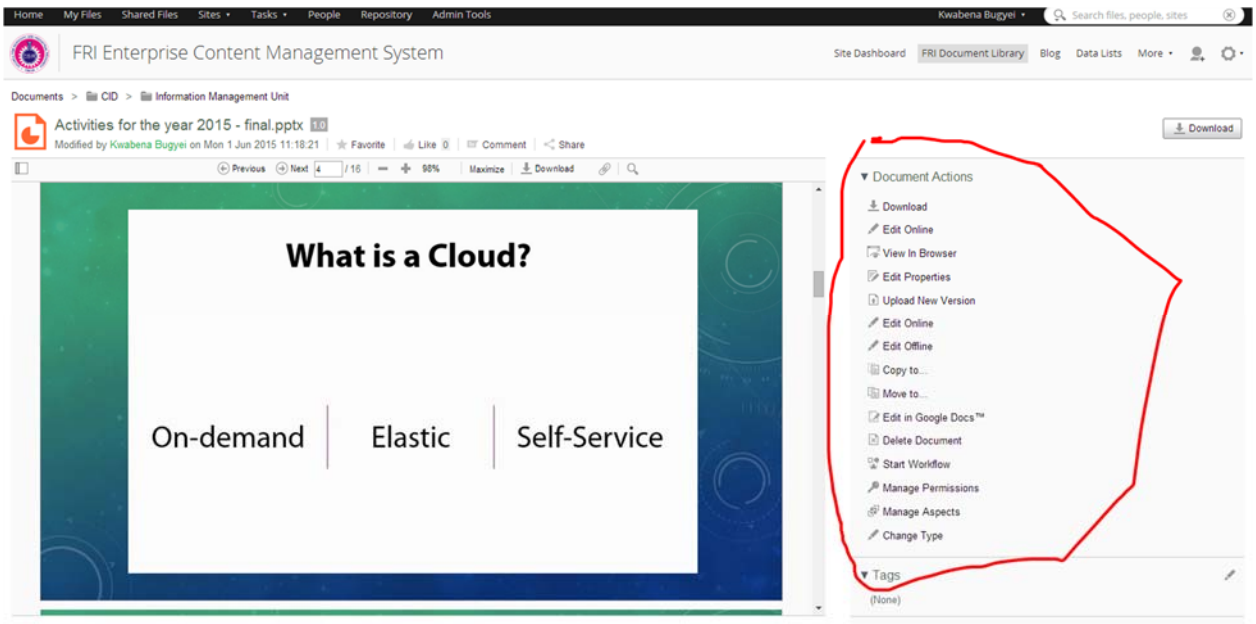
8.7 Document preview

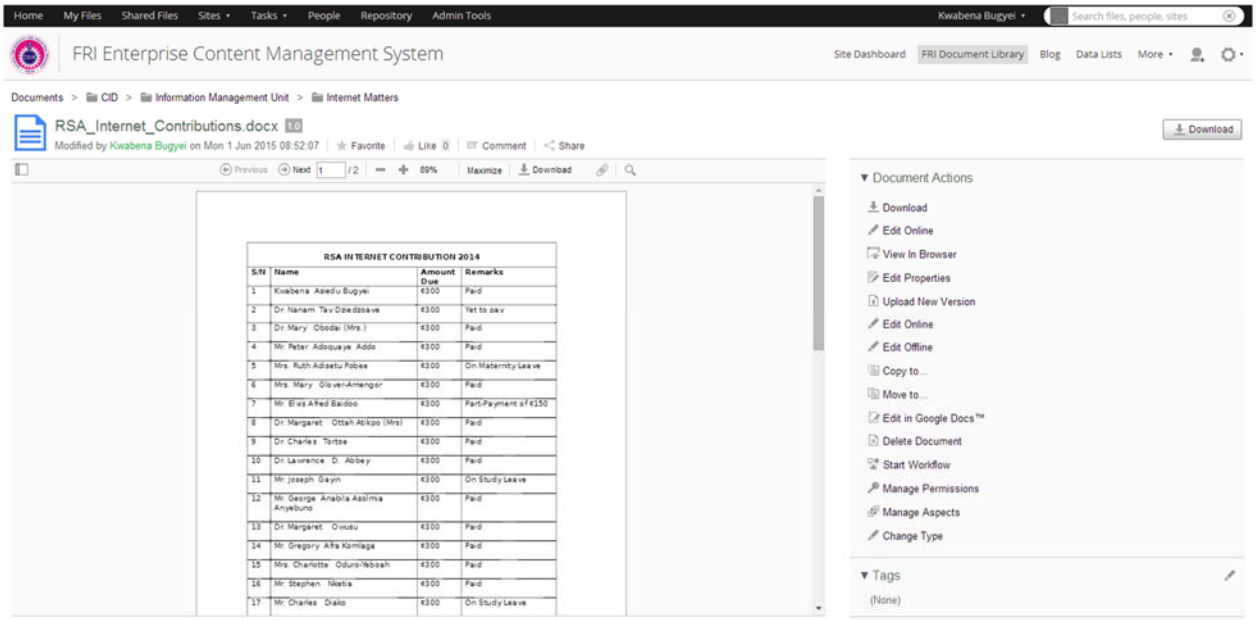




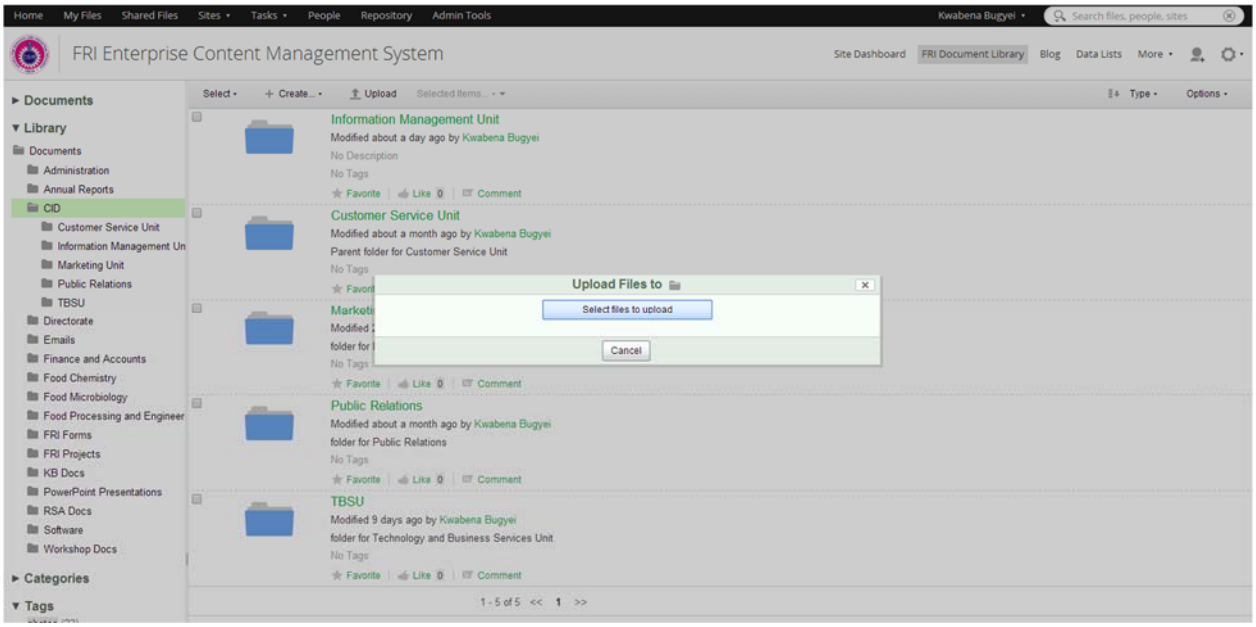


8.8 Document Actions

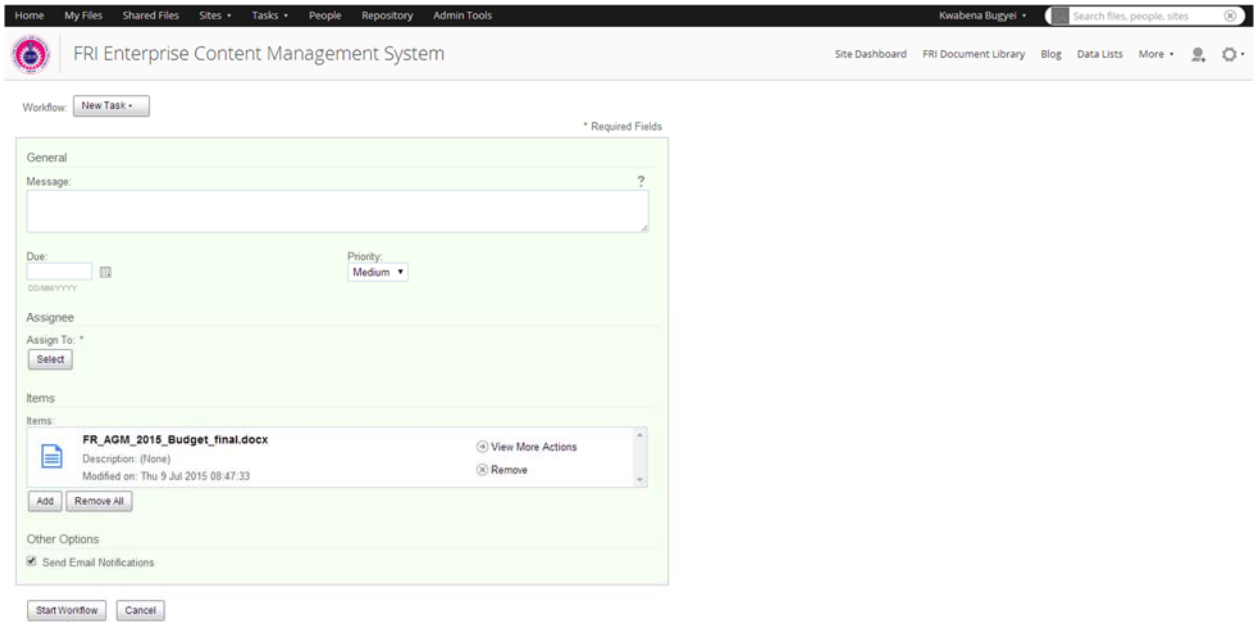




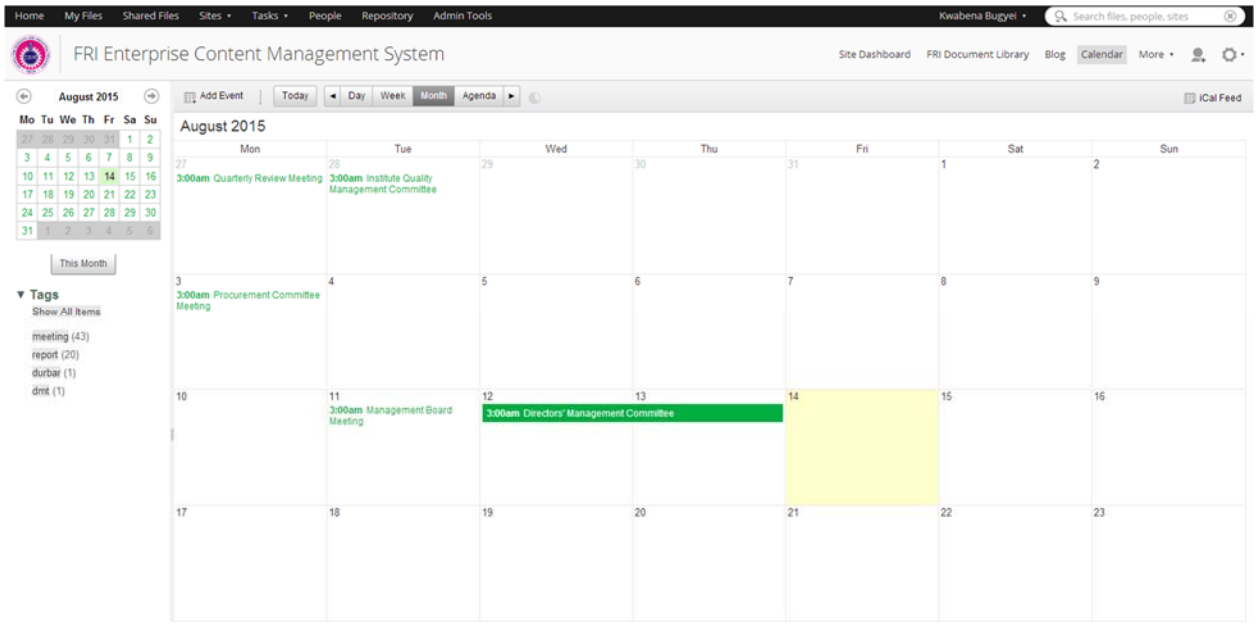
8.9 Uploading document into the system

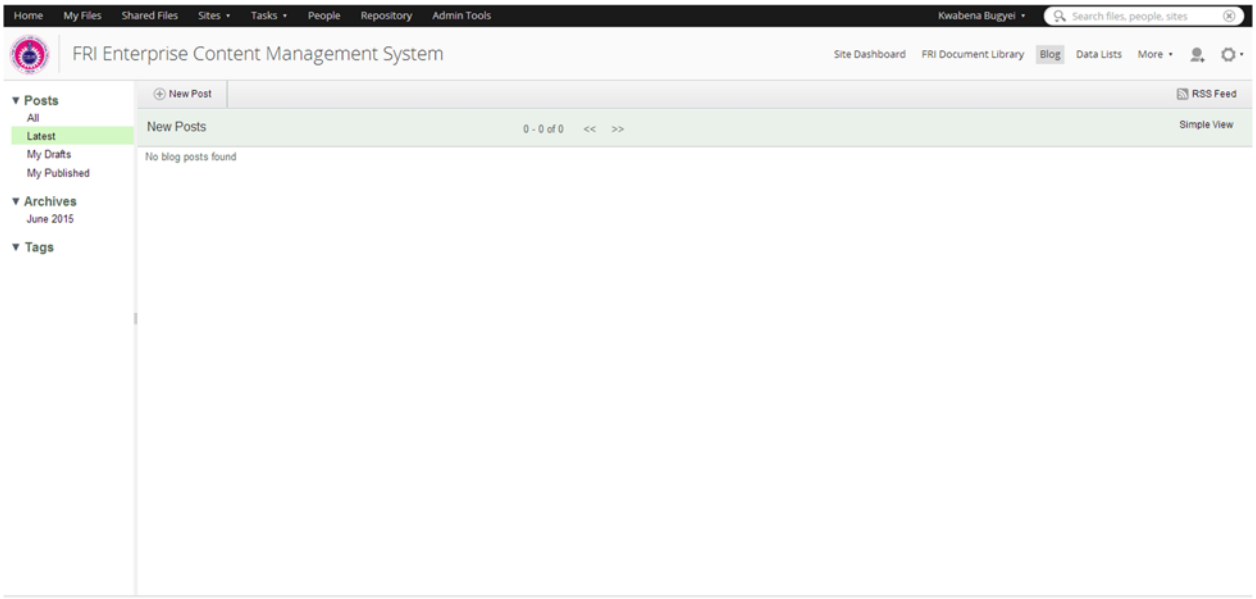


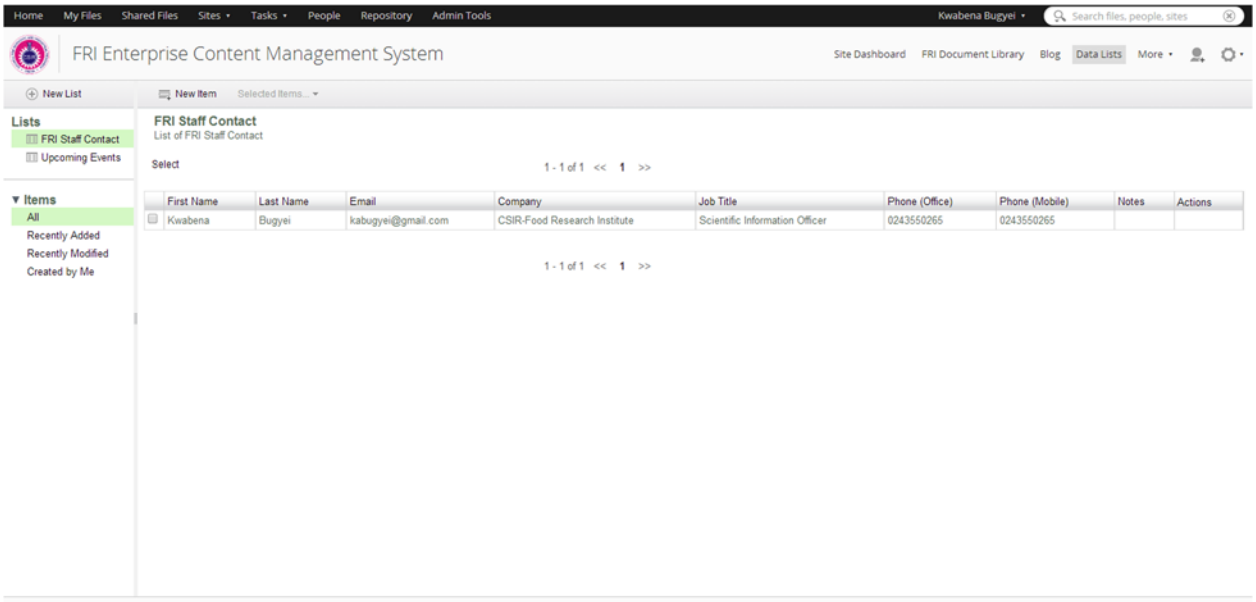
8.10 Workflow management

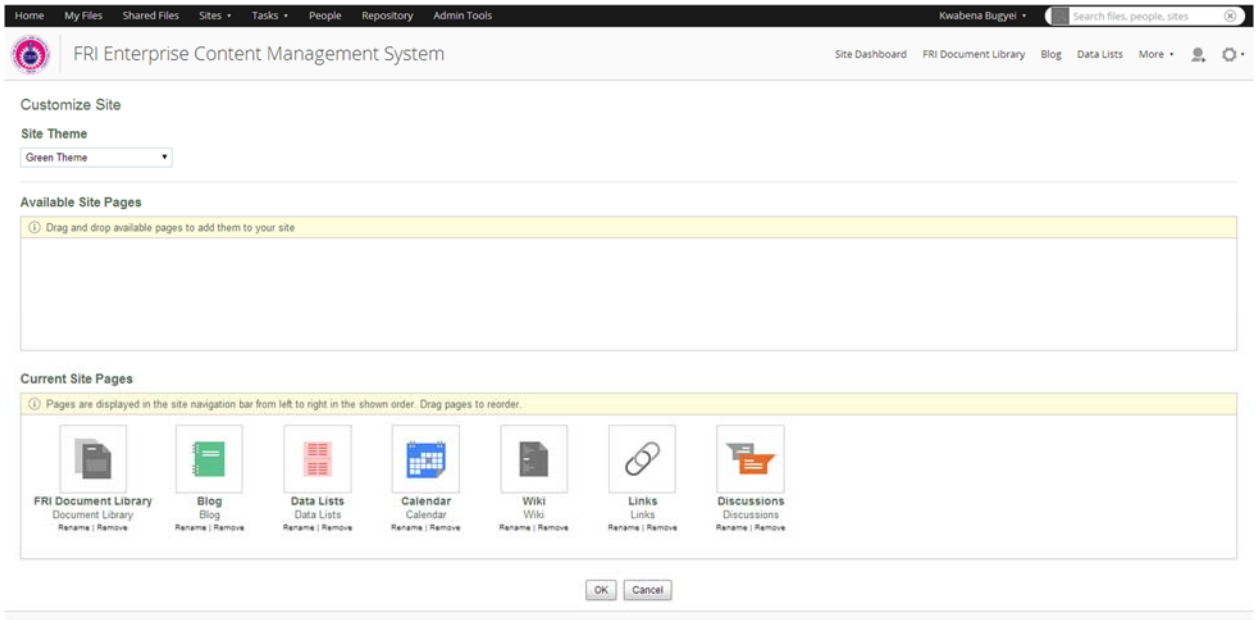


8.11 Collaboration features




Home
My Files
Shared Files
Sites
Tasks
People
Repository
Admin Tools
Kwabena Bugyei
Search files, people, sites
FRI Enterprise Content Management System
Site Dashboard
FRI Document Library
Blog
Data Lists
More
Posts
All
Latest
My Drafts
My Published
Archives
June 2015
Tags
New Post
RSS Feed
New Posts
0 - 0 of 0
<<
>>
Simple View
No blog posts found


Home
My Files
Shared Files
Sites
Tasks
People
Repository
Admin Tools
Kwabena Bugyei
Search files, people, sites
FRI Enterprise Content Management System
Site Dashboard
FRI Document Library
Blog
Data Lists
More
New List
New Item
Selected Items...
Lists
FRI Staff Contact
Upcoming Events
Items
All
Recently Added
Recently Modified
Created by Me
FRI Staff Contact
List of FRI Staff Contact
Select
1 - 1 of 1
<<
1
>>
First Name
Last Name
Email
Company
Job Title
Phone (Office)
Phone (Mobile)
Notes
Actions
Kwabena
Bugyei
kabugyei@gmail.com
CSIR-Food Research Institute
Scientific Information Officer
0243550265
0243550265
1 - 1 of 1
<<
1
>>


Home
My Files
Shared Files
Sites
Tasks
People
Repository
Admin Tools
Kwabena Bugyei
Search files, people, sites
FRI Enterprise Content Management System
Site Dashboard
FRI Document Library
Blog
Data Lists
More
Customize Site
Site Theme
Green Theme
Available Site Pages
Drag and drop available pages to add them to your site
Current Site Pages
Pages are displayed in the site navigation bar from left to right in the shown order. Drag pages to reorder.
FRI Document Library
Document Library
Rename | Remove
Blog
Blog
Rename | Remove
Data Lists
Data Lists
Rename | Remove
Calendar
Calendar
Rename | Remove
Wiki
Wiki
Rename | Remove
Links
Links
Rename | Remove
Discussions
Discussions
Rename | Remove
OK
Cancel

9.0 EDITORIAL COMMITTEE OF CSIR-FOOD RESEARCH INSTITUTE

1.	Dr. Nanam T. Dziedzoave	-	Director/Chairman
3.	Dr. W. A. Plahar	-	Chief Research Scientist
4.	Mr. Stephen Nketia	-	Scientific Secretary