Harri Viisainen

Collaborative Document Management Systems

Metropolia University of Applied Science

Degree Programme in Media Engineering

22.4.2013

Metropolia

Tekijä Otsikko Sivumäärä Aika	Harri Viisainen Dokumenttien hallintajärjestelmä pilvipalveluna 63 sivua 22.4.2013
Tutkinto	insinööri (AMK)
Koulutusohjelma	mediatekniikka
Suuntautumisvaihtoehto	digitaalinen media
Ohjaajat	sovelluspäällikkö Jorma Tiihonen yliopettaja Harri Airaksinen

Yrityksen käsittelemien sähköisten dokumenttien määrä on nykyään suuri, ja siksi dokumenttien järjestelmällisellä hallinnalla on keskeinen rooli yrityksen työnkulussa. Internetteknologiat ovat mahdollistaneet ohjelmien suorituksen pilviympäristössä, jolloin yritys hankkii käyttöoikeuden haluttuun palveluun ja maksaa vain sen käytöstä.

Insinöörityössä kartoitettiin dokumenttien hallintajärjestelmän vaatimukset pilvipalvelun näkökulmasta. Työn tarkoituksena oli dokumentoida tulevan hallintajärjestelmän toiminnalliset ja liiketoiminnalliset vaatimukset, esitellä jo olemassa olevia pilvipalveluna toimivia hallintajärjestelmiä sekä verrata niiden ominaisuuksia keskenään. Tätä varten käytiin läpi asiakasyrityksen olemassa olevaa teknistä dokumentaatiota ja haastateltiin asianomaisia. Työn lähtökohtana oli, voidaanko dokumenttien hallintajärjestelmä korvata vastaavanlaisena pilvipalveluna.

Työn tuloksena saatiin selvitys asiakasyrityksen nykyisen dokumenttien hallintajärjestelmän ominaisuuksista, joita ovat mm. tietokantaan tallentaminen, metadatan rakenteen määrittäminen, dokumenttien versionhallinta ja tunnisteiden määrittäminen, monipuoliset hakutoiminnallisuudet, dokumenttien salaaminen, dokumenttien esikatselu ja kommentointi sekä dokumenttien massasiirrot. Liiketoiminnan vaatimukset tulevalle hallintajärjestelmälle ovat, että se sopii niin pieniin kuin isoihin projekteihin ja tukee dokumenttien jakamista projekteihin osallistuvien henkilöiden kesken. Myös järjestelmän käyttöönoton ja hallinnoinnin on oltava tehokasta. Hallintajärjestelmän kustannuksia voidaan arvioida projektin koon, tilantarpeen ja dokumenttien siirtomäärien perusteella. Olemassa olevien pilvipalveluna toimivien hallintajärjestelmien vertailussa selvisi, että yksikään järjestelmä ei sellaisenaan täytä kaikkia vaatimuksia, vaan lisäominaisuuksia tarvitaan.

Insinöörityön tuloksia käytetään asiakasyrityksessä lähtökohtana siirryttäessä uuteen pilvipalveluna toimivaan dokumenttien hallintajärjestelmään.

Avainsanat	pilvipalvelut, dokumenttien hallinta, vaatimusmäärittely

Author Title Number of Pages Date	Harri Viisainen Collaborative document management systems 63 pages 22 April 2013
Degree	Bachelor of Engineering
Degree Programme	Media Technology
Specialisation option	Digital Media
Instructors	Jorma Tiihonen, Application Manager Harri Airaksinen, Head of Department

The volume of the electronic documents that the company nowadays has to manage is high. Therefore the systematic document management has a significant role in company's working process. The Internet has enabled that the software can be acquired as a service that operates in a cloud environment. In that case the company has the use of the software and pays only for the use.

This study gathered the requirements for the cloud based documentation management system. The purpose of the study was to document both functional and business requirements for the document management system and present some existing cloud based document management systems and their features. The technical documentation of the customer company's current document management system was studied and project participants were interviewed. The initial assumption for the project was to find out, if the cloud based document management system could replace the existing system.

As a result of the study the features of the customer company's current documentation management system were listed. These features among other things are storing the documents into the database, configuring the metadata and its structure, handling the documents' versions, versatile search functions, documents encryption, previewing of the documents, commenting, and the mass transfer of the documents. The business requirements for the system are that it is scalable and it supports work sharing between project participants. Also the setup and the usage of the system must be efficient. The costs for the document management system can be estimated with the information of user volumes, disk quota, and the data traffic. When comparing the features of the existing cloud based systems, it was found out that none of the systems did meet all of the requirements.

The study works as a basis for the migration to the new cloud based document management system.

Keywords	cloud computing, document management, requirements engineering

Contest

Terms, acronyms and abbreviations

Cloud computing	Data processing that is offered as a service over a network (Internet) instead of executing applications in an end user's personal computer.
ASP	Application Service Provider
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
API	Application Programming Interface is a protocol to be used as an interface by software components to communicate with each other.
VPN	A Virtual Private Network is network technology that creates a secure network connection over a public network such as the Internet or a private network owned by a service provider.
IT	Information Technology
TCP/IP	Transmission Control Protocol and Internet Protocol. It is a connection that provides end-to-end connectivity. It also specifies how data should be formatted, addressed, transmitted, routed and received at the destination.
CAD	Computer-aided design.
SLA	Service Level Agreement. An SLA is a document that describes the minimum performance criteria a provider promises to meet while delivering a service.
System	References to Collaborative Document Management system in this document.
EDP	E-mail Document Package is a tool and procedure to send documents to other project participants regardless their access to system. This tool is an effective way to send several documents, linked or zipped file by one single e-mail.
AECO	Abbreviation for Architecture, Engineering, Construction and Operations.
CAD	Abbreviation for Computer Aided Design.
CRM	Abbreviation for Customer Relationship Management.
ERP	Abbreviation for Enterprise Resource Planning.
WebDAV	Microsoft Window Web Folders

CMIS	Content Management Interoperability Service
LDAP	Lightweight Directory Access Protocol

1 Introduction

This thesis consists of a literature study and a case study. The literature study focuses on describing the principles and concepts of the cloud computing, document management systems, and the requirements engineering process. The case study describes the document management system used in Pöyry and what requirements it sets for the forthcoming document management system that preferably operates in a cloud computing environment. The case study also looks into the market to find out what kind of cloud based document management systems are available, and what kind of features they offer.

Pöyry is an international consulting and engineering company that operates globally across the energy and industrial sector. Pöyry is focused on delivering solutions for power generation, transmission and distribution, forest industry, chemicals and bio refining, mining and metals, transportation, water and real estate sectors. Pöyry is planning to renew their current MS SharePoint based document management system (DocHotel) to the cloud based service. With a cloud computing service Pöyry wants to offer additional value for its clients as a business partner, who can also offer additional services with the consulting, such as setting up and managing the documentation project.

The case study covers both the requirement analysis for the document management system operating in a cloud environment and the feasibility study of the existing cloud based documentation management systems and their features. It lists the requirements from a functional and business point of view. In the feasibility part five existing cloud based documentation management system are presented and their features are compared against each other.

2 Cloud computing

The basic idea of cloud computing is that both the data service and the needed infrastructure can be achieved from the Internet, so that the end user does not necessarily have to own either software or, in a larger scale, authorise needed data servers to perform the tasks he or she needs to do in the daily business. The same way the IT developers could design and implement new software products in a global environment. The necessary development tools are available 24/7 for production from everywhere, only the Internet connection is needed. The elastic nature of cloud computing enables also the needed data resources to be easily be scaled up or down based on the demand. [1].

Before the cloud computing term became widely known, the terms Network Computer and Network Computing were used in the last decade of the 20th century. The Network Computer term was introduced in 1996 by Oracle's chief executive officer Larry Ellison. The term Network Computer was also known as a "thin client". It is a computer without a local hard drive for storing data; instead it will use both the applications and the data from the remote data servers via TCP/IP network. Thin client is a concept that was defined by Oracle, IBM, Apple, and Netscape. [2].

Cloud computing should not be mixed with the term grid computing, though cloud computing is based on grid computing. In grid computing multiple computers are connected together to share their computing power for applications, i.e. it requires that each computer in a grid must have a framework that supports grid computing. One of the most known grid computing projects is a SETI (Search for Extra-terrestrial Intelligence) @Home, where users all over the world share their unused computing power to analyse signals from outer space. Figure 1 illustrates the grid computing network. [3].

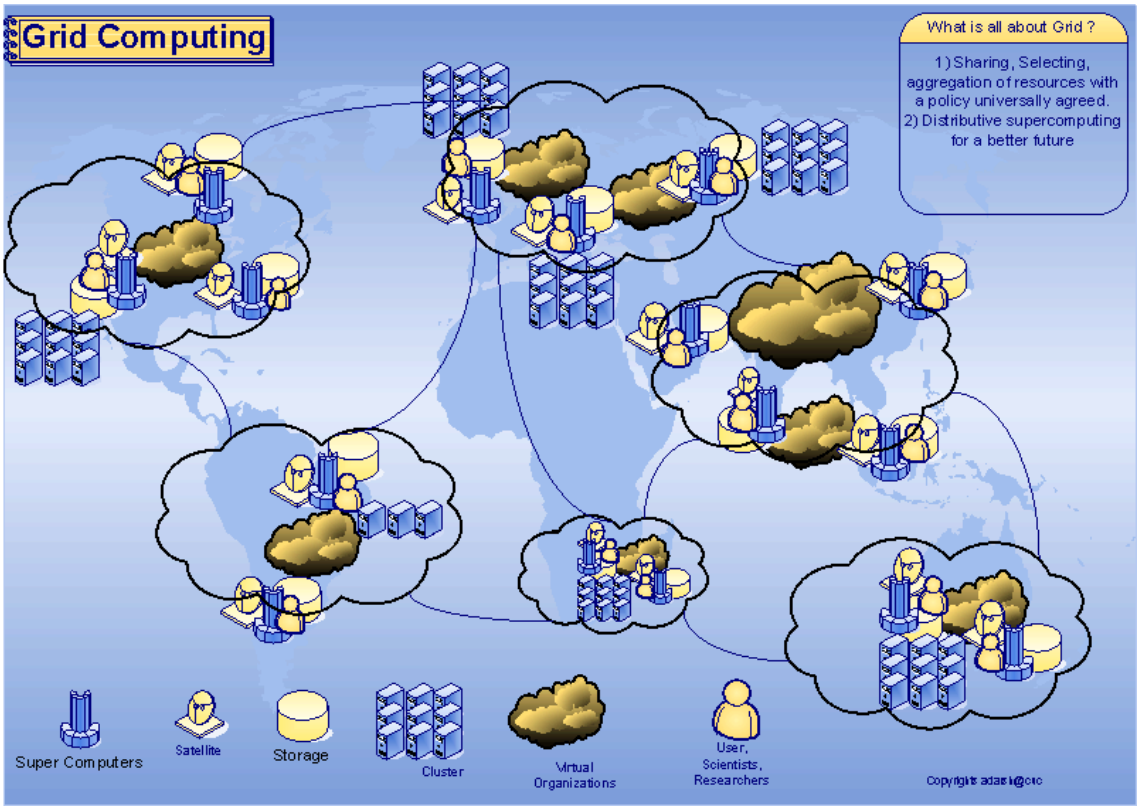

Figure 1. The grid computing network. [3].

With parallel computing power the grid network is a true supercomputer that can save a lot of money and time, if every computer's unused capacity worldwide could be harnessed to perform computation. [3].

2.1 History

Term cloud computing originates from the era when engineers started to document the structure of existing telecommunication and data communication networks. Often the nature of those networks was complicated and they included a lot of different devices, so it was more useful to start to use the cloud symbol to illustrate those networks. The cloud symbol also described the application interface between the end user and the service provider, not just the physical components of the network [4]. The first cloud computing environment included a group of clustered computers that could be used to offer both economical and effective computing power [5]. Before the cloud computing term became widely used in the middle of the first decade of the millennium, the hosting services already existed. They were referred to as ASP-services (Application Ser-

vice Provider) at the time. The term ASP was used to describe services such as email and web site hosting. The main difference between ASP and cloud computing is that in cloud computing the pricing is based on usage of the service, instead of fixed monthly or yearly based costs [6].

2.2 Background and usage

In cloud computing the end user only rents both infrastructure and applications as on-demand from the service provider via the Internet, who owns and manages the physical architecture. Because cloud computing can include anything from data storage services to the specific Web applications, the offered infrastructure must be scalable and it should adapt easily to the changes, based on end users' demands. That can be achieved by utilising virtualisation techniques to multiply one physical server into many virtual machines, and that way maximise the hardware capacity for the end users' benefit. [6].

The benefits

From a business point of view the goal of starting to use cloud computing services is to take advantage of shared resources that are in cost-effective use and by doing so reduce the overall IT-expenses. Typically companies reach following business benefits:

- A faster way to acquire latest versions of the applications compared to the traditional IT deployment model.
- To reduce overall IT-expenses.
- To free employee resources from the administration tasks to the development of the IT-infrastructure.
- To achieve more flexible IT-infrastructure by taking advantage of shared applications and straightforward information.
- To optimize business process by using standardised IT-infrastructure.

The drawbacks

The coin has the flip side too. Despite the many benefits cloud computing offers, it also has some drawbacks that should take into consideration.

The service downtime

No matter how reliable the internet connections nowadays are, still the provider outages occurs occasionally, meaning that the business suffers when offline. Especially for the smaller companies it could have harmful consequences. [7].

Security and privacy issues

How to verify that the information is secured and available only for the persons with the adequate permissions? Also in cases where the service provider will be changed to another service provider, a question of data deletion from active storage and backups arises. Not to mention the denial-of-service kinds of attacks and how the service provider has prepared for them. [7].

Cost issues

If the cloud service does not fulfil all the functionality requirements of the company, it is important to calculate the extra cost in order to achieve them. Also the comparison for the total expenses over an application's life cycle between cloud services versus on-premises version would be advisable. [7].

2.3 The service models

The cloud computing includes three different kinds of service models (figure 2) for separate end users' needs: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

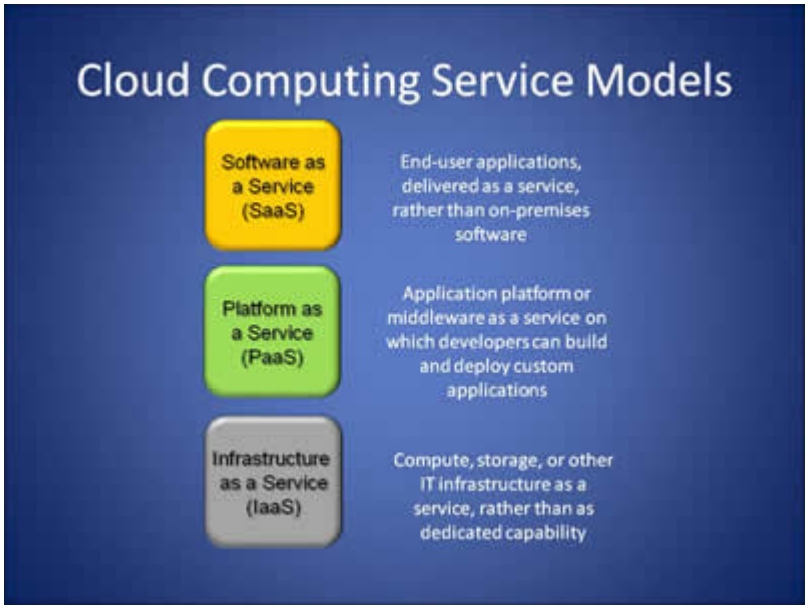

Figure 2. The service models of cloud computing [6].

The cloud computing service models are layered, by a technical implementation, so that the lowest layer contains infrastructure as a service model, which offers the IT infrastructure for computing and storage services. On top of that locates a platform as a service model that offers application platform and middleware as a service for developers. The uppermost layer contains software as a service model that offers applications for end-users. The technical structure defines what kind of computing and services the layer offers, and how it can be connected. [6].

Infrastructure as a Service

Infrastructure as a Service model offers to the end user the whole repertoire of service provider's hardware and technologies for computing power, data storage, or operating

system. Service can include virtual machines, which end user can use as a personal file server, although it locates on the Internet. Figure 3 illustrates the structure of the IaaS model [8]. For instance the Amazon Simple Storage Server represents the IaaS model. [9, 6]

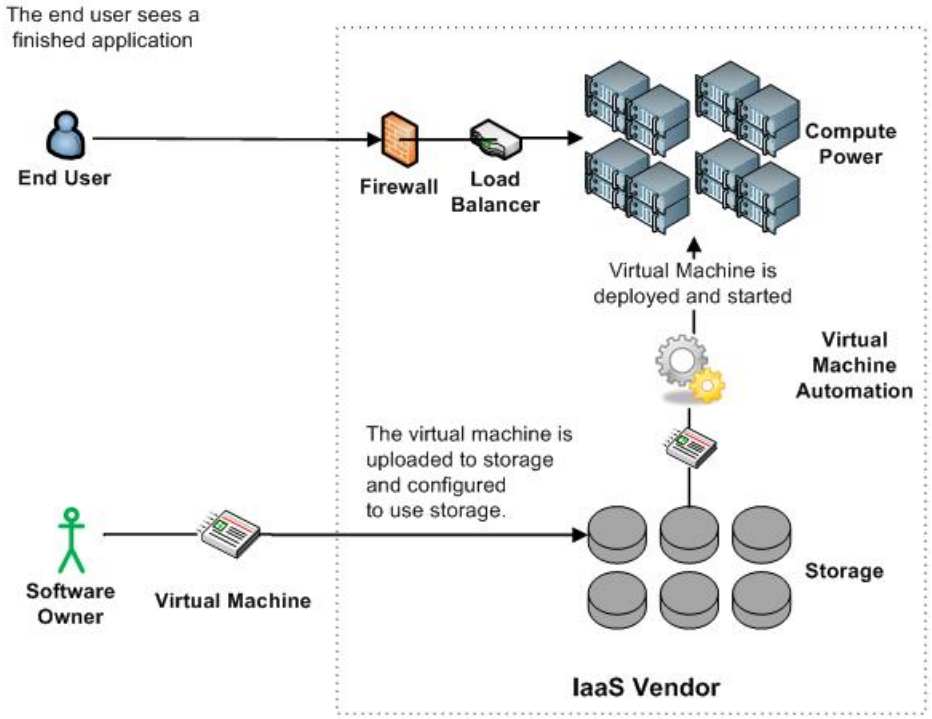

Figure 3. The structure of the IaaS model. [8].

The service provider maintains the virtual machines on the Internet and provides sliced partitions for the customer, who then uses these slices, as they were part of local infrastructure, by installing operating systems and applications on them. The customer does the infrastructure's management via console application and the end user uses the applications with the web browser. [6].

Platform as a Service

Platform as a Service model is meant for software developers, who can utilise the platform and middleware for software development purposes. Service provides the platform with necessary tools and application programming interfaces (API) to develop, build, and test database driven applications. Hybrid cloud solutions of the PaaS allow a developer to execute applications either in a cloud environment or in a local data cen-

tre. Figure 4 illustrates the structure of the PaaS model [10]. For instance the Microsoft Windows Azure platform represents the PaaS model. [9].

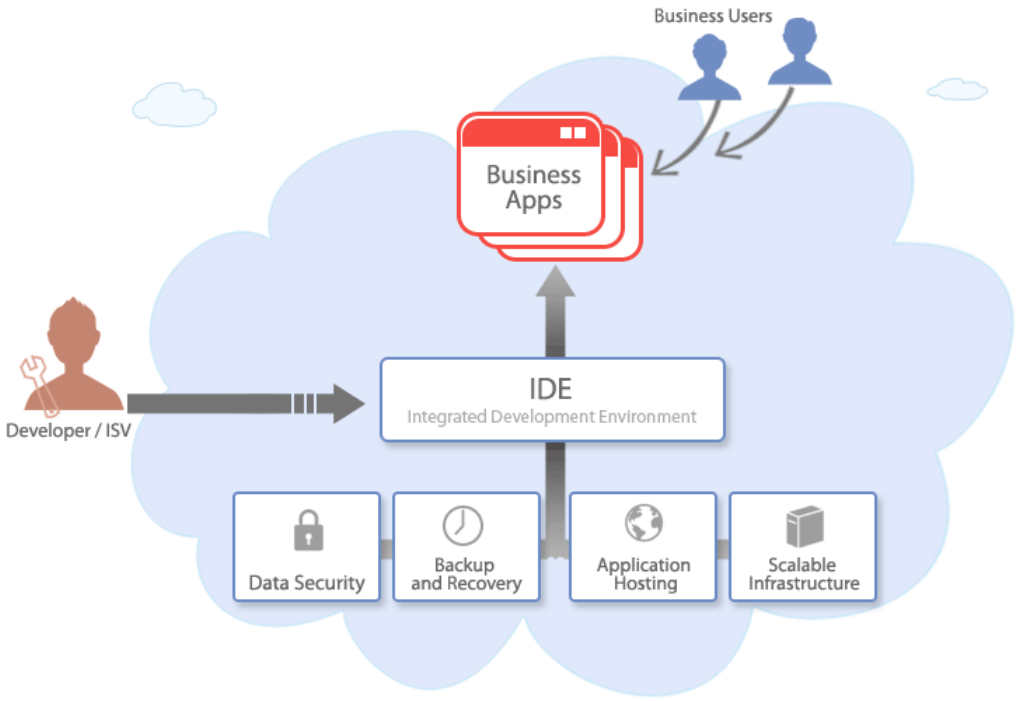

Figure 4. The structure of the PaaS model. [10].

The software developer manages the development of the business applications with the integrated development environment, which provides also services for the data security, backup and recovery functions, application hosting, and scalable infrastructure. [6].

Infrastructure as a Service

In Software as a Service model end user uses applications that locates on the Internet instead of a local computer. It enables end user to run applications that needs more computing capacity than his or her own computer has, because the needed computing is executed in the cloud. End users also do not have to worry about the application's update procedures, it is done automatically by the service provider. Figure 5 illustrates the structure of the SaaS model [11]. Adobe Creative Cloud and Google Docs are examples of the SaaS model. [6].

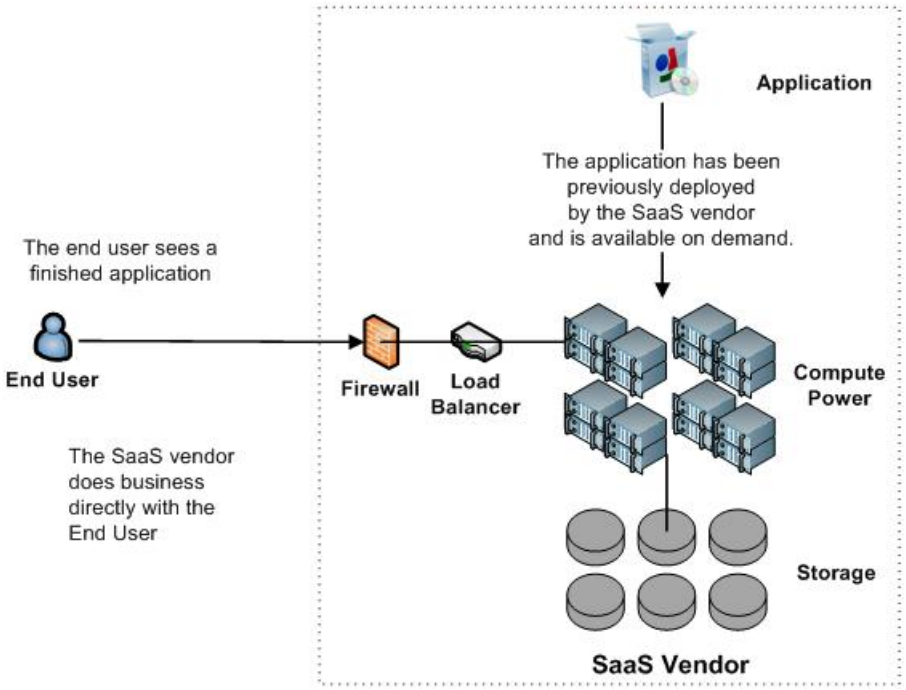

Figure 5. The structure of the SaaS model [11].

The end user has an access to the needed applications with the web browser. The service provider's role is to take care of the application management. [6].

Public cloud

With a public cloud model the user gets computing capacity and resources via Internet. The service provider owns and manages the needed infrastructure, and offers to the end user the connection, usually with the secured VPN connection. The user pays only for the usage of the service, which can be hour-, monthly-, or other time based invoicing. The service provider may also charge extra from the storage-, and server capacity, and the data communication expenses. [6].

Private cloud

In a private cloud model the user owns and manages the needed cloud infrastructure. Because the service locates in a LAN or other secured network, there is no need for an external connection. The user must cover all expenses, unless the connection is available also for some other users, when they all share the expenses. This variation is called as community cloud. For instance the government could utilise the community cloud with its bureaus, instead of building separate technical infrastructure per bureau. [6].

Hybrid cloud

The hybrid cloud model is a mixture of a public- and a private cloud. In a hybrid cloud model the company who has its own private cloud, forms a partnership with a vendor who offers public cloud services via Internet. This allows the company to manage some resources by itself and buy other resources from the service provider. For instance the company may use the Amazon Simple Storage Service for data archiving but still maintain the operational data in its local storage. By using the hybrid cloud model, the company can take the advantage of the scalability and cost-effectiveness of the public cloud and the same time keep the critical applications and data by itself. [6].

Intercloud

The intercloud model doesn't yet exist at the time of writing this. It is a hypothetic idea of the cloud model that meshes all the resources from different clouds and provides services from them based on availability, performance, user location, or business focused needs. The major hinder for the intercloud is the lack of the network automation services, because most of the work that keeps the network running and applications available at today is based on manual processes. [12]

2.4 Security of the cloud computing

The nature of the cloud computing services requires that the data and the services locate outside of the company's control. That requires companies to rethink their data

security and issues that are related to the individuals' right to privacy. The firewall is not enough anymore to protect companies' digital assets, instead the companies must renew their processes that are related to IT and the data security. The companies should standardise their practises and technologies that are related to the data security, before they start thinking the use of cloud computing services. When the organisation has a solid understanding of the information security, it would also be easier applying that policy outside of the company's network. [13].

From the security point of view the cloud computing is a suitable solution for companies, when they do not have to transfer information that is critical for business, or use applications over Internet that handle such information. That kind of policy is especially good with:

- Distributed server and data centres.
- Software development environments.
- Applications that are used seldom and do reserve lot of data resources when executing them, such as systems for testing and preproduction purposes.
- Backup systems.
- Time management applications.

The challenges in cloud computing's data security can be classified into four categories: the service management, the risk management and the continuation of the business, the propriety rights and transfer of the data, and the legal aspects. Figure 6 lists those categories in more detail. [13].

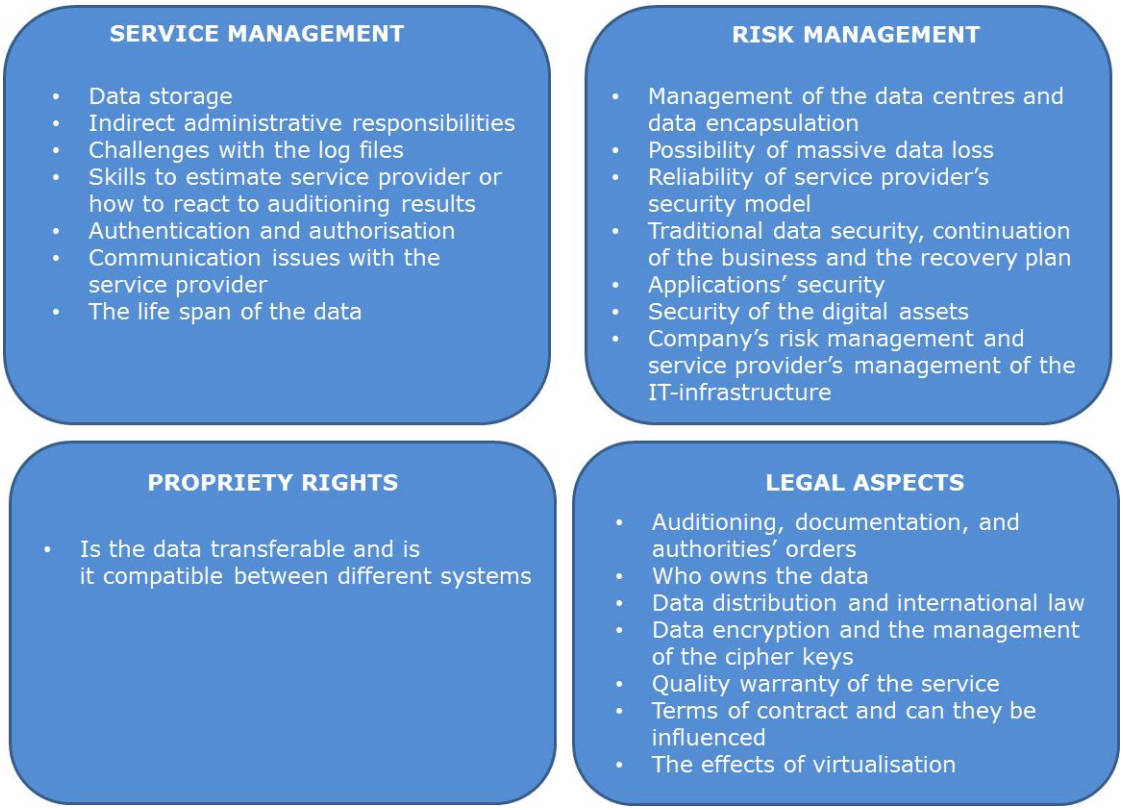

Figure 6. The four categories of the security challenges [13].

The verification of service provider's level of service and the data security may be difficult, because of the lack of standardised practises. In cases where multiple customers' data locates in a same cloud service or the customer's data is distributed to the different data centres, or where the administration is managed by someone else than service provider, the verification of data security's quality is nearly impossible. [13].

The planning of secure cloud model gets even more complicated, when only part of the IT-infrastructure is transferred to the cloud environment. In that case applications and the information, which will reside in a cloud, need specified security levels that can adapt to the changing requirements of the flexible cloud models. For instance the services inside an intranet need to be updated first so that they can be integrated to cloud service. Also the requirements from the authorities or the sensitive material that is relevant to the individual's privacy need specified security standards, procedures, and guides. [13].

When the security is in the sufficient level?

In an acquisition phase the organisation usually either accepts the service provider's service level agreement (SLA) or not. The SLA describes the content of the service and also the security and quality issues. When operating in a private cloud model, the organisation has a better chance to affect the service's security levels. But with the "big players" like Google or Amazon, the organisation has to take what it gets. For an organisation that plans to start using cloud services, it is important to understand what kind of information will be transferred to the cloud. And which criteria weigh more on a scale, the confidentiality or the availability of the data? [13].

The service documentation should be accessible from the service provider, as it can inform the buying organisation is the service architecture in a decent technological level what comes to the security issues. It is good to keep in mind that even though the documentation informs that everything looks good, it is no guarantee that the system is built as planned. The audition procedure is a good way to verify that the system is implemented as planned. The service provider should also have a process to manage vulnerabilities especially with the web applications, but also when they concern service provider's IT-infrastructure. For instance the organisation gets into a denial-of-service kind of attack. Finally the cloud service should offer tools for observation and tracking of data security threats, such as log file management, intrusion detection systems, and security incident event management. [13].

It is good to keep in mind that the good practises in data security are usually a better way to secure trusted information than just technological solutions. [13].

2.5 Amazon Web Services

Amazon.com is an eCommerce company, founded in 1994, that started as a bookstore in Internet, but nowadays sells almost all kinds of consumables online. The Amazon Web Services was started in 2002 and it is a collection of many smaller subsystems. Amazon started to share gathered data, which it collected from eCommerce, to the third-party developers. The core of the Amazon Web Services is Elastic Compute Cloud (EC2) service, which operates the clustered virtual servers needed by Amazon.

Amazon Web Services also offers long-term storage capacity for its customers. The Simple Storage Service (S3) shares containers (buckets), where customers can keep their data in a cloud. The bucket is a simple directory based solution and its storage size is limited to five gigabytes. The files in S3-service have a unique identifier that also operates as an address to the information. For instance if the user has stored a book into the S3-service, he or she could use http://s3.amazon.com/ph/book.docx address to fetch that file. The “ph” refers the bucket’s name and “book.docx” is the name of the file. [4].

Like other players in a cloud service business, also Amazon had to find a way to get income from its services. Amazon Flexible Payments System (FPS) is an interface that third-party developers can use in their applications to handle payments. The EC2 and S3 use the Amazon DevPay-interface to handle monetary transactions. Amazon charges the end user via DevPay and accounts the income to the developers. For instance the software company, who offers SaaS-service to its customers from EC2, can take the advantage of the DevPay without implementing own payment solution. Table 1 lists the sub services that are part of the Amazon Web Services. [4].

Table 1. The sub services of the Amazon Web Service

The service	The description
Amazon CloudWatch	The monitoring and reporting service
Amazon CloudFront	The content delivery networking. For instance to deliver eLearning material globally.
Amazon Virtual Private Cloud	The VPN-tunnelling solution between Amazon cloud service and customer’s local area network.
Amazon Simple Queue Service	The solution to manage messaging between applications in a cloud.
Amazon Relational Database Service	The MySQL database for the use of EC2 instances.
Amazon SimpleDB	The light database version to the cloud applications that don’t need the power of MySQL relational database.

The Amazon has been Infrastructure-as-a-Service business since 2006, and it is estimated that in 2014 its business volume will be two and half billion dollars and income about 400 million dollars. [4].

2.6 Google AppEngine

The Google AppEngine is the core of all Google services. It is a PaaS-service that offers tools to build applications and databases. Google has developed a Software Development Kit tool for third-party developers to implement and test Java and Python based applications. Every application has its own “sandbox” that separates it from the operating system, hardware, and physical server location. For this sandbox Google AppEngine offers needed services, such as layers for web services and interfaces to manage end user authentication and emails. Google AppEngine takes care of load balancing and how scalable application will be. Datastore is a Google’s way to handle saving of the permanent data. It uses a BigTable database service and the data is controlled with Google Query Language (GQL). Although Datastore is meant to store small data entities, it also enables the storage of the large data entities with the Blobstore module. The size of a one “blob” is two gigabytes. Google offers both consumer and enterprise version of the AppEngine, where enterprise version is charged on a monthly base and the consumer version is free. Table 2 lists other cloud services of the Google. [4].

Table 2. The other cloud services of the Google,

The service	The description
Google Maps	The free application for map and route information, including the Street View feature. For developers it offers an application programming interface (API)
Google Latitude	The application, where end users can share their location information.
Google Translate	The translating application.
Google Patents	The joint project between the Google and USA authorities related to patents and trademarks in Internet.
Google News	The news service, in where Google put together a news page using just an algorithm.
Google Voice	The voice mail service from Google.
Google Earth	The digital version of the earth, also Google Moon and Google Mars exist.

The other part of Google cloud services is the Google Apps. It has similar features as Microsoft Office, e.g. it contains an email (Gmail), Calendar, Drive for documents like slides, spreadsheets, drawings etc. Google Apps supports common file formats, including Microsoft Office. One can say that Google Apps is a suitable solution for households to manage electronic documentation and communication. [4].

3 Document management systems

During the last decades the amount of electronic documents has increased substantially, and at the same the need of organising the produced data has become more important. While the Internet and email-servers have enabled the document sharing easily, also the amount of nonessential data has increased. At the same time the information should be available not depending of the location or time of the day. So how to find the essential and up-to-date information from all available data that is on offer? [14].

The document itself can be understood as an entity that is examined by a person. The paper has been the traditional form of the document, but nowadays more and more of the documents are generated digitally and then stored on a computer's hard drive. The type of the electronic document can be the memo written by word processing software, or it can be a digitally scanned version of a paper document. [14].

The individual file that is stored on a computer's hard drive, without the information of the file's content, is just a file. The file's attributes only define the nature of the document, for instance that the file in a question is a memo from a company's monthly meeting held on a particular date. In a digital form the document is the sum of its attributes, and the better they are defined, the better the documents can be individualised from the vast mass of other documents. In the beginning of the nineties there was a lot of talk about paperless offices, but as we have seen they do not exist anymore. People still like to read, and even archive documents in paper form. People send and receive the documents in a digital form, but still like to print them before reading. [14]

3.1 Overview

Document management is a way to process documents' lifespan from the creation to the archiving and not just a means to organise it technically. The development of the IT has enabled more applications to produce the documentation, which has led us to the situation where it is harder and harder to find the needed document. It is also possible that the documents are overwritten accidentally, or that the document will be re-written because the existing document, or the latest version of it, is too hard to find from the mass. [14]

The computer's file architecture is not an efficient way to manage and organise company's documentation. The file name itself and the directory where the file locates are not the best way to find and classify the documents. For a one person and his or her personal files on a local hard drive it may be enough, but when it is a question to manage and organise the documentation produced by many users, then it is time to start using document management system. [14].

The document management system is an application that can both manage the documentation files and their attributes, which defines what kind of documents are in question. The documents locate in a network drive in a defined folder structure and the attributes are usually stored on a database. The user can search the documents with the attributes and also with the content of the documents. The document management system manages the versioning of the documents and takes care that one user at a time can modify the individual document. It also grants permissions to the documents so that the user or group of users can either read or modify specified documents. [14].

The document management system supports different tasks in a document's lifespan. The author creates the document and at some point marks the document ready for a review. The specified other users then act as a reviewers for the document and give their feedback to the author, who then modifies the document with a given feedback. After the review phase is over some user(s) acts an approver(s) and gives permission to publish and distribute the document to targeted users. Figure 7 shows the lifespan of the document [14]

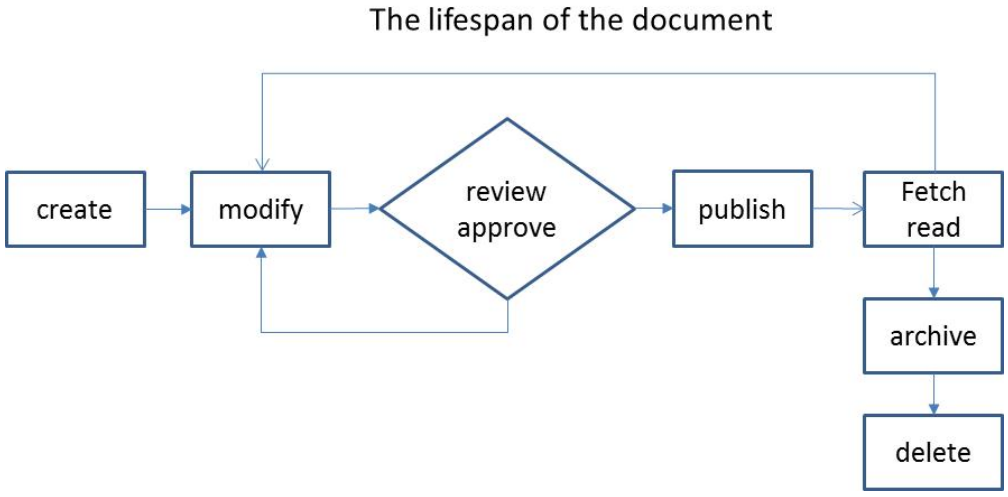

Figure 7. The document's lifespan starts from the creation and ends to the deletion [14].

With the document management systems it is also possible to organise documents in a paper format. In that case the documents' attributes are not known, we only know where the document physically locates. [14].

Basic functionalities

The difference between traditional file based document saving and saving the documents into the document management system is that instead of defining the folder and name of the document, the user enters the document's attributes. By using these attributes the user can later search and fetch the documents from the document management system. Regardless of the document's file type or who is the author of the document, the search method is always the same; either the document in question is a Microsoft Excel spread sheet, a text file produced with the Microsoft Word, or even a technical CAD drawing. [14].

The version management has also a big role in a document management system. After the document has been fetched and modified by the user, the document management system automatically creates a new version of the modified document into the system. The document management systems enable the grouping and categorising of the documents. On other words the document can be found from several folders, although it is saved only once into the system. [14].

The basic functionalities of the document management system can be summarised as following:

1. it offers an Internet browser as an graphical user interface
2. it stores the documents' attribute information into a database
3. the documents can be classified and they are searchable by the folder structure
4. the documents can be searched by the documents' attributes and content
5. it enables the tools to manage document based user rights
6. it enables the version management of the documents; check-in and check-out functionalities

The document management system manages also the permissions of the documents. It enables the administrators to define who has the rights to search, fetch and modify documents. These permissions can be granted either to the group of many users or just a single users. [14].

3.2 The document attributes

The document attributes describe what type of document is in question. As mentioned earlier they are used to ease the finding of a single document from the mass of documents that are stored into the system. Some of these attributes are defined by the system automatically, and some are defined by the user when he or she creates or modifies the document. Table 3 presents some common document attributes [14].

Table 3. The example of the document's attributes [14]

The attribute	The description	The example
Name	The document's name	Minute 7.2.2013
Author	The owner of the document	Harri Viisainen

Created	The creation date of the document	7.2.2013
Revision	The document's revision	1.1
Revision date	The revision date of the document	7.2.2013
Revision comment	Describes the changes in this revision	The summary was added
Description	The informal description of the document	Meeting with the customer
State	The state of the documents, i.e. draft, reviewed, approved etc.	Draft
Type	The type of the document, i.e. memo, minute, blueprint etc.	Minute

The document type definition is used to classify documents into different document groups. This information can be, for instance, used to select suitable document template, when creating a new document. The document template can include some attribute definitions that are specific only for this kind of document type. [14].

These kinds of definitions can be:

- The document attributes and their default and allowed values
- The tools that can be used to modify the document
- The template document as default, when creating a new document
- The rules for document's version management, like the revision numbering, the maximum amount of different revisions, and the amount of active documents
- The rules when the document becomes obsolete and when the document is archived
- The indexing of the document's content
- The definition of user permissions

- The management of the documents workflow, i.e. the management of creation, reviewing, approving, publishing, distribution and archiving

The document type classifies the usage of the documents and defines the management rules for them. The document type should individualise the document in a certain level from other documents, but too many document types can be difficult to maintain in a system. The document attributes can be used to specify the document more accurate, while the document type defines the document in a generic level. [14].

3.3 Organising the documentation

The folder structure has been used to organise the increasing number of different documents. Compared to use of operating systems' directory structure, the document management system uses dynamic folder structure, which enables the document to be located in a multiple folders, even when it physically locates inside one folder. The folders can also be used to classify documents, i.e. the dynamic folder structure allows the user to group documents differently depending on search criteria. Figure 8 illustrates the use of dynamic folders. [14].

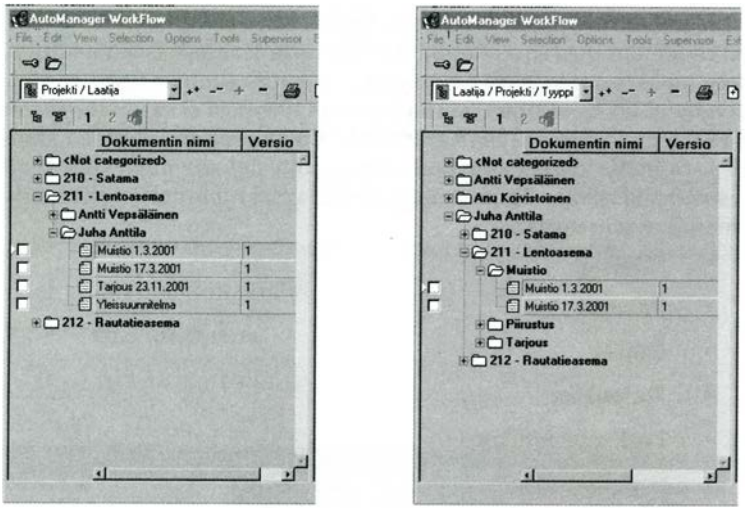

Figure 8. On the left side dialog the documents are grouped by a project and an author. The dialog on the right side groups the documents by an author. [14].

For the user the dynamic folder structure offers an easy and effective method to visualize the documents as groups that also simplify the search process, as the user knows

from what kind of groups the document can be found. Likewise the folder structure instructs with the document attributes, when saving a new document into a system. [14].

3.4 Document searching and revisions

One of the critical features of the document management system is how efficiently it offers searching methods for the user to search documents. Sometimes there is a need just to find a particular single document; on the other hand it should be possible to find multiple documents that share similar of subject matter. Usually documents can be searched by using folder structure, by defining document attributes, by free search, or by searching the content of the documents. [14].

Searching is based on folder structure.

The use of a folder structure for document searching is convenient when the user searches the group of documents. For instance the documents of the particular user locates inside an own folder, or the customer and project based documents can be found own folders. The searching process is similar to searching with windows explorer, but the document management system offers more flexible ways to search documents than traditional static folder structure. [14].

Searching is based on documents' attributes.

By using the document's attributes in a search the user gets more specific results, as the search is definite to use some particular attribute(s). For instance the documents can be searched with the name or number, or the user can combine the attributes like "I want to find all the documents of a project X, where the author is Y and Z.". The document management system can help the user by showing the attributes as preselected lists that acts as a guideline to find the right documents. Figure 9 shows an example of an attribute based search. [14].

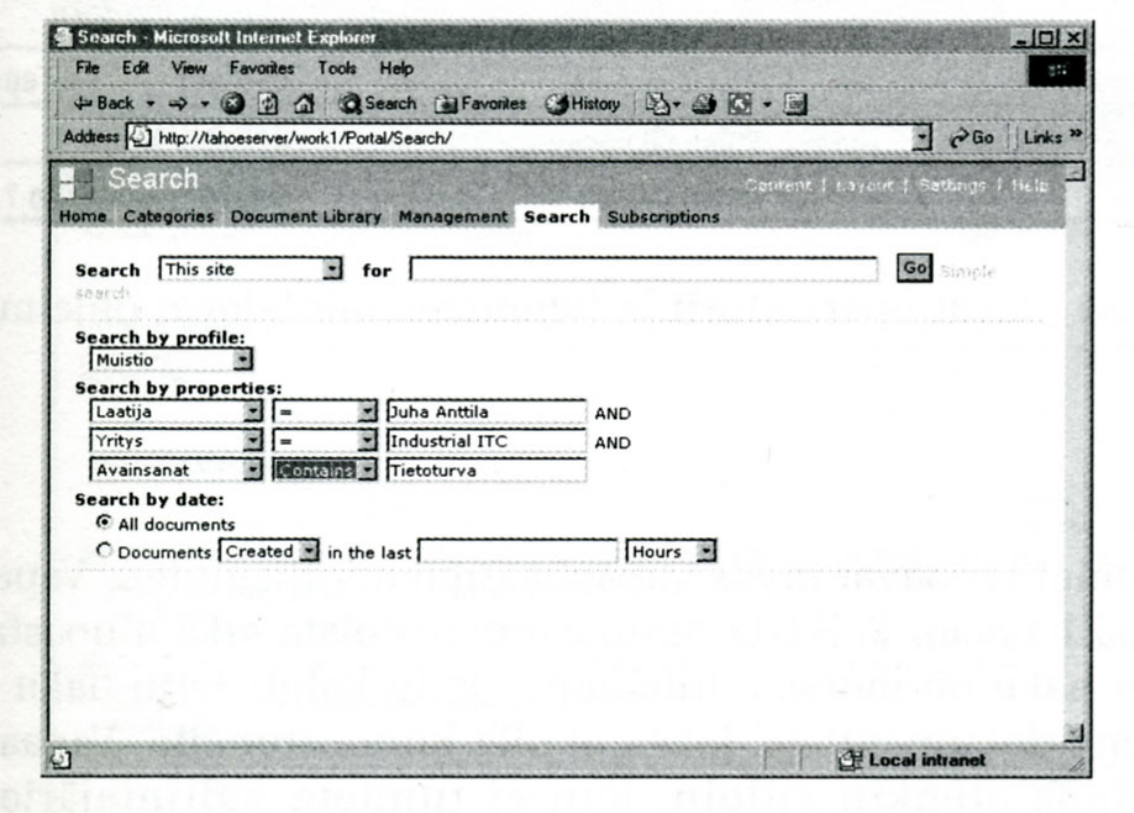

Figure 9. An example of the attribute based search by using Microsoft SharePoint Portal [14].

The document management systems also support the use of arithmetic operators and Boolean operators when defining the search attributes. With them the user gets even more specific search results. In table 4 some of the operators that can be used with the searches are listed. [14].

Table 4. Search operators.

The operator	The search method	The example
=	Equals, is	= Meeting minutes 1
<	Less than	< 10
>	Greater than	> 10
<=	Less than or equal to	<= 20
>=	Greater than or equal to	>= 20
&	AND	> 5 AND < 10

|	OR	= Matti OR = Teppo
*	Asterisk, replaces one or multiple characters	Meeting*
?	Question mark; replaces one character	Meeting minutes ?
	NOT	NOT = Meeting minutes 1
	NEAR; similar as	NEAR Meeting minutes
	Includes, Contains	Contains minutes
	Begins with, Start with	Begins with Meeting
	Ends with	Ends with 1

The supportable operators may vary between different document management systems, but with the inventive use of search operators the user can effectively find the needed documents from the vast mass of other documents. [14].

Search as a free search.

The free search method is usable when the user does not know the how the documents are classified in a system or when there is a need to find all information that is related to the search word. The free search method is slower than an attribute based search and it may also give unwanted results, in other words documents that the user is not interested in. [14].

Searching is based on documents' content.

The content based document search requires the indexing of the searchable documents, just like the information on the Internet is indexed enabling the fast search results. Without indexing the search engine must scrutinize all the documents separately, and that would reserve a lot of computing power and it would also be a very slow operation. Typically only the text based documents are indexed by the document management systems, but not necessarily the graphical documents such as drawings or blueprints. Unlike the attribute based search the content based search returns a large number of documents. Figure 10 shows an example of Microsoft SharePoint Portal's content and attribute based search. [14].

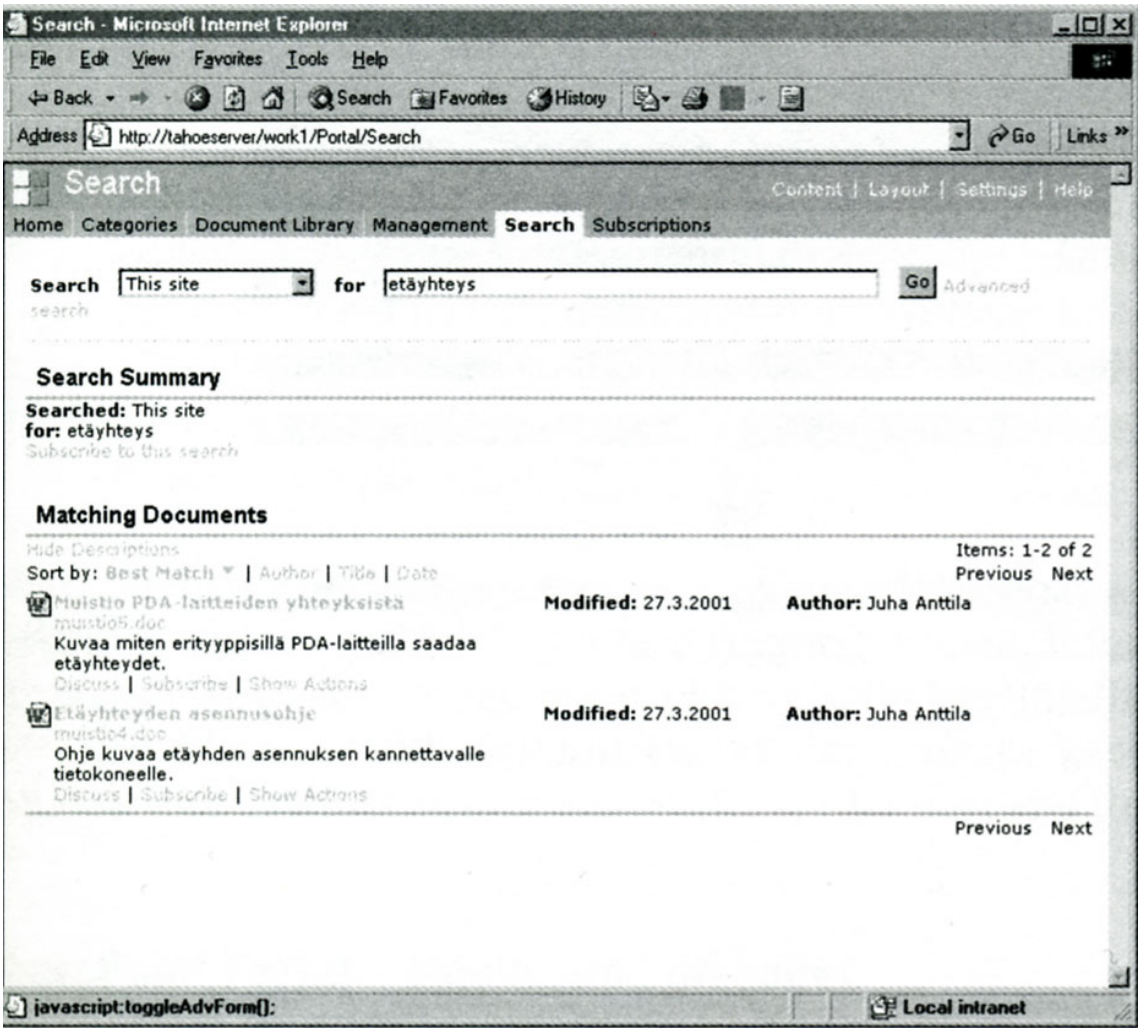

Figure 10. An example of Microsoft SharePoint Portal's content and attribute based search [14].

Some document management systems enable to combine content based search with the attribute based search. In that case the documents' attributes plays a bigger role than the documents' content, and that also gives more specific results to the user. [14].

Revision control

With a document revision control the user can keep track of changes in a document and follow the change history of the documents, and if necessary user can easily check out previous revision of a document from a system. Typically the documents are indicated with major and minor revision numbers. The table 5 shows examples of major and minor revision numbering. [14].

Table 5. Examples of different document revision methods [14]

The description	The numbers	The letters - style 1	The letters - style 2
First version	0.1	0.1	0.1
Second version	0.2	0.2	0.2
Published version	1.0	A	0
	1.1	A.1	A.1
	1.2	A.2	A.2
	1.3	A.3	A.3
Second published version	2.0	B	A
	2.1	B.1	B.1
Third published version	3.0	C	B
	3.1	C.1	C.1
	3.2	C.2	C.2
Fourth published version	4.0	D	C

Major revisions

Major revision numbering is used when the document is published, i.e. when the document has been reviewed and approved. In most cases revision is indicated with numbers or letters, or even their combinations, for instance 1.0, 2.0, 3.0 and so forth. With technical plans lettering is a common way to revision documents, like A, B, C in an ascending alphabetical order. [14]

Minor revisions

Minor revision is used with the documents that are still in a development phase. Minor revision number is added after major revision number (or letter) separated by a dot. For instance, if the major revision is 3.0 then minor revisions are 3.1, 3.2, 3.3, and so on. When the document is ready and approved, the major revision will be 4.0. In cases where major revision is a letter, minor revision can also be a letter, indicating next ma-

jor version, with a minor revision number. For instance if the major revision letter is an A, then minor revisions are B.1, B.2, B.3, and so forth. When the document is ready and approved, the major revision will be a letter B. [14].

3.5 Data security issues

For the companies the documents have a vital meaning, they may contain information that are critical to the company's existence. Therefore it is very important that the document management system can also take care of the documents' data security in addition to management of the documents. Primarily the data security is understood as a management of authorization and granting permissions to the documents, but it also contains the documents' back up processing, usage logs and removing of the obsolete documents. And when operating on the Internet also the encryption of telecommunication and the virus protection plays a significant role what comes to the data security. [14].

As a default the user does not have any permission to actual document files on a server, instead the system uses authentication to resolve user rights and allows user to view and modify only documents that are allowed to him or her. The permissions can also be used to restrict user's actions to modify documents' setting based on a document type for instance. So different users can have different roles to manage the documents, also users can be separated to groups, which offers more possibilities to grant permissions to the documents. [14].

The transaction logs informs about the usage of the documents. At least the creation of the document and the changes in document's status are recorded. Figure 11 shows an example of log file [14].

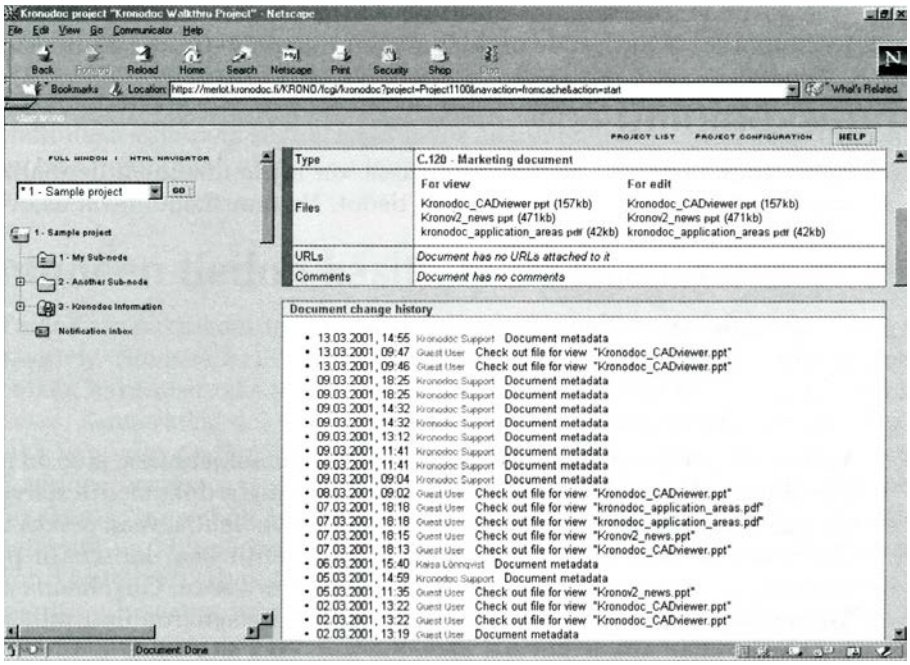

Figure 11. An example of a log file related to a document [14].

In addition following information can be logged into the system:

- authentication information, i.e. who has logged into the system
- change of documents' attributes
- viewing and downloading of documents
- printing of the documents
- change of document status, i.e. reviewed, approved
- document removing

With the documents' traceability information it is easy to check afterwards, who has removed particular document(s) or in the case of confidential information who has printed or downloaded that documentation. [14].

4 Software requirements engineering

4.1 What are requirements?

The software requirements engineering defines how the system must work and how it interacts with the end users and other systems. It may also set constraints to development, what requirements are feasible and what are not. In table 6 there is the definition of The IEEE Standard Glossary of Software Engineering for a requirement. [5]-[15].

Table 6. The IEEE Standard Glossary of Software Engineering for a requirement.

1	A condition or capability to needed by a user to solve a problem or achieve an objective.
2	A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed document.
3	A documented representation of a condition or capability as in (1) or (2).

In the software development process, requirements definition and analysis validates that the software to be developed accomplishes the given criteria, on time and on budget. The requirements are often presented in following categories:

- The business requirements that represents the high-level vision and the scope of the system from the customer. They are often abstract and approximate of the nature. They are described in a project's vision document.

- The user requirements describe how the system or the product to be developed must interact with the end user, when he or she is accomplishing tasks with it. They are described with the use cases that present the user interaction as a written story.

- The functional requirements describe how the software must be developed so that the business and the user requirements are achieved.

- The non-functional requirements contain all the other criteria that the system must fulfil, before it can be seen as a finished product. These can be requirements related to usability, reliability, performance, supportability, or regulations for instance.

The functional and non-functional requirements are described in the Software Requirements Specification document that also, in addition to the behavioural requirements, includes both the possible constraints for the development and the quality related issues. Figure 12 shows the relationship of the software requirements component. [15].

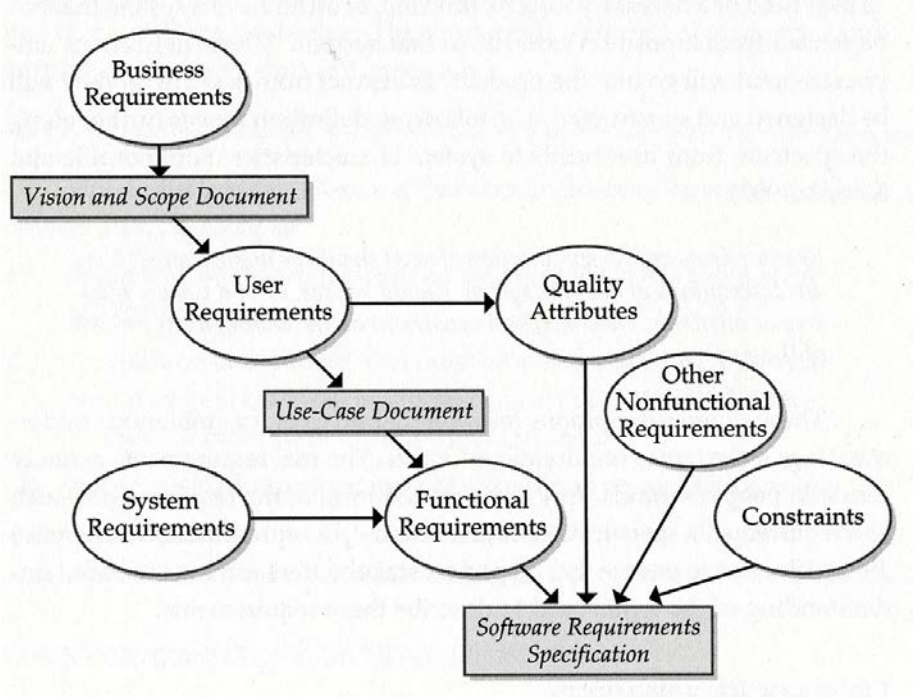

Figure 12. Relationship of software requirement components [15].

Software Requirements Specification document includes both functional and non-functional requirements and verifies that they can be implemented. By analysing the requirements, the development costs will be reduced, because the requirement errors are the most common ones and most costly to fix [16]. The requirements engineering process should continue through the whole application development process to verify that requirements are valid, and refine them if necessary [15].

4.2 Workflow and management

The requirements engineering defines the scope of the system and how it should function, both from end user perspective and as integration to the external systems. It includes the requirements' definition, documentation and maintaining from the start till the end of the development process.

The requirements engineering includes five separate phases to process requirements. [17].

- Requirements elicitation, in which the software developer communicates with the stakeholders and end users to form a common view of system's behaviour. It contains the functional needs from the business representatives and end users. This phase also defines possible development constraints.

- Requirements analysis, where the needs are analysed to become actual software requirements. It is crucial that both customer and software developer understand them same.

- Requirements specification is a document, where gathered requirements are presented formally and each requirement can be identified uniquely. This document can also act as a basis for a contract between customer and software developer.

- Requirements validation / verification ensure that the system requirements meet the software requirements. It evaluates that the requirements are feasible (cost and resource related) and correct (completeness and consistency related).

- Requirements management is a phase where the evolution of the requirements are analysed and verified through the life cycle of the development process. It includes also the change of requirements.

The requirement engineering is a process that discover, document and maintain all the requirements during a software project. It is an on-going process because the requirements are refined often during a project's life cycle. [17].

Requirements elicitation

Requirements elicitation is the process, where the user needs and constraints are identified and understood. The business requirements will serve as a base for the user requirements to be gathered. The user requirements describe the tasks, which end user needs to perform. In other words the requirements describe how the system must be implemented so that end users can succeed with their tasks. Requirements elicitation is like glue between a given problem and the solution. When the customer and the developers have a common understanding of the end user needs, it allows also the participants find alternative solutions to the problem domain. [18].

The elicitation process forwards at parallel with the requirements analysing, specification, and validation processes. In other words, when the requirements analyst communicates with the customer's stakeholders by asking questions and getting information from them, at the same time he or she analyses the given information to become software requirements. By writing down the requirements, with the needed diagrams, the user structures them as a specification, which correctness will then be verified from the customer. And when the new requirements arise, the process will start from begin. [19].

Requirement analysis

In requirements analysis the goal is to verify that all stakeholders understand them same way and also to find possible errors, omission or other deficiencies. By analysing the requirements thoroughly before proceeding further in a process, the on-going project benefits of scrutinized requirements that enables to construct more realistic work estimates for design, implementation and testing. A good practise to analyse requirements is to look at them from a different point of view, for instance in a textual and graphical form. When the stakeholders are involved to project from the beginning, it is easier for them to perceive a realistic view of the finished project. The communication with the stakeholders also helps to prioritise requirements, which requirements are more important than the others, and at the same time it clarifies the problematic issues that causes confusion and are not so easy to understand. The following methods can be used to analyse requirements more specific. [18].

- Drawing a system's context diagram that illustrates the boundaries and connections between the system and outside world.

- Defining a data dictionary, a shared repository that explains all the data elements and their structure in a system.

- Requirements modelling graphically that include data flow-, entity relationship-, state-, and class diagrams.

- Prototypes of the user interfaces that demonstrate how the finished application should behave.

The requirement analysis phase is a continuous interaction with the customer, where lot of different information need to be categorised to achieve a consistent view of the system to be developed. [18].

Requirements specification

The requirements specification is a document that lists all functional and non-functional requirements in some consistent, accessible, and reviewable way. It identifies each requirement uniquely, which enables tracking of a requirement later in a development process. It is a good way to verify that each feature in an application is originated from a requirement, and vice versa that each requirement is implemented in a final product, No loose ends. The requirement specification document should contain following information in some level. [18].

- Source of a requirement, which helps the stakeholders to understand the origin of functional and non-functional requirement, and see the relation between a requirement and a use case description, a customer input, a business rule, a standard, or some other external source.

- Labelling a requirement that makes each requirement unique, so that they are traceable and manageable to record change and status activity.

- Business rules that defines under what circumstances and by whom the application is useable. These rules can be listed under own section in a requirement specification document, or use a separate document for them.

- Traceability matrix that links the requirement to the design, implementation and the testing. It verifies that every feature in a final product has a well-founded requirement and that all requirements are implemented. Traceability matrix should be updated during the whole development, not just at the end.

It is a good practise to start using IEEE 830-1998 software requirement specification template (http://www.math.uaa.alaska.edu/~afkjm/cs401/IEEE830.pdf) and adapt that to existing needs, rather than creating one's own template from a scratch. That template contains a structure for both functional and non-functional requirements. [18].

Requirements verification

Requirements verification phase's purpose is to ensure that requirements, which are documented in a requirements specification, are complete and can be implemented and tested without any vagueness with them. Requirements that do not pass the verification should either be removed or updated to fulfil the quality characteristics for a reliable design and testing. The stakeholders participating to the verification phase is essential for the product's acceptance criteria. The requirements verification phase usually contains following actions. [18].

- Inspection of the requirements specification and other related documents. It is a good practise to form a small team of project participants that examines the documents to check if there are any defects. Such a team can consist of for instance an analyst, a customer representative, a designer, and a tester.

- Write test cases from requirements. By writing test cases that describes the product's expected functionality, from a black-box point of view, helps participants to find any illogical behaviour that can be traced back to the requirements.

- Write a user manual. A good written user manual describes the functionality of a system in an easy and understandable way. The writing of a user manual should be start at an early stage of a development, it same works as an aid for analysis of functional requirements.

- Define acceptance criteria. It is important that the customer describes their metrics to determine, whether the application fulfils their needs and is useable as planned.

The requirements verification phase ensures that all documented requirements are good and valid requirement statements and they are traceable. The requirements verification is an activity that continues through project's life cycle, when requirement changes are evaluated. [18].

Requirements management

The requirements management is the management of changes in a project's life cycle. It contains both perceiving of new requirements and what cost and other effects they do cause to the project. A specific change control board's task is to decide and prioritise what requirements will be implemented in this project, and what are implemented in following projects. Also the tracking of requirements' status in various phases of a project is a part of requirements management. Many organisations use version controlling for code based projects, but it is good practise to apply it to concern the requirement documents also. The requirement management consists of following activities. [18].

- By defining a requirements change control process, it easy to manage requirements changes as project evolves. It keeps track of requirements and what their statuses are.

- Establish a change control board, which contains few project stakeholders that goes through requirement changes in a project and decides which ones to accept and which ones to reject. The change control board also evaluates requirements to see if they are part of the project's scope and if they can be implemented.

- Defining the effects of the requirement change. For every requirement change there should be an evaluation, what are the impacts, such as cost and schedule effects, for the project. That helps the change control board to prioritise requirements and to decide which ones of the changes will be accepted.

- Trace the change effects to the other elements. With the traceability matrix it is easy to identify other elements, like requirements, components, code, or test cases, which have a relation to the changed requirement, and therefore it also may have a change impact.

The commercial requirement management software automates these tasks. With it the user can save different types of requirement into the database, define various attributes to them, track requirements' statuses, and create traceability links between requirements and other software development. [18].

Use case approach

One common method to describe user requirements is a use case scenario. The use case describes interaction between the system and actors. Actors represent the user of the system; they can be end users or external systems. The use case is a written explanation of interactions that the user must perform to get the task done. A use case can also be presented as a diagram that is a graphical form of a written description. Figure 13 illustrates the use case diagram and its elements. [16].

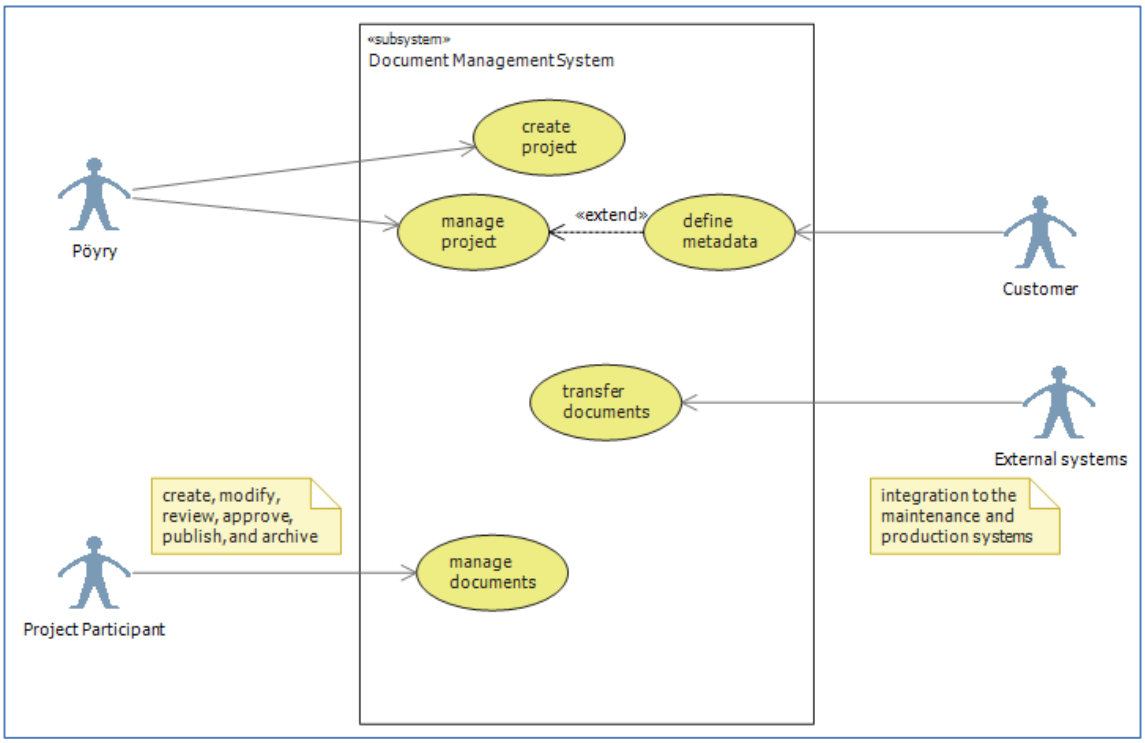

Figure 13. Major elements of the use case diagram.

The use case scenario contains the normal flow of user interactions that proceed in a chronological order explaining all the needed actions to accomplish the task. It may also contain alternative flow for interactions that varies from the normal flow of interactions. The normal flow may in some point branch off to an alternative flow, which will later join back to normal flow and ending to a successful goal. [16].

5 Collaborative document management - Case Pöyry

5.1 Introduction

This document describes and analyzes the requirements for the collaborative document management system (later called just a system in this stydy). It also presents existing cloud based document management services (SaaS) and their features available at the moment (in spring 2013).

At the moment Pöyry uses their own developed product (DocHotel) that is based on Microsoft Windows SharePoint architecture to manage and share project documentation between multiple users and user groups that are involved in global projects.

5.2 Application

Users

The system users are different companies and stakeholders that are involved into the project in various phases of its life cycle. Typically project parties publish their project documentation into a system.

One of the participating companies, such as Pöyry, does the management of projects' documentation. In other words it creates the project's work area into a system and defines the rules how different project parties can access each other's published documentation.

User management

The document management system should allow easy and effective way to manage end users and user groups that are involved in a project. It should also offer a confident method to grant and revoke access permissions to a project documentation that exists in the system.

From a system point of view following features are necessary:

- management of user groups and its users
- user group can include other user groups
- permissions to a documentation access
    - by user groups
    - by users
    - by document revisions

Users can be separated to three kinds of roles:

- readers, who have an access to browse and read permitted documents
- editors, who have an access to browse and modify permitted documents
- managers, who manages documents in a project, e.g. by granting rights to users and user groups

The system should enable group users into user groups by allowing a user group to contain other user groups.

For instance Pöyry has defined two user groups: Pöyry-readers (read access to Pöyry's documentation) and Pöyry-editors (update privileges to Pöyry's documentation). Now let's assume that a project has company X, which needs read privileges to Pöyry's documentation. Project manager just adds company X into Pöyry-readers user group to grant them read access.

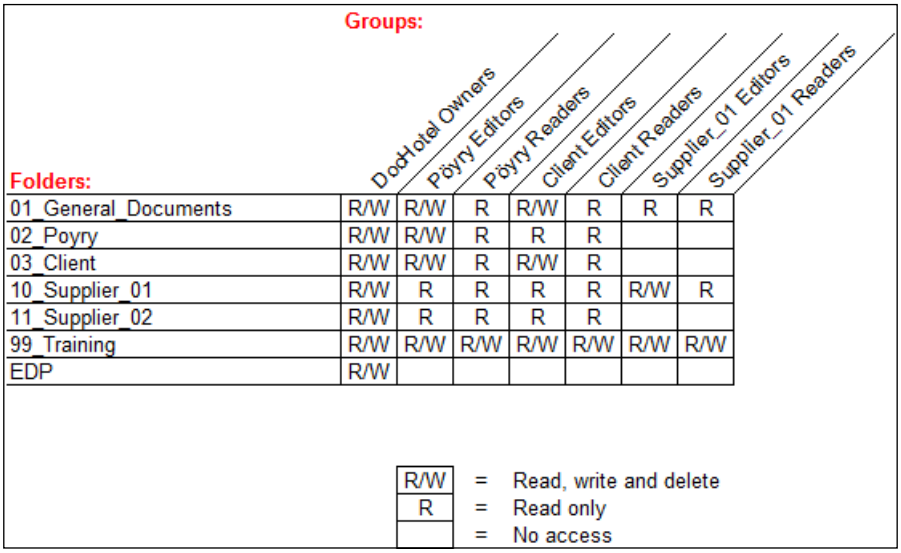

Figure 14. Folder structure and defined permissions of user groups on Pöyry's projects (Pöyry).

Figure 14 shows the permissions' structure used in Pöyry at the moment. It lists the user groups and their permissions in a project's folder structure.

Documentation workflow

In documentation workflow has several user roles that take part of a document's creation process.

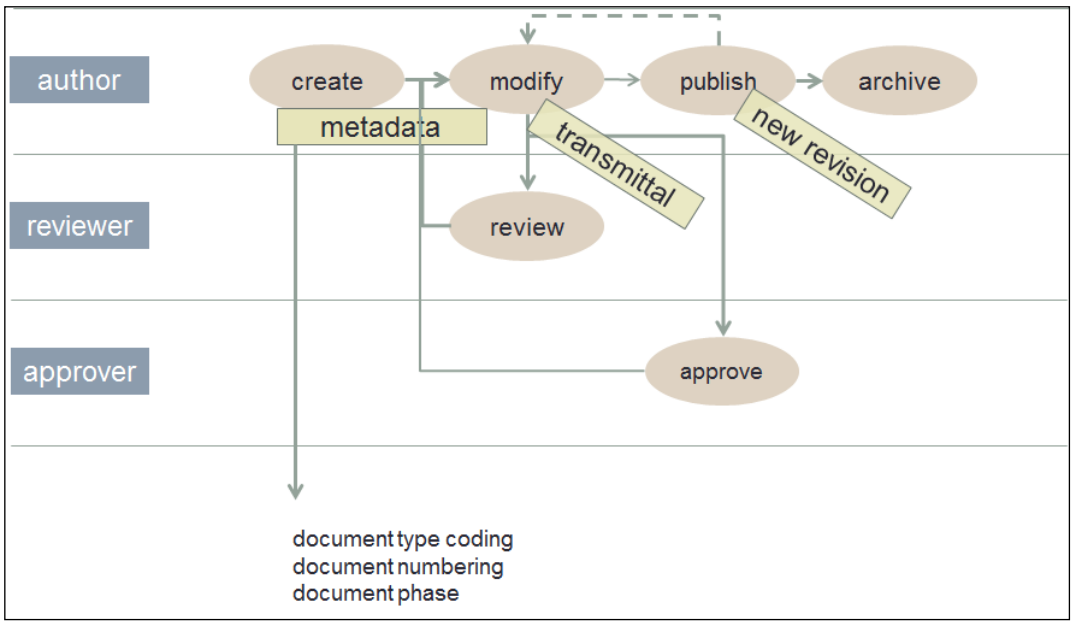

Figure 15. Overview of a documents life cycle and different user roles.

Figure 15 shows a typical life cycle of the document and actions of different user roles.

Project setup overview

When a new project begins, Pöyry defines, in collaboration with a customer, the following issues concerning documentation:

- what kind of metadata are used in a project
- who are the project participants and what are their privileges
- structure of the document management system

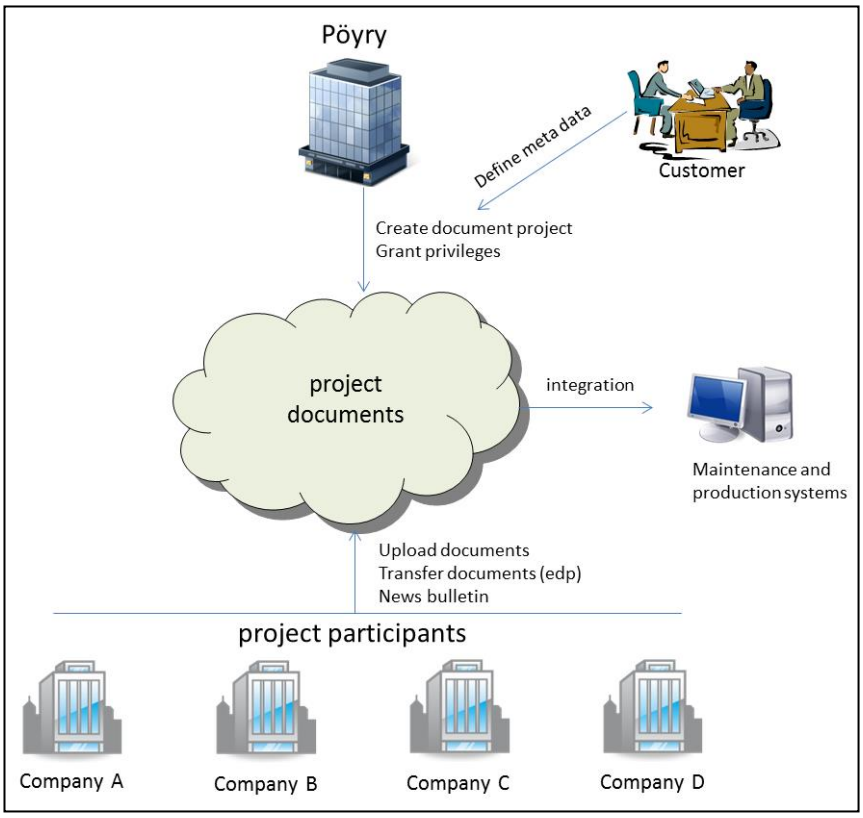

Figure 16. Overview of a documentation project.

Figure 16 shows a documentation project and its participants.

Environment

The document management system should run as a cloud computing service available via Internet that offers authoring tools for data management.

Language and culture

Because Pöyry operates in a global environment and with different cultural aspects, the system must support multi-language and take into account separate ways to display dates, digits and other relevant, culture specific information including metadata if possible.

In most projects the documents exists in two languages: native and English. For instance if the project operates in China, the project documentation is both in Chinese and English. Therefore the system's user interface should also be both in Chinese and English.

Architecture

The system preferably runs software as a service (SaaS), so that hosting and daily based IT-management is operated by a service provider.

The architecture should enable integration to other plants' externals systems such as maintenance and production systems (plants) for reliable transfer of documentation when the project is finished.

Operation of the application

The system should operate on 24/7 basis and it also should offer global end user support.

5.3 Functions

Basic functions

Captive systems

The system should store the documents into a data warehouse instead of a file server. This means that the users use documents' to check in and check out functionality, when they are modified.

Configurable metadata and structure

The system should support functionality that enables end user to add, modify, and remove metadata used for documents' definition and how they are organised. The system should also support the use of templates in customer specific projects, as they do not necessarily use same metadata in their documentation.

Check in / check out

The system enables the check in / checks out functionality that prevents users to overwrite other users' modifications to the documents.

Version handling

The system is capable to keep track of different versions of the documents, and enables users to browse documents' version history.

Revision handling

The system offers tools for user to define the format of documents' revision marks.

Number banks (issue / client)

The system should support to define number banks that are used to identify the technical drawings in a project. The number bank offers the next available sequence, which

is customer and project based, to be attached to the drawing. The drawing may contain multiple sequences per drawing, for instance one for the customer and one for internal use.

Search functions

The system offers versatile search methods for user to find documents from the system, based on documents' metadata and structure.

Document preview / thumbnail

The system enables user to preview documents without downloading them. A document preview is not modifiable. When user moves the cursor over a thumbnail, an informal window opens showing relevant information of a document in question (author, revision, date etc.).

Data encryption

The system should enable the encryption and decryption of the document.

Handling of documents including several files

The system enables the user to handle several files included in the document as one document, e.g. minutes of meeting including memo and appendixes.

E-mail handling

The system offers a feature for users to send e-mail from their personal e-mail system or project's e-mail box to the document management system and these e-mails will be saved automatically as a document into the document management system.

Printing services

The system should be capable to print both an individual document and number of documents (mass printing) to a printer the user has defined.

Folder structure template

The system should offer tools to define versatile and project specific folder structure templates to be used in the project.

URL for document

The system should enable to define an URL for targeted users / user groups, i.e. they can access the document behind a link. The system sends the link via e-mail.

Commenting of the document

The system enables the user to comment documents via document preview. The commenting does not change the content of the document, instead it is separate information attached to the document. The document's author is notified, when someone has commented the document.

Document mass transfer (EDP)

The system should enable functionality for users to gather multiple documents from a system and send them to the selected project participants who may or may not have an access to the system.

The user can mass collect documentation that the user has permission to. If gathered documents are sent as a link list to a receiver, the system allows downloading of granted documents only. The receiver must have permission to the document.

But if documents are sent as a zip-package, then the user who has gathered the documentation is also responsible for the validity of a mass sending of documents.

Transmittal

The system should enable functionality to add dispatch note as part of document mass transfer that includes necessary metadata. The user should easily define what metadata are included into a transmittal.

Mass upload / download tool

The system should enable a project participant to upload several documents into the system at a time. The sender defines what documents and what metadata are included to be uploaded into a system. The system then updates (or add) sent documents and their metadata.

The system should also enable a project participant to download several documents from a system at a time. The user defines what documents and what metadata are included to be downloaded from the system.

Project news

With project news functionality project participants can easily publish announcements and similar kind of messages that are shown to all other project members in a project's workspace.

Different views to the document

The system offers possibility to configure different views to document, e.g.

- Recently modified
- I am editing
- Documents by Doc Type / Area / Supplier etc.

Marking of favourites / tags

Users are enabling to mark documents as Favourite or define Tags for documents. These can be used as search criteria in search function.

5.4 Business requirements

Common requirements

The collaborative document management system will be a commercial product that operates Software as a Service in a cloud environment. It is available via Internet and the end user pays only for the usage.

Table 7. Pöyry's project categories

Discipline	Users	disk quota	project's life cycle
small	20 or less	30 GB or less	less than 6 months
medium	21 - 100	31 - 100 GB	6-12 months
large	101 or more	101 - 150 GB	over 12 months

Pöyry has categorized projects in three types based on amount of users, disc quota and project's life cycle.

Supports work sharing

The system streamlines the documentation process in a various projects and enables the sharing of work and documents between project participants.

Scalability

The system should be scalable for the needs of the various sizes projects, which can be either small, medium, or large ones.

Cloud based system

The collaborative document management system runs on the Internet as a cloud service and it is used with a web browser. Pöyry's interest is to use a commercial product that fulfils the requirements in the best possible way.

Configurable

The document management system should be configurable for projects of various types and sizes.

Cost effective

As the system operates as a cloud service, it reduces the cost of the organisation's IT-infrastructure (servers, data networks, replication, and maintenance) as the service provider takes care of them and the company pays only for the usage of the service.

The setup and usage

The setup of the collaborative document management system should be easy, and its functionality should support the learning process so that the end users can use it in the most efficient way.

The external systems

The system should be capable of integrating with other external systems (e.g. CAD, plant's data management, process accounting, and maintenance systems).

Data security

The system should enable the reliable way to maintain documentation so that only those participants, who are granted to handle documents, can do so. Also the service provider must have a process how to act with denial-of-service attacks and other threats.

Costs

Operating costs

Typically the operating costs of the cloud service are based on

- needed disk quota
- amount of data transfer per month
- amount of users
- needed computing capacity

As an example, the Amazon S3 has following pricing in the EU region. (http://aws.amazon.com/s3/pricing/)

Table 8. Storage pricing

size	pricing
1 TB / month	$0.095 per GB
49 TB / month	$0.080 per GB
450 TB / month	$0.070 per GB
500 TB / month	$0.065 per GB
4000 TB / month	$0.060 per GB
>5000 TB / month	$0.055 per GB

Table 9. Request pricing

request	pricing
PUT, COPY, POST, or LIST	$0.01 per 1000 requests
Glacier archive and restore requests	$0.55 per 1000 requests
Delete requests	Free *
GET and all other requests	$0.01 per 10000 requests
Glacier data restores	Free **
* No charge for delete requests of standard or RRS objects. For objects that are achieved to Glacier, there is a pro-rated charge of $0.033 per gigabyte for objects deleted prior to 90 days	
** Glacier is designed with the expectation that restores are infrequent and unusual, and data will be stored for extended periods of time. You can restore up to 5% of your average monthly Glacier storage (pro-rated daily) for free each month. If you choose to re-	

store more than this amount of data in a month, you are charged a restore fee starting at $0.011 per gigabyte.

Table 10. Data transfer (out from S3 to Internet) pricing

size	pricing
1 TB / month	$0.000 per GB
10 TB / month	$0.120 per GB
40 TB / month	$0.090 per GB
100 TB / month	$0.070 per GB
350 TB / month	$0.050 per GB
>350 TB / month	negotiable

Formula for costs

The following example uses Amazon S3 prices to estimate documenting costs in a cloud environment.

- Cost data
- Project duration: 24 months
- Users in a project: 100
- Monthly disk quota: 135 GB

Table 11. An example of cost calculation.

cost	estimate	price	per month	total
Storage		$0,095	$12,825	$307,80
Requests / day (1 user)				
PUT, COPY, POST, LIST	10	$0,01	$0,21*	$5,04
GET	20	$0,01	$0,042	$1,01
Data transfer (GB per month)	1500	$0,12	$180	$4320
			Project total cost	**$4633,85**
* Average 21 work-days per month				

Formula for storage

$$Storage = disc\ quota\ per\ month\ \times\ price\ \times project\ duration$$

Formula for requests / day (for 1 user)

$$Requests_{PUT} = \frac{(users \times estimate \times price \times 0.21)}{1000} \times project\ duration$$

$$Requests_{GET} = \frac{(users \times estimate \times price \times 0.21)}{10000} \times project\ duration$$

Formula for data transfer

$$Data\ transfer = estimate\ \times\ price\ \times project\ duration$$

Expectations for the system

Though both the customer and Pöyry will be using the collaborative documentation management system, they have different expectations for it.

Table 12. The expectations for the system

Expectation	Customer	Pöyry
cost effective	Decision point to choose Pöyry (Pöyry can offer solution to document management in the project)	important, reduces overall costs
easy to learn and use	important, personnel uses	important, Pöyry setup and configures the project
integration to external systems	important, supports business process	important, business value for Pöyry
configurable	Client's standards can be implemented into the system	important, suites for versatile customers and projects
data protection	critical, matter of trust	critical, negative publicity
life cycle	plant operates 30 years	size of a project defines (2 years)
additional value to customer	Handling of docu-	Business value for Pöyry.

	ments during the project, the continued use of the maintenance phase.	Pöyry can offer additional services for the client.

5.5 Document management systems as a cloud service

This chapter describes existing cloud based services for collaborative document management and their features.

ProjectPlace

ProjectPlace is a Swedish company that offers cloud services for different areas including document management.

In addition to document management ProjectPlace also offers tools for project planning, resource management, online meetings, meeting managements, project templates, and issue management. It runs also on iPhones, iPads and Android apps.

Document management system allows to:

- Lock files while you work and prevent others from making changes at the same time.
- Keep track of the changes made, both by you and by others, using version management and document history.
- Send document reviews to your partners and/or employees, and collect all comments and changes in one place.
- Set access rights on folders that contain sensitive information, to be sure that no unauthorized persons have access.
- Share documents in one central place.

ProjectPlace has a Support Centre that operates 24/7 and they provide an e-mail response within an hour. They also offer both online and classroom based training.

ProjectPlace has over 700000 users in 150 countries.

Table 13. ProjectPlace technical recommendations.

	Operating system	Web browser	E-mail program
PC	windows XP or later	IE7 or later, Firefox	Outlook 2000 or later
Mac	Mac OS X	Firefox, Safari	Mail 1.3.4. or later

ProjectPlace web site: https://www.projectplace.com/.

Autodesk Buzzsaw

Fully integrated into the Autodesk portfolio, Buzzsaw SaaS includes document model and data management tools to support Building Information Modeling (BIM) workflows with access from desktop, mobile device, or the web.

Buzzsaw Features:

- View and exchange project documents, designs and models from any modern web browser.
- View thumbnails, images 2D and 3D DWF files and edit DWG files.
- Buzzsaw Sync automatically synchronizes from desktop to Buzzsaw.
- Site- and project-level user and group capabilities along with 8 permission levels to access project information.
- Activity log and audit trail. Know who accessed, edited, uploaded, or downloaded document and when.

- Version control. Track and manage multiple document versions, add comments and send notifications when a new version is created.

Table 14. System requirements.

Desktop PC	windows XP or later, 2 GHz processor or higher, 35 MB free disk space, VGA 1024x768 or higher
Buzzsaw web access	IE7 or higher, Firefox 4 or higher, Apple Safari 5 or higher, Google Chrome 10 or higher
Buzzsaw mobile app	Apple: IOS 4.2 or higher, Android: OS 2.2.1 or higher

Autodesk Customer Support is available Monday to Friday in business hours except Autodesk holidays.

Autodesk Buzzsaw web site: http://usa.autodesk.com/buzzsaw/.

LogicalDOC Cloud

LogicalDOC Cloud is a document management software solution from a USA based company Logical Object. It was founded in 2006 and it addresses the needs of the enterprise document management on the market across broad range of different sectors and vertical markets.

LogicalDOC Cloud contains a web user interface for sharing, setting security roles, auditing and finding enterprise records and registers. The flow of the documents through the business cycle can be managed and tracked from their initial creation through sharing and collaboration, across approval, and though reviews and revisions.

LogicalDOC Cloud supports the following web browsers:

- Firefox 10+

- Internet Explorer 8+
- Google Chrome 8+
- Safari 5+

LogicalDOC web site: http://www.logicaldoc.com/document-management-system.html.

Alfresco

Alfresco is an open source document management system that uses Amazon Web Services (AWS) infrastructure within the USA. It leverages the Amazon Simple Storage Service (S3) technology to provide reliable and secure storage for documentation according to EU standards. It offers 256-bit advanced encryption standard (AES) by using a paid network.

Alfresco has over seven million business users in 75 countries that manage four billion files and processes daily.

Alfresco web site: http://www.alfresco.com.

The basic features of Alfresco are:

- Document management including version control, on-line previews, configurable workflows, integration to MS Office and Google Docs, definable property groups and types, and comprehensive search functions based on Apache Soil platform
- Business process automation including rules driven task assignments, user dashboards, folder rules definitions, customization options for business processes, and business process integration with ERP and CRM applications
- Record management including simple records filing, flexible file plans to model the system, records management for scheduled tasks, content based search, and freeze feature to lock records and folders to prevent modifications

- Team collaboration including project collaboration and team sites, support for Wiki, blogs, lists, FAQ's and links, commenting, Facebook's Like and Follow features, notifications and activity feeds, synchronization features

- Web publishing including publishing to social channels (Facebook, Twitter, LinkedIn, YouTube etc.), integration with existing applications, and content deploying from Alfresco to content server

- Secure file sharing including hybrid ECM that runs both on-premise and in the cloud, public file share, full audit control, cloud security, customizable access levels, access from PC, tablet and smartphone, and viral user adoption

- Mobile content management that supports the use of mobile devices

Alfresco offers following support packages:

- Standard support with telephone access and four business hour response time

- Enterprise support on 24/7 basis, severity 1 issues are responded within two hours and severity 2-3 issues are responded within two business hours

- Premiere support based on Enterprise support that has two separate tiers: Premier and Premier Advantage

The comparison matrix of support packages can be found here: http://www.alfresco.com/services/subscription/technical-support#network-services.

The pricing and subscription information of different options can be found here: http://www.alfresco.com/products/compare.

OMNISTAR

OMNISTAR document management software is a web based file manager and storage utility that offers employees and customers easy access to files in a secure online repository. It is used over 14000 companies over 80 countries.

OMNISTAR web site: http://www.omnistardrive.com.

OMNISTAR's basic features are:

- User account management that gives users to specific access and permission levels to any files and folders
- Power features like multiple file download, drag and drop files, file check-in and check-out, document search, large file upload tool, file recycle bin, secure server support, customizable interface, document upload email notification, and optional file approval
- Data security including secure data center, secure network, SSL security
- Supports IE, Firefox, Safari and Google Chrome
- Statistics and reporting including system usage overview, upload/download statistics, pending documents, pending document check-in list, and file access dates
- 24/7 service and support including helpdesk ticket, knowledge base, telephone, and support videos

Pricing information can be found here: http://www.omnistardrive.com/pricing.php.

5.6 Comparison and risk analysis

The following table compares above mentioned cloud based solutions and their features.

Table 15. Comparison table of the cloud services.

Features	ProjectPlace	Auto-desk Buzzsaw	Logi-calDOC Cloud	Alfres-co	OM-NISTAR
Configu-rability	project templates	workflow tem-plates	metada-ta, tem-plates	configu-rable workflow	no
Number banks	no	no	yes	no	no
Check In / Check Out	yes	yes	yes	yes	yes
Version handling	yes	yes	yes	yes	yes
Revision handling	no	no	yes	yes	no
Multiple files in one doc-ument	no	no	no	no	no
User rights manage-ment	yes	yes	yes	yes	yes
Data pro-tection	yes	yes	yes	yes	yes
Com-menting	yes	yes	yes	yes	no
Integra-tion to other systems	MS Office	Autodesk	Web-DAV, CMIS, MS Of-fice, Amazon S3, LDAP,	Amazon S3, Drupal, Ephe-soft, Google Docs, jive,	

			MS Active Directory	Liferay, MS Office, Mulesoft, QuickOffice, SalesForce	
Pricing model and prices	Enterprice package: based on number of users Team edition: for 3-50 users, monthly fee, starting at $29.95 per user per month (source: http://zincubate.com/managing-projects-online-with-projectplace)	Standalone licence; for individual and small offices Network licence: for larger organizations to access Buzzsaw from multiple locations	Contact supplier for pricing information	Enterprice Network: 500 users, unlimited storage, 2GB file size limit	Basic hosted: $17.00 / month, 1GB storage, 1GB bandwidth, unlimited users Premium hosted: $37.00 / month, 3 GB storage, 3GB bandwidth, unlimited users

Risk analysis

If Pöyry decides to choose one of the above mentioned document management systems for further inspection, it will be a good practice at the time to ask for technical auditioning from a service provider, especially in issues that concern data security. For instance how the service provider has prepared against possible denial-of-service kinds attacks.

6 Summary

This study lists Pöyry's requirements for their next document management system that will replace the existing DocHotel that is a Microsoft SharePoint based application. The new collaborative document management system will operate in a cloud environment as Software as a Service. Pöyry operates as a service provider for its customers setting up the projects' workspaces (folder structure, document templates, and metadata) and defining permissions for project participants.

This study also presents a few worthy candidates and their features that are available at the moment, in spring 2013, offering document management in a cloud environment.

References

1 Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. 2009. Above the Clouds: A Berkeley View of Cloud Computing. URL: http://x-integrate.de/x-in-cms.nsf/id/DE_Von_Regenmachern_und_Wolkenbruechen_-_Impact_2009_Nachlese/$file/abovetheclouds.pdf. Accessed 31.1.2013.

2 Lautanala, Janne. 1997. NC Verkkotietokoneet. Espoo: Suomen ATK-kustannus Oy. Read 31.1.2013.

3 Yao, Chao-Hsu. 2006. Grid Computation: The Fastest Supercomputer in the World. URL: http://www.csa.com/discoveryguides/grid/reviewf.php. Accessed 27.3.2013.

4 Heino, Petteri. 2010. Pilvipalvelut. Helsinki: Talentum. Read 18.1.2013.

5 Kukkonen, Juha. 2011. Tutkimussuunnitelma - Pilvipalvelut. URL: myy.haaga-helia.fi/~a0903114/Pilvipalvelut.docx. Accessed 20.2.2013.

6 Salmio, Petri. 2012. Pilvipalvelut. Thesis. URL: https://publications.theseus.fi/handle/10024/41634. Accessed 8.1.2013.

7 Ward, Susan. 2011. 5 Disadvantages of Cloud Computing. URL: http://sbinfocanada.about.com/od/itmanagement/a/Cloud-Computing-Disadvantages.htm. Accessed 26.3.2013.

8 Tantow, Martin. 2011. What to consider before choosing an IaaS provider. URL: http://cloudtimes.org/2011/02/27/what-to-consider-before-choosing-an-iaas-provider/. Accessed 24.1.2013.

9 Konyoumjian, Victor. January 2010. The New Age of Cloud Computing and GIS. URL: http://www.esri.com/news/arcwatch/0110/feature.html. Accessed 10.1.2013.

10 Color & Cloud. 2012. Platform as a Service overview. URL: http://www.coloandcloud.com/editorial/platform-as-a-service-overview-paas/. Accessed 24.1.2013.

11 Department of Electronics and Information Technology, Ministry of Communications & Information Technology, Government of India. 2010. Glossary-s. URL: http://deity.gov.in/content/glossary-s. Accessed 24.1.2013.

12 MacVittie, Lori. 2009. Intercloud: The evolution of Global Application Delivery. URL: https://devcentral.f5.com/blogs/us/intercloud-the-evolution-of-global-application-delivery. Accessed 25.1.2013.

13 Nixu. 2010. Pilvipalvelujen tietoturvasta yritysjohdolle. URL: http://www.nixu.fi/julkaisut/2010/pilvipalvelujen-tietoturvallisuudesta/. Accessed 20.2.2013.

14 Anttila, Juha. 2001, Dokumenttien hallinta. Helsinki: Oy Edita Ab. Read 6.2.2013.

15 The Institute of Electrical and Electronics Engineers. 1990. The IEEE Standard Glossary of Software Engineering Terminology. URL: http://www.idi.ntnu.no/grupper/su/publ/ese/ieee-se-glossary-610.12-1990.pdf. Accessed 22.1.2013.

16 Leffingwell, Dean, Widrig, Don. 2000. Managing Software Requirements: a unified approach. Addison Wesley Longman Inc. Read 15.1.2013.

17 Hokkanen, Mia. 2001. Requirements Traceability. Thesis. URL: http://www.doria.fi/bitstream/handle/10024/35358/nbnfi-fe20011576.pdf?sequence=1. Accessed 15.1.2013.

18 Wiegers, karl E. 1999. Software Requirements. Microsoft. Read 5.3.2013.

19 Masters, Morgan. 2010. An Overview of Requirements Elicitation. URL: http://www.modernanalyst.com/Resources/Articles/tabid/115/articleType/ArticleView/articleId/1427/An-Overview-of-Requirements-Elicitation.aspx. Accessed 27.3.2013.