TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

JUHA BOHM

DOKUMENTTIEN METADATA JA SEN HYVÄKSIKÄYTTÖ VAKUUTUSLIIKETOIMINNASSA

Diplomityö

Tarkastaja: professori Hannu Jaakkola
Tarkastaja ja aihe hyväksytty
Talouden- ja rakentamisen tiedekunta-neuvoston kokouksessa 7. lokakuuta 2015

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Tietotekniikan koulutusohjelma
BOHM, JUHA: Dokumenttien metadata ja sen hyväksikäyttö vakuutusliiketoiminnassa
Diplomityö, 78 sivua
Lokakuu 2015
Pääaine: Ohjelmistotuotanto
Tarkastaja: professori Hannu Jaakkola

Avainsanat: dokumentti, metadata, attribuutti, raportointi, vakuutusliiketoiminta

Tämän diplomityön tarkoituksena oli tutkia dokumenttien metadatan hyödyntämismahdollisuuksia. Tarkastelussa käytettiin hyväksi kirjoittajan työelämässä kertynyttä kokemusta vakuutusliiketoiminnasta sekä metadatan merkityksestä erilaisissa yhteyksissä keskittyen liiketoimintaan ja asiakkuuteen liittyviin dokumentteihin. Työssä keskeistä oli tutkia erilaisista metadata-attribuuteista ja niiden kombinaatioista saatavaa tietoa sekä säännöllistä ja toistuvaa raportointia. Lisäksi tarkasteltiin näin saadun tiedon käyttöä joko itsenäisesti tai yhdistettynä muualta saatavaan tietoon. Tutkimuksen tuloksena oli, että dokumenttien metadatasta saadaan monipuolista säännöllistä raportointia. Tätä raportointia voidaan sellaisenaan tai yhdistettynä muualta saatavaan tietoon käyttää vakuutusliiketoiminnan kehittämiseen sekä kustannusten seurantaan ja hallintaan. Tietotekniikan operatiivisia kustannuksia voidaan seurata ja vaikuttaa niiden suuruuteen. Dokumenttien metadatasta saadaan tietoa, jota vakuutusliiketoiminta voi käyttää suunnitellessaan uusia tuotteita ja asiakaspalveluita sekä paljastamaan mahdolliset väärinkäytökset. Tämän tutkimuksen perusteella saatu lopputulos on, että hyvälaatuista dokumenttien metadataa ja siitä saatavia raportteja voidaan käyttää monipuolisesti yhtenä tietolähteenä ohjaamaan liiketoiminnan suunnittelua ja hallintaa sekä hallitsemaan tietotekniikan kustannuksia.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Information Technology
BOHM, JUHA: Document metadata and its utilization in insurance business
Master of Science Thesis, 78 pages
October 2015
Major: Software Engineering
Examiner: Professor Hannu Jaakkola

Keywords: document, metadata, attribute, reporting, insurance business

The purpose of this master's thesis was to study possibilities to utilize document metadata. Writer's experience from insurance business was used in the study and metadata's importance in different business cases. The study focus is on business and customer documents but also company's internal documents are slightly looked into. The main focus in the study was to investigate different kinds of metadata attributes and information and reporting which can retrieved by combining attributes together. The result of the study was that versatile and regular reporting can be obtained by using document metadata. This reporting then can be used as such or with data coming from other sources to develop and manage insurance business, follow up business costs and business planning. Obtained reporting can be used to follow up and manage IT operation costs and to plan how to influence the future IT infrastructure costs. Information retrieved from document metadata insurance business can use to develop new products and services for their customers and to reveal frauds and abuses. The end result based of this study is that high quality document metadata and reporting using this metadata can be used in many ways as an information source to control and steer the business and IT.

ALKUSANAT

Tämän työn kirjoittaminen on ollut mielenkiintoinen oppimisprosessi, jossa olen voinut käyttää työelämässä oppimiani taitoja ja tietoja uuden näkökulman kautta. Olen joutunut pohtimaan työelämässä opittua ulkopuolisin silmin eikä sisältä käsin omien työtehtävien hoidon kannalta. Haluankin kiittää Hannu Jaakolaa hyvästä ohjauksesta ja asiantuntevista kommenteista työn loppuun saattamiseksi. Kiitosta ansaitsee myös Kari Laine, joka on tarkastanut työni entisenä kollegana sekä antanut kannustusta.

Työssä tutkittava aihe on monille jollakin tavalla tuttu, sillä se käsittelee dokumentteja ja niihin liittyvää metadataa. Vaikka aihe onkin tuttu, sitä ei ole tutkittu tai hyödynnetty täysin läpikotaisin. Erityisesti dokumenttien metadatan hyödyntämismahdollisuudet ohjaustietona tai pohjatietona suunnittelussa ja hallinnoinnissa on alue, jossa voidaan vielä tehdä paljon selvitystä ja innovaatioita. Työssä perehdytän lukijan metadataan, sen keräämiseen ja luotiin, metadatan laatuun vaikuttaviin tekijöihin ja erilaisien tiedostoformaattien metadata ominaisuksiin. Samalla pyrin perustelemaan miksi dokumenttien metadataa tulee kerätä ja miten varmistetaan kerätyn metadatan laatu. Työssäni annan lukijalle oikeita esimerkkejä siitä, miten metadataa on hyödynnetty ja voidaan hyödyntää sekä tietotekniikassa että liiketoiminnassa. Esimerkit ovat vakuutusliiketoiminnasta ja vakuutusyhtiöiden tietotekniikkaympäristöistä. Näissä esimerkeissä hyödynnän tietoja ja taitoja, joita olen kerryttänyt työskenneltyäni 25 vuotta erään suuren vakuutusyhtiön palveluksessa monissa erilaisissa tehtävissä. Tavoitteena on herättää lukija ajattelemaan dokumenttien metadataa ja miksei myös vakuutusliiketoimintaa uudesta näkökulmasta. Tämän tavoitteen lisäksi oma tavoitteeni oli tarkastella työni aihetta yleisellä tasolla hyödyntäen käytössä olevia menetelmiä ja käytäntöjä.

Dokumenttien muodot muuttuvat ja määrät kasvavat edelleen, mistä johtuen niihin liittyvän metadatan tulee muuttua mukana. Tämän muutoksen myötä avautuu uusia mahdollisuuksia metadatan hyödyntämiseen ja automaatioon. Tästä syytä toivon diplomityöni innostavan myös muita ideoimaan ja jatkamaan aiheen tutkimista.

Naantalissa, 17.10.2015

Juha Bohm

SISÄLLYSLUETTELO

1. JOHDANTO 1
2. DOKUMENTTIEN METADATA JA TIEDOSTOFORMAATIT 3
2.1 Mitä on metadata? 3
2.1.1 Määritelmä 3
2.1.2 Miten metadata kerätään ja luodaan? 4
2.1.3 Metadatan standardointi 5
2.1.4 Metadatan yhtenäistäminen 6
2.2 Tiedostoformaatit ja tiedostomuunnokset 7
2.2.1 Massatulostusformaatit 7
2.2.2 Arkistointiformaatit 8
2.2.3 Avoin tiedostomuoto 9
2.2.4 Käytännön de facto formaatit 10
2.2.5 Tiedostomuunnokset 10
3. METADATAN HYÖDYNTÄMINEN DOKUMENTTIEN ELINKAARRESSA 12
3.1 Vakuutustoiminta 12
3.2 Metadatan hyödyntäminen dokumenttien elinkaaren hal-linnassa 14
3.3 Aktiiviset ja liiketoimintakriittiset dokumentit 19
3.4 Passiiviset ja liiketoimintakriittiset dokumentit 20
3.5 Dokumenttien lyhytaikainen arkistointi 21
3.6 Dokumenttien pitkäaikainen arkistointi 24
3.7 Dokumenttien pysyvä tuhoaminen 26
4. METADATAN HYÖDYNTÄMINEN LIIKETOIMINNASSA 30
4.1 Metadatan hyödyntäminen vakuutusliiketoiminnassa 30
4.2 Raportointi 31
4.3 Tiedolla ohjaaminen ja johtaminen 36
4.4 Ulkoinen ja sisäinen asiakas ja asiakkuus 37
4.5 Uudet palvelut ja tuotteet 38
4.6 Liiketoimintaprosessit 41
4.7 Petosten havaitseminen ja ennalta ehkäisy 43
4.8 Tietotekniikan operatiiviset kustannukset 46
5. METADATAN HYÖDYNTÄMINEN YRITYKSEN SISÄISESSÄ PÄÄTÖKSENTEOSSA JA SUUNNITTELUSSA 54
5.1 Liiketoimintastrategian suunnittelu ja seuranta 54
5.2 Operatiivisten kustannusten suunnittelu ja seuranta 55
5.3 IT-strategian suunnittelu ja seuranta 58
5.4 Projektisalkun suunnittelu ja seuranta 58
5.5 Metadatan hyödyntäminen osana tietoturvaa 59
5.5.1 Käyttäjien tunnistus ja oikeudet 59
5.5.2 Tietoturvan vaatimukset dokumenttien metadatalle 60

6. YHTEENVETO 63
6.1 Yhteenveto työstä ja työn tuloksista 63
6.2 Tulevaisuus 64
6.3 Jatkotutkimuskohteet 64
LÄHTEET 66

KUVALUETTELO

Kuva 1. *Kuvassa esitetään informaation elinkaaren vaiheet sekä miten informaatio liikkuu vaiheesta toiseen. Tätä elinkaarimallia voidaan soveltaa suoraan dokumenttienelinkaareen.(Baldeón 2010) 17*

Kuva 2. *Kuvas esittää talletettujen tiedostojen käyttöä niiden luonnin jälkeen..... 26*

Kuva 3. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön EMC Documemtum infrastruktuurissa olevan dokumenttivaraston dokumenttien kokonaismäärän kehitys..... 33*

Kuva 4. *Pohjoismaissa toimivan vakuutusyhtiön kuukausittaiset massatulostusmäärät kolmen vuoden ajalta..... 35*

Kuva 5. *Kuvassa esitetään tavoitetila, jossa asiakas saa saman informaation ja palvelukokemuksen riippumatta palvelukanavasta tai asiakkaan päätelaitteesta.(Laine 2012) 42*

Kuva 6. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön EMC Documemtum infrastruktuurissa olevan dokumenttivaraston dokumenttien kokonaismäärän kehityksestä liiketoimintayksiköittäin. 49*

Kuva 7. *Pohjoismaissa toimivan vakuutusyhtiön kuukausittaiset massatulostusmäärät ja postitukseen käytettyjen kirjekuorien määrät kolmen vuoden ajalta maittain..... 51*

Kuva 8. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön EMC Documemtum infrastruktuurissa olevan dokumenttivaraston levyn käytöstä ja kasvusta liiketoimintayksiköittäin kolmelta vuodelta..... 53*

Kuva 9. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön talletustilan käytöstä ja siihen liittyvistä kustannuksista..... 57*

LYHENTEET JA MERKINNÄT

AFP — engl. anvanced function printing, massatulostukseen IBM:n kehittämä tiedostoformaatti, jolla päästään suuriin tulostusnopeuksiin ja hyvään tulostuksen hallintaan.

Arkistonmuodostussuunnitelma
Arkistonmuodostussuunnitelman perustehtävä on ohjata organisaation arkistonmuodostusta, joka koostuu asiakirjatietojen suunnitelmallisesta arkistoinnista, rekisteröinnistä ja seulonnasta. Sen avulla voidaan lisäksi tehostaa organisaation tietopalvelua.

Arkistointisuunnitelma
Arkistointisuunnitelma on korvannut yleisterminä arkistointimuodostussuunnitelman, jonka alun perin tuli julkishallinnosta.

Asiakirja — synonyymi dokumentille.

Auto-tiering — datan automaattinen siirto eri tallennuskerrosten välillä.

Automated tiered storage
Tietovarasto, jossa data siirretään sääntöpohjaisella automatiikalla varastoon kuuluvien eri tallennusvälineiden ja kerrosten välillä

Massatieto — engl. big data, tieto, jota kertyy suuria määriä jatkuvasti suurella nopeudella ja joka on jäsentymätöntä (Gartner, Big Data 2015)

CRM — engl. customer relationship management, asiakkuudenhallinta.

Codecs — Koodekit, algoritmi tai tietokoneohjelma, joka pakkaa, purkaa ja muuntaa kuva- ja äänisignaalia.

EMC — 1979 perustettu amerikkalainen yritys, joka on erikoistunut tietovarastoihin sekä sisällön-, dokumenttien- ja datan hallintaan.

EXIF — engl. exchangeable image file format, standardi valokuvien ja kuvatiedoistojen metadatalle.

Dokumentti — Tietokokonaisuus, joka on tallennettu ja tarkoitettu ihmisen ymmärrettäväksi sekä aistittavaksi. Dokumentin rakenne on hierarkkinen koostuen osista, jotka voivat olla tekstiä, kuvaa, ääntä tai videota.

Digitaalinen dokumentti
Dokumentti, joka tuotetaan, tallennetaan ja esitetään tietotekniikan avulla. Dokumentit luodaan tuottamisohjelma kirjoittamalla ja eri lähteistä koostamalla. Digitalisen dokumentin tulostusohjelmalla dokumentin sisäinen esitysmuoto muunnetaan ihmisen ymmärtävään muotoon. Usein yksi ohjelma sisältää molemmat ohjelmat sekä dokumenttien talletuksen. Tässä työssä dokumentilla tarkoitetaan digitaalista dokumenttia.

Digitalisaatio — engl. digitalization, digitaalisen teknologian hyväksikäyttö uudistamaan ja muokkaaman olemassa olevia liiketoiminta – ja palvelumalleja sekä luomaan uutta liiketoimintaa (Gartner, Digitalization 2015).

Disaster recovery test
Testissä testataan miten yrityksen toiminta voi jatkua, kun osa tietotekniikkainfrastruktuurista yllättäen on poissa käytöstä.

Dokumentin esitysmuoto
Digitaalisesta dokumentista voidaan tehdä useita erilaisia ihmisen ymmärtämiä esitysmuotoja (engl. rendition). Esitysmuoto riippuu välineestä, jolla esitysmuoto näytetään ihmiselle. Välineitä ovat mm. paperi, nettisivusto tai kännykkä.

Folksonomia	engl. folsonomy, yhteisöllisesti tuotettu avoin luokittelujärjestelmä.
IPTC	engl. international press telecommunications council, metadatastandardi kuvatiedostoille.
ISO	engl. international organization of standardization, kansainvälinen standardointijärjestö.
Master data	Ydintieto on itsenäistä tietoa, joka on välttämätön yrityksen toiminnan ylläpitämiseksi esim. asiakastiedot
Metadata	Tietosisältöä kuvaavaa ja merkitystä selittävää tietoa. Tässä työssä sillä tarkoitetaan dokumentteihin liitettävää tietoa.
Metadatan attribuutti	Metadatamallissa oleva tietoelementti, joka sisältää tietoa. Käytännössä usein attribuutti vastaa yhtä metamallin tietokenttää.
Mpeg4	engl. motion picture expert group, ISO-14496 standardi videon ja äänen pakkaamiseksi.
OASIS	engl. organization for advancement of structured information standards, kansainvälinen järjestö, jonka tavoitteena on avointen IT – alan standardien kehittäminen ja edistäminen.
PDF	engl. portable documet format, eri järjestelmien väliseen dokumenttien vaihtoon Adode Systemsin kehittämä tiedostoformaatti, joka on sovellus -, laitteisto - ja käyttöjärjestelmä riippumaton.
Rakenteellinen dokumentti	Dokumentti, jolla on ennalta sovittu rakenne ja sisällön muoto.
Robotisaatio	engl. robotization, tarkoittaa työtehtävien ja prosessien automatisointia siten, että robotilla voidaan korvata ihminen.
Semi-rakenteellinen dokumentti	Dokummentti jonka rakenne ja sisällön muoto on osittain ennalta sovittu.
Strukturoitu data	engl. structured data, informaatio, joka on talletettu ennalta sovitun rakenteen ja muodon mukaisesti.
Taksonomia	engl. taxonomy, hierarkkinen luokittelumalli
TIFF	eng, tagged image file format, kuvien tallennukseen käytettävä häviötön tiedostomuoto.
XML	engl. extensible markup lauguage, standardoitu merkintäkieli tiedon esittämiseen ja kuvaamiseen

.

1. JOHDANTO

Diplomityössä tutkitaan miten dokumentteihin liittyvää metadataa voidaan saada ja miten sitä voidaan hyödyntää. Tavoitteena on selvittää lukijalle menetelmiä metadatan keräämiseen, sen laadun varmistamiseen liittyviä keinoja sekä dokumenttien metadata hyödyntämismahdollisuuksia. Työssä olevien metadatan keräämisen ja laadun varmistamisen menetelmien tarkastelussa pyritään antamaan lukijalle pohjatiedot, joita voidaan jatkokehittää ja tutkia lisää. Dokumenttien metadatan hyödyntämisestä, niin tietotekniikan kuin liiketoiminnan suunnittelussa, seurannassa ja hallinnassa, annetaan erilaisia näkökulmia. Jokaisessa kohdassa esitellään käytännön esimerkkejä vakuutusliiketoiminnasta. Tekijän motivaationa työn kirjoittamisella on ollut parantaa lukijoiden ymmärrystä dokumentteihin liittyvästä metadatasta ja sen käytettävyydestä osana yrityksen päätöksentekoa. Motivaationa oli myös mahdollisuus käyttää työn kirjoittamisessa työelämässä kertynyttä osaamista tällä alueella.

Datan määrä maailmassa ja yrityksissä kasvaa kiihtyvällä vauhdilla. Tämä tarkoittaa, että myös dokumenttivarastoissa olevien dokumenttien määrä kasvaa ja dokumenttien muoto ja rakenne muuttuvat ja monipuolistuvat. Dokumentit eivät ole enää vain tekstiä vaan kuvaa, ääntä ja videota sekä niiden yhdistelmiä. Dokumenttien monipuolistuminen ja varastojen kiihtyvä kasvu asettavat vaatimuksia varastojen hallintaa ja oikean tiedon löytymisellä. Yritykset hakevat uusia mahdollisuuksia pienentää tietotekniikkakustannuksia samalla kun ne hakevat kilpailuetua kehittämällä uusi palveluita asiakkailleen entistä nopeammassa tahdissa. Näihin vaatimuksiin voidaan vastata käyttäen dokumentteihin liittyvää metadataa, joko sellaisenaan tai yhdistettynä muualta saatavaan tietoon. Tämän mahdollisuuden käyttäminen kuitenkin edellyttää toimenpiteitä tarvittavan metadatan keräämiseen ja toimintatapojen muutoksia. Tekniset ratkaisut dokumenttien metadatan luomiselle ja keräämiselle ovat usein helppoja toteuttaa verrattuna muutoksiin yrityksen toimintatavoissa.

Diplomityössä luvussa 2 kuvataan metadataa ja erilaisia tiedostoformaatteja. Tarkoituksena on antaa metadatalle määritelmä ja menetelmät laadukkaan metadatan keräämiselle. Luvussa esitellään myös miten erilaiset tiedostoformaatit vaikuttavat kerättävän metadatan laatuun ja määrään. Luvussa 3 esitellään ehkä tunnetuin ja eniten käytetty dokumenttien metadatan hyödyntämiskohde, joka on dokumenttien elinkaarenhallinta. Erityisesti kiinnitetään huomiota dokumenttien omistajuuteen ja liiketoiminta-arvoon sekä automaatioon. Luvussa 4 esitetään dokumenttien metadatan hyödyntämismahdollisuuksia liiketoiminnassa perustuen dokumenttivarastoista saatavaan metadata pohjaiseen raportointiin, jota voidaan käyttää pohjatietona. Luvussa tutkitaan mm. metadatasta saatavan tiedon vaikutuksia uusiin palveluihin ja petosten havaitsemiseen. Luvun 5 tarkoituksena

on esitellä käytännön esimerkein dokumenttien metadatasta saatavan raportoinnin ja tietojen käyttömahdollisuuksia yrityksen sisäisessä toiminnassa. Huomiota kiinnitetään mm. tietotekniikan operatiivisiin kustannuksiin ja tietoturvaan. Luvussa 6 esitetään yhteenveto diplomityöstä sekä työn aikana havaitut jatkotutkimuskohteet.

2. DOKUMENTTIEN METADATA JA TIEDOSTOFORMAATIT

Luvussa 2 selvitetään metadatan luonnetta, rakennetta ja sisältöä. Luvussa käsitellään dokumenttien luokittelua ja dokumenttiautomatiikka. Tarkoituksena on esitellä, miten metadatan sisältö riippuu näistä ja mitä vaikutuksia erilaisilla tiedostoformaateilla on. Asiaa lähestytään jakamalla metadataan liittyvät asiat omiin kohtiinsa sekä esittelemällä dokumenttien luokittelu ja sen tavoitteita. Tiedostoformaattien vaikutuksia käsitellään tarkastelemalla eri formaattien soveltuvuutta erilaisiin käyttötilanteisiin.

2.1 Mitä on metadata?

Kohdassa 2.1 esitellään lukijalle, mitä metadata on, miten sitä voidaan tuottaa ja hallita. Lukijalle esitellään dokumenttityypitystä ja sen suhdetta dokumenttien metadataan.

2.1.1 Määritelmä

Alakohdassa 2.1.1 käydään läpi muutama metadatan määritelmä ja sekä esitellään tässä työssä käytetty määritelmä perusteluineen.

Metadatalle löytyy helposti paljon erilaisia määritelmiä. Metadata on tietoa tiedosta, jota voidaan liittää minkä tyyppiseen tietoon tai tietovarastoon tahansa. Systemaattinen metadatan tuottaminen ja kerääminen on edellytys tietovarastojen hallinnalle ja ylläpidolle. Ilman hyvää metadataa tietovarastot rämettyvät ja muuttuvat vähitellen käyttökelvottomiksi. Tietovarastojen eheys rapistuu ja tietojen haku ei anna oikeita tuloksia (Salminen 2005).

Arkistolaitos (2015) määrittelee ”Keskeisiä käsitteitä”-sivustollaan: *Metadata on tietoa tiedosta, tietosisältöä kuvaavaa ja sen merkitystä selittävää informaatiota. Metadata helpottaa aineiston hakua, paikallistamista, tunnistamista ja säilyttämistä sähköisessä muodossa. Metadatalla on tärkeä merkitys myös dokumenttien käsittelyvaiheessa.* Myös tässä määritelmässä esille tulee metadatan merkitys - kuvata tietoa sekä edesauttaa tiedon hallintaa ja hakua. Arkistojen hallinnan kannalta yksi keskeisimmistä metadatan hyödyntämiskohtaista on tiedon elinkaarenhallinta.

Tämän työn kannalta molemmat määritelmät pitävät paikkaansa. Dokumenttivarastot sisältävät erilaisia digitaalisia dokumentteja. Nämä dokumentit voivat olla hyvinkin erilaisia vaihdellen määrämuotoisesta vakiodokumentista valokuviin ja videoihin. Kaikkiin näihin erilaisiin dokumentteihin voidaan liittää metadataa, jotka kuvaavat niitä ja auttavat niiden hallinnoinnissa.

Metadatalla itsellään on tunnistettuja ominaisuuksia, jotka voivat vaihdella eri tietovarastojen välillä. Metadatan ominaisuuksia ovat mm. seuraavat (Salminen 2005):

- tuottaja
- tuottamisajankohta
- muutostenajankohta
- formaaliusaste
- pysyvyys
- käyttäjä

Metadataan liittyvien ominaisuuksien tarkempi tarkastelu vaatisi oman tutkimuksensa. Tämän työn kannalta riittää tieto, että myös metadatalla itsellään on ominaisuuksia ja että näihin ominaisuuksiin voidaan vaikuttaa.

2.1.2 Miten metadata kerätään ja luodaan?

Alaohdan 2.1.2 tarkoituksena on esitellä miten dokumentteihin liittyvää metadataa voidaan kerätä ja luoda. Siinä esitellään yhteisöllinen (folksonomiaan perustuva) metadatan luonti sekä automaatioon ja sääntöihin (taksonomiaan perustuva) perustuva metadatan luonti. Molempien menetelmien huonoja ja hyviä puolia tarkastellaan metadatan laadun näkökulmasta.

Metadataa tulee kerätä systemaattisesti. Sitä voidaan kerätä käyttäjiltä tai luoda automatiikan avulla. Tyypillisesti käyttäjien antama metadata on yhteisöllisesti tuotettua (engl. *folksonomy*) ja se perustuu vapaasti määriteltäville avainsanoille (engl. *key words*) ja tageille (engl. *tags*). Näin käyttäjiltä kerätyssä metadatassa ei ole hierarkiaa, josta selviäisi mitkä metadata-attribuutit ovat ylätason tietoja ja mitkä alatason tietoja. Tämä hierarkian puute johtuu siitä, että metadatalle ei ennen sen keruuta määritelty rakennetta ja mallia vaan ne muodostuvat vähitellen käyttäjien lisätessä yhteisöllisesti metadataa. Yhteisen metadatamallin (engl. *metadata schema*) puuttumisen myötä ei metadatan sisältöä voida standardoida tai yhtenäistää. Vaikka käyttäjien kirjoittama metadata on rakeenteellisesti yksinkertaista, sen tulkitseminen koneellisesti on hyvin vaikeata ja paljon laskentatehoa vaativaa. Tämä puolestaan voi aiheuttaa ongelmia tiedon etsinnässä kuten hakuaikojen pidentymistä ja puutteellisia tai vääriä hakutuloksia. Sen sijaan ennalta määriteltyyn taksonomiseen luokitteluun perustuvassa tiedon haussa voidaan hakuaikoja optimoida ja minimoida puutteelliset ja väärät tiedot. Tässä menetelmässä määritellään metadatamalli, joka voidaan vakioida ja yhtenäistää. Tämä metadatamalli puolestaan voidaan siirtää säännöiksi tietovarastojen hallintajärjestelmiin, sovelluksiin ja liiketoimintaprosesseihin, jolloin metadatan kerääminen voidaan automatisoida ja koneellistaa. Automatiikalla kerätty metadata on tasalaatuista ja yhtenäistä, mutta heikkoutena ovat mahdolliset virheet ja puutteet metadatamallissa, joita on vaikea havaita tai paikallistaa. Tästä syystä metadatamallia ja siihen perustuvaa automatiikkaan tulee tarkistaa ja tarvittaessa päivittää

säännöllisesti esimerkiksi liiketoiminnan muuttuessa tai järjestelmämuutoksien yhteydessä.

Dokumenttivarastoissa yleensä käytetään standardoitua metadatamallia ja siihen perustuvaan automatiikkaa metadatan luonnissa ja keräämisessä. Kirjoittajan tietojen mukaan dokumenttivarastojen yhteydessä on harvinaista käyttää puhdasta folksonomiaan perustuvaa metadatan keräämistä. Tämä johtuu dokumenttivarastojen luonteesta ja käyttötarkoituksesta. Käytössä voi kuitenkin olla välimalli, jossa valtaosa metadatasta kerätään metadatamalliin perustuvalla automatiikalla, mutta käyttäjät voivat myös antaa joitakin metadatatietoja. Tällaisia käyttäjiltä kysyttäviä tietoja voivat olla dokumenttiin liittyvä vapaamuotoinen kuvaus, käyttäjätunnus, dokumentin versio tai joissakin tapauksissa myöskin tietoturvataso.

2.1.3 Metadatan standardointi

Tämän alakohdan tarkoituksena on esitellä dokumentteihin liittyvän metadatan standardointia sekä sen vaikutuksia metadataan. Laatukriteereinä käytetään metadatan täsmällisyyttä, täydellisyyttä ja oikea-aikaisuutta.

Dokumenttivarastot sisältävät erilaisia dokumentteja, jotka voivat olla vakiomuotoisia asiakirjoja, valokuvia, sähköposteja, esityksiä, sopimuksia, videoita jne. Yhteistä kaikille näille erilaisille dokumenteille on, että niillä on jonkin tyyppinen rakenne ja rakenteesta johtuvia ominaisuuksia. Näitä hyväksikäyttäen voidaan määritellä hierarkkinen dokumenttien tyypitys ja luokittelu. Esimerkiksi vakuutusehdot on yksi dokumenttityyppi, jolla on ominaisuuksina liiketoiminta-aluetieto (yksityis, yritys tai teollisuus) ja vakuutustyyppi (esim. palo –, irtaimisto – tai varkausvakuutus). Nämä ominaisuudet voivat jakaantua aliominaisuuksiksi. Esimerkiksi maatalousyrittäjä on aliominaisuus liiketoiminta-aluetiedolle yritys. Näistä tiedoista voidaan rakentaa erilaisiin dokumentteihin liittyvä ominaisuusluokitus, josta puolestaan voidaan rakentaa näihin dokumentteihin liittyvä metadatamalli attribuutteineen. Tämä metadatamalli sisältää usein päällekkäistä tietoa. Saman tiedon määrittäminen ja lopuksi tallentaminen useaan kertaan ei ole käytännöllistä eikä suotavaa. Tästä syystä metadatamalli tulee standardoida ja optimoida poistamalla sieltä päällekkäiset ja tarpeettomat tiedot. Ennen rakenteellisen metadatan standardointi voidaan metadatalle ja sen sisällölle asettaa rajoitteita ja vaatimuksia (Pirttimäki 2012). Vakuutusliiketoiminnassa päällekkäisyyttä aiheuttavia metadata-attribuutteja ovat mm. asiakasnumero, vakuutusnumero ja osoitetiedot. Esimerkkinä tarpeettomasta tiedosta metadatassa ovat valokuvien resoluutiotiedot, joilla ei vakuutusliiketoiminnan kannalta ole merkitystä ja ne jätetään tallentamatta metadataan. Liiketoiminta ja käyttötarkoitus siis ohjaavat metamallin standardointia ja optimointia. Lisäksi tulee huomioida erilaisten tietojärjestelmien tekniset vaatimukset metadatamalliin.

Standardoitu metadatamalli voidaan siirtää erilaisiin järjestelmiin ja automatisoida. Näin päästää tilanteeseen, jossa dokumenttien metadataa luodaan, muokataan ja ylläpidetään

koneellisesti. Dokumentin siirtyessä liiketoimintaprosessissa vaiheesta toiseen siihen liittyvää metadataa muutetaan vastaamaan kyseistä prosessin ja dokumentin elinkaaren vaihetta. Koneellisesti tuotettu metadata on täsmällistä ja standardoidun metadatamallin mukaista. Dokumenttien metadata on myös täydellistä, koska se täyttää metadatamallin vaatimukset. Tämä ei kuitenkaan tarkoita, että metadatamallia ei tulisi päivittää tai että yhdelläkään dokumentilla ei olisi tyhjiä attribuutteja metadatassaan. Käytännössä standardoitu metadatamalli on elävä malli, jonka toimivuutta ja laatua tulee tarkkailla ja päivittää aktiivisesti säännöllisin väliajoin. Dokumenttien tyyppi ja luonne määräävät, mitkä metadatan attribuutit ovat käytössä. Esimerkiksi asiakkaan lähettämän sähköpostin metadataan ei ole tarkoituksen mukaista liittää vakuutussopimuksen tunnusta. Tälle tunnuksella varattu metadata-attribuutti jätetään tyhjäksi sähköpostin metadataa luotaessa. Voidaankin siis sanoa, että dokumenttivarastossa olevien dokumenttien metadata on täydellistä standardoidun metadatamallin puitteissa. Metadatan laatukriteereistä oikea-aikaisuudella tässä työssä tarkoitetaan metadatan luontia ja muokkausta sekä sen suhdetta dokumentin elinkaaren vaiheisiin. Standardoidulla metadatamallilla varmistetaan, että dokumenttien metadata elää dokumentin elinkaaren ja dokumenttiin tehtävien muokkausten tahdissa. Oikea-aikaisuusvaade tulee täytettyä, kun dokumenttien metadataa muutetaan ja tuotetaan koneellisesti perustuen metadatamalliin. Automatiikalla tuotettuun metadataan voidaan tarvittaessa lisätä käyttäjiltä kysyttävää lisätieto. Tätä varten on metadatamalliin määriteltävä tarvittavat attribuutit. Näin käyttäjiltä kerätty lisätieto ei pysy ajan tasalla automaattisesti vaan se vaatii tarvittaessa käyttäjiltä manuaalista päivitystä. Kirjoittajan tietojen mukaan ei ole kovin yleistä, että standardoituun metadatamalliin pyydetään käyttäjiä lisäämään metadataa, koska standardoidulla metadatamallilla ja metadatalla juuri pyritään mahdollisimman suureen automatiikkaan ja sitä kautta saavutettavaan korkealaatuiseen oikea-aikaiseen metadataan.

2.1.4 Metadatan yhtenäistäminen

Alakohdan 2.1.4 tarkoituksena on tarkastella standardoidun metadatan yhtenäistämistä ja sen tarpeellisuutta. Yhtenäistämisen mahdollisuudet ja tavoitteet esitellään liiketoimintanäkökulmasta.

Metadatan standardoinnin tarkoituksena on määritellä ja rakentaa metadatamalli attribuutteineen, jonne sitten voidaan järjestyksessä ja luokitellusti laittaa sisältöä. Attribuutteihin laitettava sisältö eli varsinainen metadata sisältää tietoa erilaisissa muodoissa. Saman attribuutin sisältämän tiedon muoto, pituus ja rakenne voivat vaihdella. Esimerkiksi asiakasnumeroattribuutin sisältämä tieto voi olla pituudeltaan vaihteleva ja sisältää vain numeroita tai numeroita ja kirjaimia riippuen, minkä yhtiön liiketoiminta-alueen asiakkaasta on kyse. Vakuutusyhtiön yksityisasiakkailla on erilainen asiakasnumero kuin yritysliiketoiminta-alueen asiakkailla. Lisäksi eri mailla on erilaiset asiakasnumeroinnit, jos yritys toimii useassa maassa.

Metadatan yhtenäistämisen tavoitteena on vakioida ja yhtenäistää metadatamalliin laitettava tieto. Tämä tarkoittaa, että jokaiseen attribuuttiin laitetavan tiedon muoto, pituus ja rakenne määritellään ja sovitaan yhtiön liiketoiminnan kanssa. Esimerkkinä asiakasnumero voidaan yhtenäistää sopimalla, että kaikissa liiketoiminta-alueissa asiakasnumero on pituudeltaan kuusi merkkiä, sisältää vain numeroita eikä sisällä maakoodia. Tämän lisäksi voidaan vielä sopia, että asiakasnumero on generoitu numerosarja, joka ei sisällä organisaatio- ja liiketoiminta-aluetunnuksia tai muutakaan tietoa rakenteessaan. Näille tiedoille on standardoidussa metamallissa varattu omat attribuutit, jotka puolestaan on yhtenäistetty.

Liiketoimintanäkökulmasta metadatan standardointi ja sen edut ovat selkeitä ja paremmin ymmärrettäviä kuin yhtenäistämisen. Metadatan yhtenäistäminen on huomattavasti vaikeampi ja työläämpi tehtävä. Usein liiketoiminta ei halua edetä metadatan standardointia pidemmällä. Liiketoimintanäkökulmasta tämä on ymmärrettävää, koska liiketoimintaorganisaatiot ovat tottuneet käyttämään epäyhtenäistä metadataa. He osaavat sujuvasti tulkita eri liiketoiminta-alueiden metadataa ja myöskin liiketoimintaprosessit on rakennettu käyttäen epäyhtenäistä metadataa. Tästä johtuen metadatan yhtenäistäminen ei tuo välitöntä etua tai parannusta liiketoiminnalle. Metadatan yhtenäistäminen tuo liiketoiminnalle välillisiä parannuksia. Voidaan rakentaa vain yksi ohjelmistokomponentti usean sijaan käsittelemään kaikkien liiketoimintayksiköiden metadataa. Liiketoiminta prosessit voidaan yhdistää ja luopua päällekkäisistä työkuluista prosessien sisällä. Tämä sovellus- ja prosessirakenteiden yksinkertaistaminen ja yhtenäistäminen laskee sovelluskehityksen kustannuksia, nopeuttaa sovelluskehitysprojekteja, yksinkertaistaa virheen selvitystä ja parantaa järjestelmien käytettävyyttä. Tämän kehityksen tuloksena liiketoiminnan maksamat tietotekniikkakustannukset laskevat ja tuovat näin kilpailuetua kilpailijoihin nähden sekä mahdollistavat nopeamman palvelukehityksen.

2.2 Tiedostoformaatit ja tiedostomuunnokset

Kohdassa 2.2 käsitellään, erilaisia dokumenttien tiedostoformaatteja ja tiedostoformaattien muunnosta. Kohdan tavoitteena on antaa lukijalle käsitys, miten metadataa voidaan hyväksikäyttää, muuntaa ja luoda tiedostomuunnoksien (engl. file conversion) yhteydessä.

2.2.1 Massatulostusformaatit

Alakohdassa 2.2.1 esitellään massatulostukseen liittyvät tiedostoformaatit sekä niiden metadata.

Massatulostuksessa käytettävän tiedostoformaatin tulee olla tehokkaasti pakattua, mahdollistaa suuret tulostusnopeudet, tulostuksen virheenkorjauksen sekä tulostuksen seurannan ja hallinnan. Massatulostusformaatti, joka sisältää nämä ominaisuudet ja soveltuu parhaiten massatulostukseen, on AFP-formaatti (engl. *advanced function printing*). AFP-

formaatilla päästään suuriin tulostusnopeauksiin virheenkorjauksella (AFP Consortium 2015). Tulostus- ja painotalot haluavat lähes poikkeuksetta massatulostettavat dokumentit AFP-eräajoina. Esimerkiksi Pohjoismaissa toimiva vahinkovakuutusyhtiö lähettää ulkopuoliselle tulostustalolle massatulostukseen menevän materiaalin AFP-eräajoina. Yhtiön tulostusmäärät ovat miljoonia A4-sivuja kuukaudessa.

AFP-formaatin rinnalla käytetään PDF-formaattia (ISO 2008). Tällöin tulosmäärät ovat pieniä, sillä PDF-formaattia käytettäessä ei päästä suuriin tulostusnopeuksiin, jossa tulostetaan satoja sivuja minuutissa. Lisäksi tulostuksen hallinta ja virheenkorjaus eivät ole samalla tavalla mahdollisia kuin AFP-formaattia käytettäessä. Tästä syystä useat yritykset tekevät ennen massatulostusta tiedostomuunnoksen ja uudelleen paketoinnin PDF-formaatista AFP-formaattiin. Esimerkkinä PDF-formaatin käytöstä on asiakkaille tulostettavat ja lähetettävät laskut, joiden sivu- ja kokonaismäärä ovat pieniä, jolloin ne voidaan tulostaa käyttäen PDF-tiedostomuotoa.

Kirjoittajan tietojen mukaan PostScript-formaattista ennustetaan uutta massatulostusformaattia, joka mahdollisesti jollakin aikavälillä voisi syrjäyttää AFP-formaatin ensisijaisena massatulostusformaattina. PostScript sisältää mm. eräajojenhallinnan, muistin- ja tulosteiden hallinnan käytön hallinnan (Adobe 2015 ja Fermilab 2015). Tällä hetkellä ProstScriptiä käytetään vain vähän massatulostuksessa.

Massatulostus ei aseta suuria vaatimuksia metadatalle. Riittää, että tulostusmateriaali saadaan poimittua esimerkiksi vakuutuslajin, päivämäärän, vakuutusnumeroin tai näiden yhdistelmien perusteella. Tästä syytä AFP-muotoisten dokumenttien dokumenttivarastojen metadata on rakenteeltaan yksinkertaista ja määrältään vähäistä.

2.2.2 Arkistointiformaatit

Alakohdassa 2.2.2 esitellään arkistointiin liittyvät tiedostoformaatit ja arkistoinnin vaatimukset dokumenttien metadataan.

Arkistointi tehdään arkistointisuunnitelmassa kerrotulla tavalla. Suunnitelmassa otetaan kantaa mm. arkistoitavien dokumenttien arvoon, arkistoinnissa käytettäviin tiedostoformaatteihin ja arkistoitujen dokumenttien hävitykseen. Dokumentteja kannattaa säilyttää arkistoissa vain niin kauan kuin niillä on arvoa, mutta ei sen pidempään kuin vaaditaan (Lyytikäinen 2004). Paperiarkistossa dokumentit säilyivät niin kauan kunnes ne aktiivisesti hävitettiin. Digitaalisessa arkistoissa asiat ovat päinvastoin. Dokumentit säilyminen edellyttää aktiivisia toimenpiteitä, sillä muuten ne häviävät arkistosta. Dokumenttien arvoon otetaan kantaa dokumenttien metadatassa, mutta arkistoinnissa käytettävät tiedostoformaatit vaikuttavat dokumenttien säilyvyyteen, käytettävyyteen ja luettavuuteen. Suositeltavia tiedostoformaatteja pitkäaikaissäilytykseen, jotka täyttävät edellä olevat vaatimukset, ovat PDF/A-1a, PDF/A-1b, PDF/A-2a, PDF/A-2b tekstimuotoisille dokumenteille (Alliance Group 2015), TIFF (rev 5 ja rev6) kuville (Digital Preservation Org

2002) sekä rekisterimuotoiselle tiedolle Arkistolaitoksen (2009) SÄHKE-normit määrittelevät XML-rakenteen. Videoiden ja äänitteiden pitkäaikaissäilytyksessä käytettävien formaattien tilanne ei vielä ole sammalla tavalla vakiintunut kuin dokumenttien ja kuvien kohdalla. Esimerkiksi mpeg4-muotoisen tiedoston tekemiseen voidaan käyttää useita keskenään erilaista koodekia (engl. codecs). Sama tilanne on äänitteiden suhteen.

Tärkeimmät arkistoinnin metadatalle asettamista vaatimukset liittyvät säilytysaikaan, säilytettävän tiedon arvoon, elinkaareen ja omistajuuteen. Näiden tietojen avulla voidaan erilaisten dokumenttien elinkaarta ja niistä muodostuvia arkistoja hallita.

2.2.3 Avoin tiedostomuoto

Alakohdan 2.2.3 tarkoituksena on antaa yleiskatsaus avoimiin dokumenttiformaatteihin, jotka ovat laite- ja ohjelmistoriippumattomia. Lisäksi käsitellään avoimien dokumenttiformaattien vaikutukset metadataan.

Avoin tiedostoformaatti on tiedostomuotostandardi, jonka on kehitetty muokattavien toimistoasiakirjojen kuten teksti-, taulukkolaskenta- ja esitysgrafiikkadokumenttien tallentamiseen. Standardin on kehittänyt OASIS industry consortium. Standardin lähtökohtana on XML-formaatti (W3C 2015). Avoimen tiedostomuodon standardi löytyy ISO standardeista (ISO 2006). Avoin ODF -tiedostoformaattia (engl. *open document format*) mahdollistaa ohjelmisto- ja laitteistoriippumattoman dokumenttien käsittelyn ja talletuksen. ODF-formaattiin on mahdollista tallentaa dokumentteja monista toimisto-ohjelmista kuten *Microsoft Office*-, *LibreOffice*- ja *IBM Lotus Symphony*-ohjelmista. Myöskin nettitoimisto-ohjelmistot kuten *Google Docs* tukevat ODF-formaattia (OASIS 2015).

Avoin tiedostoformaatti ja sen käyttö on vapaa lisenssimaksuista, mikä tekee sen käytöstä yrityksille houkuttelevaa. Markkinoiden toimisto-ohjelmistojen johtaja Microsoft kehitti oman ODF-formaatin. Tämä formaatti on lisenssimaksullinen, jonka takia se ei ole houkutteleva vaihtoehto käyttäjille.

ODF-dokumenttien metadata on erillisessä tiedostossa, joka on paketoitu yhteen varsinaisen dokumentin kanssa. Metadata tiedostoja voi olla useita ja ne ovat XML-pohjaisia. Tämä mahdollistaa monipuolisen metadatan liittämisen dokumentteihin (OASIS 2007). Sovelluksien, jotka käsittelevät ODF-dokumentteja, tulee varmistaa, että metadata säilyy ja kehittyy dokumentin elinkaaren mukaisesti.

Kirjoittajan tietojen mukaan ODF-formaattia ei ole saanut kovinkaan paljon suosiota yritysmaailmassa. Usein yrityksissä on ODF-formaattia kokeiltu, mutta laajamittaiseen käyttöön sitä ei ole otettu. Syynä on ollut, että yrityksissä on pitkää totuttu käyttämään ja investoimaan *Microsoft Office*-ohjelmistoon. Näin on varmistettu, että yrityksen dokumentit avautuvat oikein kaikissa ohjelmistoissa sekä voitu maksimoida henkilöstön koulutukseen ja osaamiseen tehdyt investoinnit.

2.2.4 Käytännön de facto formaatit

Alakohdassa 2.2.4 esitellään lukijalle käytännössä käytettävät tiedostoformaatit sekä tarkastellaan, miten ne tulevat tulevaisuudessa muuttumaan.

Yrityksen dokumentit talletetaan dokumenttivarastoihin toimisto-ohjelmiston käyttämissä tiedostoformaateissa. Tällä hetkellä Microsoftin toimisto-ohjelmistot ovat valta-asemassa yritysmaailmassa, joten käytössä olevat talletusformaatit ovat doc-, docx-, xls-, xlsx-, ppt- ja pptx-muotoisia. Käytännössä myös muiden kuin Microsoftin toimisto-ohjelmistojen tulee pystyä käsittelemään ja tuottamaan näitä tiedostoformaatteja. Voidaan siis sanoa, että näistä formaateista on tullut käytännön standardeja, joita ilman on hankala tulla toimeen.

Lyhyellä aikavälillä on vaikeata nähdä muutoksia käytännön tiedostoformaattien käytössä. Pidemmällä aikavälillä avoimilla tiedostoformaateilla on mahdollisuuksia korvata nykyset monet erilaiset formaatit ja yhtenäistään formaatit. Ohjelmistoista ja laitteistoista riippumattomat formaatit helpottaisivat dokumenttien ja niiden metadatan siirtoa järjestelmistä ja palveluista toiseen, koska tiedostoformaattien muunnoksia ei enää tarvittaisi. Kirjoitta mielestä avoimien tiedostoformaattien yleistyminen tulee olemaan hidasta, koska yritykset ovat investoineet paljon nykyisiin käytännön tiedostoformaatteihin. Yritysten IT-infrastruktuuri ja henkilöstön osaaminen on rakennettu nykyformaattien varaan ilman varautumista nopeisiin muutoksiin.

2.2.5 Tiedostomuunnokset

Alakohdan 2.2.5 tarkoituksena on selventää tiedostokonversion vaikutuksia metadataan sekä antaa käsitys, milloin tiedostomuunnoksia tehdään. Tarkastelussa keskitytään tietoteknisen infrastruktuurin muutoksiin sekä monikanavaiseen asiakaskommunikointiin (engl. *omnichannel customer communication*).

Dokumentteja muunnetaan tiedostoformaatista toiseen tehtäessä muutoksia yrityksen tietotekniseen infrastruktuuriin, ohjelmistojen vaihdon yhteydessä tai lähetettäessä dokumentteja asiakkaalle. Erityisen tärkeäksi on viime aikoina noussut saman palvelukokemuksen tarjoaminen asiakkaille riippumatta käytössä olevasta päätelaitteesta. Tämä vaatii asiakkaan päätelaitteen tunnistusta ja dokumenttien muuntamista päätelaitteelle sopivaan formaattiin. Olipa tiedostomuunnoksen syy mikä tahansa, muunnoksissa metadata tulee ottaa huomioon. Metadatan määrä ja muoto vaihtelevat tiedostoformaatista toiseen. Tämä tulee ottaa huomioon tiedostomuunnoksia tehtäessä. Esimerkiksi AFP-formaatissa metadataa on vähän, koska sitä käytetään massatulostuksessa. Muunnettaessa dokumentteja AFP-formaatista PDF-formaattiin, joudutaan tilanteeseen, jossa PDF mahdollistaa AFP:tä laajemman metadatan käytön. Teknisesti on mahdollista luoda lisää metadataa tiedostomuunnosten yhteydessä ja näin ottaa käyttöön toisen formaatin laajempi meta-

data. Liiketoimintatarve ja käyttötarkoitus ratkaisevat, siirretäänkö lähdeformaatin metadata sellaisenaan kohdeformaatin metadataksi, vai muutetaanko metadataa tiedostomuunnoksen yhteydessä luomalla uutta tai muokkaamalla vanhaa. Esimerkkinä olevassa AFP-muunnoksessa PDF-formaattiin useimmiten metadata siirretään sellaisenaan ilman muutoksia, koska dokumenttien liiketoiminta-arvo ja käyttötarkoitus eivät sitä vaadi muutoksia metadataan. Sen sijaan esimerkiksi dokumenttivarastojen yhdistämisissä lähdejärjestelmästä siirrettävien dokumenttien metadataa muokataan ja muutetaan vastaamaan kohdejärjestelmän metadatamallia ja vaatimuksia.

Yhteenvetona voidaan todeta, että tiedostomuunnoksia tehtäessä tulee metadata ottaa huomioon määrittelemällä yhdessä liiketoiminnan kanssa, miten metadataa käsitellään tiedostomuunnoksissa. Dokumenttien metadata voidaan siirtää muuttumattomana formaatista toiseen tai lisätä ja muuntaa sitä tarvittaessa.

3. METADATAN HYÖDYNTÄMINEN DOKUMENTTIEN ELINKAARRESSA

Luvussa 3 käsitellään dokumentteihin liittyvän määrämuotoisen ja yhtenäisen metadatan hyödyntämistä dokumenttien elinkaarenhallinnassa. Tämä tapahtuu perehdyttämällä lukija eri näkökulmiin käyttää hyväksi metadataa ja antamalla käytännön esimerkkejä vakuutusliiketoiminnan tavasta käyttää metadataa. Tarkoituksena on myös antaa uusia ideoita ja selvittää lukijalle, miten monipuolisesti dokumenttien metadataa voidaan käyttää, joko sellaisenaan tai yhdessä muualta saatavan tiedon kanssa.

3.1 Vakuutustoiminta

Kohdan 3.1 tarkoituksena on antaa lukijalle tietoa vakuutusliiketoiminnasta kertomalla vakuutuksista ja vakuuttamisesta.

Vakuutus voidaan määritellä usealla tavalla. Riskien kannalta vakuutus on menettely, jossa tietyn riskin tai tiettyjen riskien alaiset henkilöt tai yritykset sopivat vahinkojen tasaamiseen erikoistuneen laitoksen kanssa. Sopimuksessa riskin toteutuessa laitos korvaa aiheutuneen vahingon. Näin syntyvää sopimusta kutsutaan vakuutukseksi ja sopijaosapuolena olevaa laitosta vakuutusyhtiöksi. Vakuutusta voidaan tarkastella myös liiketalouden näkökulmasta. Tällöin liiketaloudellisesti vakuutus määritellään keinoksi, jolla ennalta arvaamattomasti ja sattumanvaraisesti toteutuvasta vahingosta aiheutuvat kustannukset voidaan muuntaa vuosittain jatkuvaksi tasaiseksi kustannukseksi. Näitä perusteita käyttäen voidaan taas määritellä asiakkaalta perittävä vuosimaksu (Hoppu, Hemmo 2006).

Vakuutukseen liittyy joukko ominaisuuksia. Näitä ominaisuuksia ovat:

- Vakuutuksen kohteena olevan riskin toteutuminen riippuu ennalta arvaamattomasta ja epävarmasta tapahtumasta tai tapahtuma sarjasta. Riskin toteutuminen on siis sattumanvaraista.
- Vakuutukseen liittyy vahingonmahdollisuus. Mahdolliset vahingot on kuvattava vakuutuksessa, sillä vakuutus ei korvaa kuin ennalta määritellyt vahingot.
- Vakuutusmaksujen ja vakuutuksesta maksettavien korvausten tulee olla tasapainossa tarkasteltaessa pitkällä aikavälillä. Tässä tarkastelussa otetaan huomioon myös vakuutuksen liittyvät hoitokustannukset esimerkiksi säilytyskustannukset ja erilaiset postituskustannukset. Vakuutusmaksun ja riskin tulee siis vasta toisiaan.
- Usein vakuutettava riski on samankaltainen suurella joukolla vakuutuksenottajia, jolloin riski ja kustannukset voidaan tasata tämän suuren joukon kesken. Tasaus mahdollistaa alhaiset vakuutusmaksut verrattuna vakuutettuun riskiin.
- Riippumattomuus vakuutuksenottajista. Tällä tarkoitetaan, että vakuutuksenantajan ja -ottajan välillä ei ole eikä saa olla riippuvuuksia.

Edellä olevia vakuutukseen liittyviä ominaisuuksia voidaan käyttää vakuutusdokumenttien metadatan attribuutteina. Nämä ominaisuudet ovat yhteisiä kaikille vakuutuslajeille. Näiden yhteisten ominaisuuksien lisäksi on vakuutuslajikohtaisia ominaisuuksia, joita voidaan myös käyttää dokumenttien metadata-attribuutteina.

Vakuutukset voidaan jakaa kahteen luokkaan, jotka ovat yksityisvakuutukset ja sosiaalivakuutukset. Usein puhutaan yritysvakuutuksista omana vakuutuslajina. Yritysvakuutuksella tarkoitetaan vakuutusta, joka kattaa yrityksen käyttö- ja vaihto-omaisuuden sekä rakennukset. Yritysvakuutuksella siis voidaan vakuuttaa samoja riskejä ja kohteita kuin yksityisvakuutuksessa. Tässä työssä ei yritysvakuutusta käsitellä erillisenä itsenäisenä vakuutuslajina vaan osana yksityisvakuutuslajia.

Yksityisvakuutukset voidaan taas jakaa neljään luokkaan riippuen vakuutettavasta kohteesta. Nämä luokat ovat seuraavat:

- Henkilövakuutukset
- Vahinkovakuutukset
- Liikenteen ja kuljetusten vakuutukset
- Yhdistelmävakuutukset

Kaikki edellä olevat vakuutusluokat jakautuvat alaluokkiin, jotka kertovat, mitä kohdetta tai riskiä vakuutetaan. Henkilövakuutukset käsittävät henki-, tapatuma- ja eläkevakuutukset. Vahinkovakuutukset sisältävät palo-, oikeusturva-, vastuu-, takaus-, luotto-, laina-, maksuturva- ja keskeytysvakuutukset.

Sosiaalivakuutukset liittyvät kansalaisten sosiaaliturvaan ja niillä varaudutaan sosiaalisiin henkiriskeihin. Tällaisia riskejä ovat työkyvyttömyys, työttömyys ja vanhuus. Sosiaalivakuutukset eivät kata tai vähennä omaisuuteen kohdistuvia vahinkoja. Riskien toteutuessa sosiaalivakuutuksen tarkoituksena on turvata toimeentulo ja korvata aiheutuneita kuluja.

Sosiaalivakuutuksia ovat seuraavat vakuutukset:

- Sairausvakuutus
- Työeläkevakuutus
- Työtapaturmavakuutus
- Kansaneläkevakuutus
- Työttömyysvakuutus
- Työntekijöiden ryhmähenkivakuutus

Vakuutuksilla on, riippumatta siitä mikä vakuutus on kyseessä, joukko yhteisiä attribuutteja, joita voidaan käyttää metadatassa. Tällaisia yhteisiä attribuutteja ovat esimerkiksi asiakasnumero, asiakkaan nimi, osoite ja puhelinnumero. Vakuutuksiin liittyy myös erilaisia dokumenttityyppejä, joista osa on kaikille yhteisiä. Yhteisiä dokumenttityyppejä

ovat vakuutuskirja, vakuutusehdot ja vahinkoilmoitus. Näihin yhteisiin dokumenttityyppeihin liittyy metadata-attribuutteja, jotka ovat yhteisiä. Näiden yhteisteisten attribuuttien lisäksi jokaisella vakuutuksella on kyseiseen vakuutukseen ja vakuutustyyppiin liittyviä omia attribuutteja, joita voidaan käyttää dokumenttien metadatassa. Esimerkkejä tällaisista attribuuteista ovat tapaturmavakuutukseen liittyvät asiakkaan yksilölliset terveystiedot tai palovakuutuksessa kohteena olevan omaisuuden tiedot. Vakuutuksilla voi olla vain siihen liittyviä dokumenttityyppejä, joilla on omat metadata-attribuutit. Tällaisia dokumenttityyppejä ovat esimerkiksi erilaiset omaisuusluettelot ja lääkärinlausunnot.

Vakuutusluokat ja -tyypit eivät ole staattisia, vaan ne muuttuvat ajan kuluessa. Osa vakuutuksista poistuu käytöstä, uusia vakuutuksia syntyyn tai olemassa olevien vakuutusten sisältö muuttuu. Esimerkkejä aikaisemmin käytössä olleita vakuutuksia, jotka ovat poistuneet käytöstä:

- Myötäjäisvakuutus
- Eläinvakuutus
- Kavallusvakuutus
- Lasivakuutus

Vakuutuksien poistuessa käytöstä se vaikuttaa dokumenttien elinkaareen ja metadataan. Dokumentit ja niihin liittyvä metadata tulee tuhota pysyvästi. Tämän lisäksi muutoksia tulee metadataan liittyviin standardeihin ja metadatamalliin (engl. *schema*).

Yleisesti voidaan todeta, että vakuutuksiin liittyvien dokumenttien metadata koostuu kaikille vakuutuksille yhteisistä attribuuteista ja kullekin vakuutustyypille ominaisista attribuuteista. Dokumenttien metadata talletetaan yhdessä dokumentin kanssa ja se elää ja muuttuu dokumentin elinkaaren mukana.

3.2 Metadatan hyödyntäminen dokumenttien elinkaaren hallinnassa

Kohdassa 3.2 tarkoituksena on esitellä dokumenttien elinkaaren hallintaa ja siihen liittyvää metadataa.

Dokumenttien elinkaaren hallinta on ehkä tunnetuin ja käytetyin metadatan hyödyntämisen tapa. Elinkaaren hallinta on tärkeää kaikille yrityksille. Hyvällä dokumenttien ja siihen liittyvän metadatan elinkaaren hallinnalla saadaan isotkin dokumenttimassat hallintaan ja sitä voidaan hyödyntää tehokkaasti. Dokumentit saadaan nopeasti ja helposti haettua sekä voidaan olla varmoja, että haettu dokumentti on ajan tasalla ja viimeisintä versiota. Päätöksenteossa kaikilla on samat ajantasaiset tiedot käytettävissään nopeaan ja laadukkaaseen päätöksentekoon, jolla ohjataan yritystä ja organisaatioita.

Dokumenttien elinkaaren hallintaan voidaan käyttää erilaisia teknisiä ratkaisuja ja tuoteperheitä (engl. *product suite*). Ratkaisut tarjoavat välineitä dokumenttien luonnista aina

pitkäaikaisarkistointiin ja pysyvään tuhoamiseen asti. Kaikissa näissä ratkaisuissa on seuraavat osat:

- Dokumenttien luonti-, versiointi- (engl. *document composition tools*) ja hallintavälineet (engl. *document management tools*)
- Fraasien ja vakioitujen tekstikappaleiden (engl. *standard text blocks*) luonti-, talletus- ja hallintavälineet
- Tietovarasto dokumenttien säilytykseen ja hallintaan (engl. *document repository*)
- Dokumenttien arkistointi (engl. *document archive*)
- Metadatan standardointi, luonti ja hallinta välineet
- Välineet dokumenttimassan sääntöpohjaiseen ja automaattiseen käsittelyyn
- Dokumenttien indeksointi- ja hakuvälineet

Edellä mainittujen osien lisäksi tuoteperheet voivat sisältää tekniset ratkaisut dokumenttien tiedostomuunnoksille (engl. *transformation tools, rendention tools*), video- ja äänitiedostojen käsittelylle, kuvien käsittelylle sekä ratkaisut monikanavaiseen asiakaskommunikointiin (engl. *multichannel customer communication, omnichannel customer communication*).

Dokumentit kulkevat elinkaarensa aikana läpi erilaisia vaiheita kunnes ne lopulta tuhotaan pysyvästi. Dokumentit siirretään elinkaaren vaiheesta toiseen automaattisesti ennalta määriteltyjen sääntöjen mukaisesti. Säännöt puolestaan on rakennettu käyttäen dokumenttien metadataa. Yksinkertaisimmillaan dokumentteja vaiheesta toiseen siirtävä sääntö käyttää hyväkseen vain yhtä metadatan attribuuttia, joka on dokumentin ikä. Usein tämä on aivan riittävän hyvä sääntö, jos kyseessä ovat dokumentit, joihin ei kohdistu viranomaissäädöksiä tai lakisääteisiä vaatimuksia. Tällaisia dokumentteja ovat esimerkiksi yrityksen sisäiset kokousmuistiot ja kokouksien esityslistat.

Yleensä siirto kuitenkin tehdään useiden eri metadatan attribuuttien yhdistelmänä. Dokumentin iän lisäksi säännöissä usein käytettyjä metadatan attribuutteja ovat dokumenttityyppi, dokumentin versionumero ja organisaatio. Irtaimistovakuutuksen vakuutusmaksusta asiakkaalle lähetetty lasku esimerkiksi siirtyy elinkaaren vaiheesta toiseen säännöllä, jossa käytetään laskun luontiaikaa, laskun eräpäivää, tyyppiä lasku ja vakuutuksen kohdetta irtaimisto. Kuten esimerkistä huomataan dokumentilla voi olla monia aikaleimoja (engl. *time stamps*), jotka se saa kulkiessaan elinkaaren vaiheesta toiseen. Esimerkissä olevaa dokumenttityyppiä ei versioida, sillä asiakkaalle lähetettävä lasku ei muutu kulkiessaan elinkaaressaan vaiheesta toiseen. Mahdollisesti asiakkaalle lähetettävät muistutus- tai korjauslaskut ovat uusia dokumentteja, eivät alkuperäisen laskun versioita. Näin ollen dokumentin version kertovaa attribuuttia ei voida käyttää säännössä, joka siirtää laskun dokumentin elinkaaressa eteenpäin.

Muita dokumentin elinkaarenvaiheen määrääviä attribuutteja ovat dokumenttien liiketoiminta-arvo (engl. *bussiness value*), dokumenttien voimassaoloajat, lainsäädännön vaatimukset, dokumenttien tyyppi, liiketoiminta-alueen vaatimukset, maakoodi, dokumentin tekijä, dokumentin omistaja ja viranomaisvaatimukset ja määräykset.

Organisaatiotiedon käyttö metadata-attribuuttina on yleistä ja usein myös perusteltua. Organisaatiotiedon käyttöön dokumenttien metadatassa liittyy riskejä ja vaaroja. Yritysten organisaatiot eivät ole pysyviä, vaan ne vaihtuvat joko sisäisten organisaatiouudistusten toimesta tai yrityskauppojen myötä. Organisaation muuttuessa myös dokumenttien metadatan sisältö muuttuu. Organisaatiokerrosten määrän lisääntyessä tai pienentyessä dokumenttien metadatamallissa olevia attribuutteja tulee korjata vastaavasti. Tapauksessa, jossa ei tehdä muutoksia organisaatiokerroksiin tai -hierarkiaan, mutta tehdään muutoksia organisaatioyksiköiden nimeämiseen ja sisältöön, myös metadatan attribuuttien sisältämä tieto muuttuu. Kummassakin tapauksessa joudutaan ottamaan kantaa siihen, muutetaanko dokumenttivarastossa jo olevien dokumenttien metadata vai ei. Kummatkin ratkaisut ovat teknisesti mahdollista toteuttaa. Olemassa olevien dokumenttien metadata voidaan päivittää automaattisilla dokumenttimassan huoltoajoilla. Ongelmana tässä on metadatan päivittämisen vaatima aika isoilla dokumenttimassoilla. Yleensä tällaiset ajot suoritetaan yöllä, jolloin dokumenttivaraston käyttäjiä on vähän. Käytännössä yleisempi ratkaisu on jättää dokumenttivarastossa oleva dokumenttien metadata päivittämättä ja ottaa uusi metadatamalli ja sisältö käyttöön jostakin sovitusta ajankohdasta lähtien. Yhteenvetona voidaan todeta, että organisaatiotiedon käyttöä dokumenttien metadatassa tulee harkita tarkkaan. Yksistään itsenäisenä organisaatiotieto dokumentin metadatassa ei useinkaan ole tarpeellinen. Yhdistettynä organisaatiotieto dokumenttien omistajuuteen tuo paljon lisäarvoa dokumenttien elinkaaren hallintaan. Lisäksi organisaatiotieto voidaan usein korvata liiketoiminta-aluetiedolla, joka on organisaatiotietoa pysyvämpi metadata-attribuutti, sillä harvoin yrityksen liiketoiminta-alue tai liiketoimintaidea muuttuvat.

Usein myös dokumenttien omistajalla voi olla ja yleensä on oma näkemys siitä milloin dokumentti voidaan siirtää sen elinkaaressa vaiheesta toiseen. Erityisesti dokumentin elinkaaren viimeisessä vaiheessa, jolloin dokumentti pitäisi pysyvästi tuhota dokumentin omistaja määrää tai vahvistaa tuhoamisajankohdan.

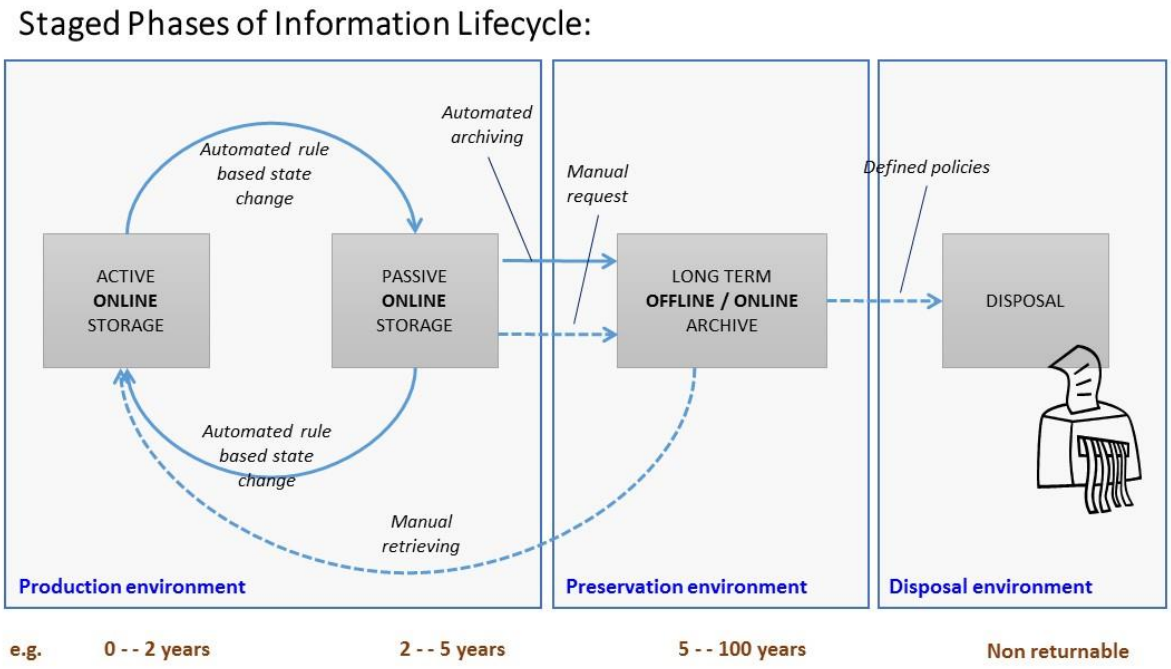
Dokumenttien omistajuus on yksi tärkeimmistä metadata-attribuutteista. Dokumenttien omistajat ovat tärkeässä roolissa, kun automaation vaatimia sääntöjä laaditaan, testataan ja hyväksytään tuotantoon. Omistajien tulisi tulla siitä organisaatiosta, joka dokumentit luo. Tämä tarkoittaa käytännössä liiketoimintayksikköjä ja omistajat ovat liiketoimintayksikköjen henkilöitä. Dokumenttien omistaja voi olla:

- yksi nimetty henkilö
- erikseen nimetty ryhmä henkilöitä
- jokin organisaation osa

Dokumenttien elinkaaren hallinnassa dokumenttien omistajat ovat avain asemassa. Omistajat tekevät yhteistyössä tietotekniikka-ammattilaisten kanssa käytettävän metadatamallin attribuutteineen sekä vastaavat automatiikan käyttämien sääntöjen oikeellisuudesta, toimivuudesta ja laadusta.

Dokumenttien elinkaaren vaiheet ovat yhteydessä liiketoimintaprosesseihin. Suhde ei ole yksi yhteen suhde missä dokumentti siirtyisivät elinkaaressaan eteenpäin aina kun niiden paikka liiketoimintaprosesseissa vaihtuu. Liiketoimintaprosesseissa on enemmän vaiheita kuin dokumenttien elinkaaressa. Liiketoimintaprosessin alussa luotu uusi dokumentti pysyy samassa elinkaarensa vaiheessa vaikka se kulkee liiketoimintaprosessissa usean vaiheen läpi. Uusi tarjous asiakkaalle käy läpi liiketoimintaprosessin vaiheet tarjousluonnos, lopullinen tarjous ja tarjouksen hyväksyminen, mutta dokumentin elinkaaren kannalta se pysyy samassa vaiheessa elinkaarensa alussa. Kuitenkin jokainen liiketoimintaprosessin vaihe aiheuttaa dokumentin metadatan attribuutteihin muutoksia, joka kertovat dokumentin tilan (engl. *document status*). Jokainen liiketoimintaprosessin vaihe aiheuttaa siis muutoksen dokumentin metadataan ja dokumentti siirtyy dokumentin elinkaaressa eteenpäin kun dokumentin metadata on tietyllä tasolla. Esimerkkitapauksessamme tarjousdokumentti siirtyy elinkaaressa eteenpäin, kun liiketoimintaprosessin tarjousvaiheeseen liittyvät vaiheet on suoritettu ja niitä vastaava metadata on muodostunut.

Dokumenttien elinkaaren vaiheet ja niiden hallinta voidaan kuvata seuraavalla kuvalla. Kuvassa dokumentit siirtyvät eri vaiheiden (engl. *stages*) välillä automaattisesti ja tarvittaessa myös manuaalisesti.



Kuva 1. *Kuvassa esitetään informaation elinkaaren vaiheet sekä miten informaatio liikkuu vaiheesta toiseen. Tätä elinkaarimallia voidaan soveltaa suoraan dokumenttienelinkaareen.(Baldeón 2010)*

Kuvassa 1 dokumenttien elinkaaren eri vaiheet on ryhmitelty kolmeen ryhmään. Nämä ovat:

- Tuotantoympäristöön (engl. *production environment*) liittyvät dokumentin elinkaaren vaiheet
- Säilytykseen ja arkistointiin (engl. *preservation environment*) liittyvät dokumentin elinkaaren vaiheet
- Pysyvään tuhoamiseen (engl. *disposal environment*) liittyvät dokumentin elinkaaren vaiheet

Kuvassa 1 tuotantoympäristössä on kaksi tietovarastoa, jotka molemmat on kytketty verkkoon. Nämä ovat aktiivinen tietovarasto (engl. *active online storage*) ja passiivinen tietovarasto (engl. *passive online storage*). Tällä halutaan korostaa eri elinkaaren vaiheessa olevien dokumenttien paikkaa tuotantoympäristössä. Käytännössä tietovarasto on molemmissa tapauksissa sama verkkoyhteydellä varustettu tekninen ratkaisu ja tietokanta, jossa dokumentit on erotettu niiden sisältämän metadatan avulla. On myös mahdollista rakentaa tekninen infrastruktuuri, missä on kaksi erillistä tietovarastoa. Tällainen ratkaisu monimutkaistaa tietotekniikkainfrastruktuuria ja aiheuttaa korkeat käyttökustannukset (engl. *operational costs*) johtuen infrastruktuurin monimutkaisemmasta hallittavuudesta.

Useamman dokumenttitietovaraston perustaminen on perusteltua, kun halutaan erottaa asiakasdokumentit ja yrityksen sisäisistä dokumenteista. Tietoturvan näkökulmasta tällainen erottaminen on järkevää ja suotavaa, koska näin varmistetaan, että asiakastiedot ja yrityksen sisäiset tiedot eivät vahingossa sekoitu keskenään. Molemmille dokumenttivarastoille voidaan määritellä oma standardi metadatamalli ja dokumenttien elinkaari sääntöinen. Tarvittaessa näiden kahden tietovaraston sisältämiä dokumentteja voidaan käyttää ristiin tai yhdessä. Käytännössä kuitenkin tällaisia tarpeita on vähän johtuen tietovarastoissa olevista dokumenttien luonteesta ja käyttötarkoituksesta. Pohjoismaisessa vahinkovakuutusyhtiössä on käytössä tämä kahden tietovaraston malli. Asiakasdokumentit ovat omassa dokumenttivarastossa ja yrityksen sisäiset dokumenteista ovat omassa tietovarastossa. Molemmat varastot on rakennettu samaan tekniseen ympäristöön erillisinä dokumenttivarastoina. Molempiin varastoihin on määritelty omat metadatamallit ja dokumenttien elinkaarimallit. Vakuutusliiketoiminnassa näiden kahden dokumenttivarastojen yhteiskäyttö tulisi kyseeseen tilanteessa, jossa vakuutusetsivät tutkisivat vakuutuspetosta tai muuta väärinkäyttöepäilyä, jossa olisi mukana yhtiön henkilökuntaa.

Kuvassa 1 keskellä on dokumenttien arkistointi ja säilytys. Pitkäaikaisarkistointi on jaettu aktiiviseen arkistoon (engl. *online archive*) ja passiiviseen arkistoon (engl. *offline archive*). Aktiivisella arkistolla tarkoitetaan verkkoyhteydellä varustettua arkistoa, josta voidaan hakea tietoa käyttäen päätelaitteelta lähetettyjä kyselyjä. Passiivisessa arkistossa ei ole verkkoyhteyttä ja tietoa ei voida hakea päätelaitteelta lähettävillä kyselyillä. Vakuutusmaailmassa passiivinen arkisto on paperiarkiston ja varmistusnauhat. Aktiiviarkistoja taas ovat kaikki digitaalisia dokumentteja sisältävät verkkoyhteydelliset dokumenttivarastot.

Kuvassa 1 viimeisenä on dokumenttien lopullinen tuhoaminen dokumenttiarkistosta. Dokumentit ja niihin liittyvä metadata tuhotaan pysyvästi, kun dokumentteihin ei enää kohdistu säilytysvelvollisuutta. Liikatoiminta-alue, lait ja viranomaissäädökset määrittelevät säilytysvelvollisuuden ja -ajan. Dokumentit tuhotaan ennalta määriteltyjen sääntöjen mukaisesti automaattisesti. Vakuutusliiketoiminnassa dokumenttien tuhoamista ei ole toteutettu kovinkaan tehokkaasti, vaan pikemminkin kaikki halutaan säilyttää varmuuden varalta. Tällainen toiminta voi kuitenkin olla ristiriidassa vakuutusalaa säätelevien määräysten ja säännösten kanssa, joka velvoittaa vakuutusyhtiöitä tuhoamaan vanhentuneet dokumentit.

Seuraavissa kohdissa käydään dokumenttien elinkaaren vaiheet läpi. Tavoitteena on antaa lukijalle kokonaisvaltainen käsitys dokumenttien elinkaaresta ja sen vaiheista.

3.3 Aktiiviset ja liiketoimintakriittiset dokumentit

Kohdan 3.3 tarkoituksena on esitellä lukijalle ja määritellä dokumentit, jotka ovat aktiivisia ja liiketoiminta kriittisiä.

Usein kuulee puhuttavan kuumasta datasta tai kuumista dokumenteista. Tällä tarkoitetaan dokumentteja, joita tarvitaan päivittäisessä liiketoiminnassa eli ne ovat aktiivisia liiketoimintakriittisiä dokumentteja. Nämä dokumentit ovat iältään nuoria. Tyypillisesti niiden ikä on muutamista päivistä viikkoihin tai kuukausiin riippuen dokumenttityypistä. Niiden sisältö muuttuu ja tarkentuu sitä mukaa, kuinka liiketoimintaprosessi etenee. Dokumentteihin tietoa voivat lisätä yrityksen työtekijät tai erilaiset tietojärjestelmät, jotka käsittelevät dokumentteja liiketoimintaprosessien aikana. Dokumenttien versiointi tapahtuu yleensä tietojärjestelmien toimesta automaattisesti. Sen sijaan asiakassegmentointiin perustuvaan personointiin tarvitaan työtekijää. Dokumenttien kulkiessa liiketoimintaprosessien läpi muuttuu myös niissä oleva metadata. Minimissään muutokset koskevat dokumenttien versio tietoa tai aikaleimaa, mutta lähes aina siihen lisätään myös dokumenttien tilaa (engl. *status*) kuvaava metatietoa. Tilatietoa ovat mm. luonnos, hyväksymätön, hyväksytty ja hylätty. Tarjoukseen liittyviä tilatietoja taas ovat tarjousluonnos, lopullinen tarjous, asiakkaalle lähetetty tarjous, hylätty tarjous, vanhentunut tarjous ja hyväksytty tarjous.

Vakuutusliiketoiminnassa aktiivisia liiketoimintakriittisiä dokumentteja ovat tarjoukset, vahinkoilmoitukset, laskut, korvauspäätökset, mainokset ja erilaiset kampanjadokumentit. Jokaisella näillä dokumenttityypillä on erilaiset elinkaaret. Lyhimmillään dokumentin aktiivinen liiketoimintakriittinen vaihe voi olla vain muutamia kymmeniä minuutteja ja pisimmillään muutamia kuukausia. Vahinkoilmoituksien elinkaaret voivat vaihdella paljonkin.

Esimerkki tällaisesta lyhytkestoisesta dokumentin elinkaaresta on vahinkotapaus, jossa kuljettaja peruuttaa autonsa parkkihallissa betonipylvääseen. Kuljettaja ei saa vammoja

eikä vahingossa ole muita osapuolia. Kuljettaja tekee nettipalvelussa korvausilmoituksen tapahtuneesta ja lähettää sen vakuutusyhtiölle käsiteltäväksi. Vakuutusyhtiön tietojärjestelmän automaattinen sääntö- ja päättelykoneisto käsittelee vahinkoilmoituksen ja tekee korvauspäätöksen. Tieto korvauspäätöksestä lähetetään autonkuljettajalle tekstiviestinä. Käsittelyaika on muutamia kymmeniä minuutteja, joka on myös tehdyn vahinkoilmoituksen aktiivinen liiketoimintakriittinen aika. Päätöksen jälkeen vahinkoilmoitus siirtyy elinkaaressaan seuraavaan vaiheeseen. Vahinkoilmoitus voi liikkua myös takaisin aktiiviseksi liiketoimintakriittiseksi dokumentiksi, jos vahinkoilmoituksen tekijä on tyytymätön korvauspäätökseen ja haluaa uudelleen käsittelyn ilmoitukselleen. Uudelleen käsittelyyn päätyneen vahinkoilmoituksen metadataa muutetaan vain dokumentin tilaa kuvaavien attribuuttien osalta.

Esimerkiksi taas tarjoukset ja erilaiset mainokset pysyvät aktiivisina ja liiketoimintakriittisinä niin pitkään kunnes niiden voimassaoloaika on päättynyt. Tarjousten elinkaari vaihtelee muutamasta päivästä kuukausiin riippuen tarjouksen tyypistä ja kohteesta. Tuona aikana kaikki tarjoukseen liittyvä dokumentaatio on aktiivista ja sitä tarvitaan päivittäisessä työskentelyssä. Usein tarjouksiin liittyy myös mainoskampanja, joka on myös aktiivista koko tarjouksen keston ajan. Mainoskampanja liittyy monen tyyppisiä dokumentteja ja tiedostoja. Videoita, animaatiota ja äänitiedostoja käytetään paremman näkyvyyden saamiseksi. Nämä ovat myös aktiivista liiketoimintakriittistä materiaalia, jotka tulee käsitellä samalla tavalla kuin muutkin dokumentit. Vakuutusliiketoiminnassa asiakkaalle lähetetty mainos tarjouksesta tulkitaan sopimuksesi ja mainoksiin sovelletaan samoja käsittely- ja elinkaarisääntöjä kuin sopimuksiin.

Vakuutusliiketoiminnassa oman luokkansa muodostavat vakuutusehdot ja vakuutuskirjat, jossa on koottuna asiakkaan vakuutukset ja niihin liittyvät turvat. Vakuutusehdot ovat aktiivisia ja liiketointakriittisiä dokumentteja niissä määritellyn voimassaoloajan. Ehtoja käytetään tulkittaessa erilaisia asiakastilanteita. Vakuutusehdot ja vakuutuskirjat muuttuvat harvoin. Muutoksia ehtoihin tehdään kerran vuodessa, tai kun viranomaismääräykset tai lait muuttuvat. Vakuutuskirjaan tehdään muutoksia tarvittaessa esimerkiksi, kun asiakas haluaa muuttaa vakuutuksiaan tai turvatasojaan. Vakuutusehdot ovat vakioituja tekstikappaleita (engl. *standard text blocks*), jotka on talletettu käytössä olevaan dokumenttien luonti ja hallintavälineen tietovaraston. Näistä vakioiduista tekstikappaleista voidaan koostaa tarvittavat dokumentit automatiikalla käyttäen ennalta määriteltyjä dokumenttipohjia ja pohjiin sisältyvää dokumenttilogiikkaa. Vastaavasti asiakkaan vakuutuksien tiedot on talletettu vakuutusyhtiön asiakasjärjestelmään, josta asiakkaalle voidaan luoda ja lähettää ajantasainen vakuutuskirja. Vakuuskirjaa ja vakuutusehtoja ei siis säilytetä luotuna dokumenttina dokumenttitietovarastossa ja näin ollen niillä ei myöskään ole omaa dokumentin elinkaarta metadatoineen.

3.4 Passiiviset ja liiketoimintakriittiset dokumentit

Kohdassa 3.4 esitellään dokumentit, jotka ovat passiivisia mutta liiketoimintakriittisiä.

Passiiviset ja liiketoimintakriittiset dokumentit ovat dokumentteja, joita tarvitaan päivittäisessä liiketoiminnassa taustamateriaalina. Nämä dokumentit ovat valmiita ja ne eivät enää muutu. Dokumentteihin ei enää tehdä muutoksia automaattisesti tietojärjestelmien toimesta tai henkilökunnan toimesta. Dokumentit eivät siis enää ole luonnoksia vaan valmiita virallisia versioita dokumenteista. Päivittäisessä liiketoiminnassa näitä dokumentteja käytetään tietojen tarkistukseen, asiakkaan viimeisimpien yhteydenottojen tarkastelussa sekä taustamateriaalinen uusmyynnissä. Metadata pysyy tässä vaiheessa muuttumattomana, kunnes dokumenttiin tehdään muutoksia. Metadata muuttuu myös, kun dokumentteja ei enää tarvita päivittäisessä työskentelyssä taustamateriaalina ja ne siirtyvät elinkaarensa seuraavaan vaiheeseen.

Vakuutusliiketoiminnassa tällaisia passiivisia liiketoimintakriittisiä dokumentteja ovat esimerkiksi vahinkoilmoitukset, joista on lähetty korvauspäätös asiakkaalle, jolla on määrätty aika vaatia muutosta tehtyyn päätökseen. Toinen esimerkki ovat tarjouspyynnöt, joissa on määritelty tarjouksen voimassaoloaika, jonka aikana asiakas voi hyväksyä, hylätä tai pyytää uutta tarjousta. Korvauspäätöksen valitusajan tai tarjouksen voimassaoloajan jälkeen dokumentit ovat ohittaneet liiketoimintaprosessin viimeisen vaiheen ja ovat menettäneet arvonsa liiketoimintakriittisinä dokumentteina. Dokumentit siirretään elinkaaressaan eteenpäin lyhytaikaiseen arkistointiin. Siirron yhteydessä dokumenttien metadata muutetaan vastaamaan lyhytaikaisen arkistoinnin vaatimuksia.

3.5 Dokumenttien lyhytaikainen arkistointi

Kohdassa 3.5 käsitellään lyhytaikaista dokumenttien arkistointia sekä siihen liittyvää metadataa ja liiketoiminta-arvoa. Kohdassa käydään läpi yleisiä arkistointiin liittyviä asioita kuten tavoitteita, huomioon otettavia asioita ja hallintaa. Lopuksi annetaan käytännön esimerkkejä vakuutusliiketoiminnasta.

Kuten kohdassa 3.4 todettiin passiiviset ja ei-liiketointakriittiset dokumentit voidaan siirtää lyhytaikaiseen arkistointiin. Vaikka nämä dokumentit eivät enää ole liiketoimintakriittisiä, on niillä edelleen liiketoiminnan kannalta arvoa. Liiketoiminta saa dokumenteista tietoa asiakkastietoa, jota käyttäen voidaan palvelua kohdentaa ja täsmentää. Dokumenteista saadaan mm. tietoa asiakaan lähivakuutushistoriasta, korvauksista, vakuutusturvista ja vahinkoista. Tätä tietoa haetaan ja raportoidaan käyttäen arkistossa olevien dokumentteja ja niiden metadataa.

Dokumenttien arkistointiin on monta näkökulmaa ja määritelmää. Arkistolaitoksen asiakirjahallinnon ja arkistoimen oppikirjan alusta löytyy hyvä määritelmä. Kirjassa määritellään arkistot yhteiskunnan toimivaksi muistiksi. Tämä kuvaa arkiston perustehtävää. Yritystoiminnassa arkistot ovat yrityksen toimiva muisti, joita voidaan hyödyntää liiketoiminnan ohjauksessa ja tulevaisuuden suunnittelussa. Vakuutusliiketoiminnassa yrityksen toimiva muisti on asiakastiedot ja kaikki asiakkuuteen liittyvät dokumentit.

Erilaisten tahojen ylläpitämiä arkistoja säädellään ja hallintaa monilla säädöksillä, lailla ja asetuksilla. Arkistolaitos (2006) määrittelee Asiakirjahallinnon ja arkistoinnin oppikirjassa arkistotoimeen ja arkistointiin liittyvää sääntelyä seuraavasti:

”Arkistot liittyvät kiinteästi hallintoon, kansalaisten oikeusturvaan ja demokratian toteutumiseen (viranomaisten toiminnan valvonta, vapaa mielipiteenmuodostus), minkä vuoksi niiden muodostumista, säilyttämistä ja käyttöä ohjaa lainsäädäntö. Erityisesti tämä koskee viranomaisarkistoja, mutta tietyiltä osin myös yksityisiä arkistoja.”

Arkistolainsäädäntö on kansainvälistä, sillä kaikista maista löytyy arkistoja koskevia lakeja ja säädöksiä. Iältään nämä kansainväliset arkistolainsäädökset ovat nuoria, vaikka arkistojen on säädelty jo kauan eri maissa käyttäen moninaisia alemman tason normeja. Lisäksi Arkistolaitos määrittelee arkistojen juridisen sääntelyn seuraavasti:

”Tyypillistä arkistotoimen ja arkistojen juridiselle sääntelylle on se, että yhden peruslain (arkistolain) ohella siihen vaikuttaa muitakin lakeja, erityisesti julkisuus- ja tietosuojalainsäädäntö. Tämän lisäksi myös hallinnonalakohtaiset erityislait on välttämätöntä ottaa huomioon.”(Arkistolaitos 2006).

Tästä juridisen sääntelyn moninaisuudesta, johtuen arkistot ja arkistotoimi erottuvat muusta informaatioaloistaan omaksi erityiseksi alakseen omine erityisvaatimuksineen ja tavoitteineen.

Dokumenttiarkistojen muodostamiseen tarvitaan arkistointisuunnitelma, joka palvelee organisaatiota tai yritystä. Arkistointisuunnitelman perustehtävänä on ohjata organisaation arkistonmuodostusta määrittelemällä suunnitelmallisen dokumenttien arkistoinnin, dokumenttien rekisteröinnin, dokumenttien seulonnan ja luokittelun. Suunnitelmassa on pysyvä ja ajantasainen kuvaus organisaation tehtävistä, tehtäviin sisältyvistä käsittelyvaiheista ja käsittelyvaiheiden yhteydessä kertyvistä dokumenteista sekä niiden käsittelystä. Arkistonmuodostussuunnitelmassa asetetaan vaatimuksia dokumenttien metadatalle. Arkistointisuunnitelma ei välttämättä itsessään sisällä arkistoinnin vaatimien metadata-attribuuttien tarkkaa määrittelyä. Arkistointisuunnitelmassa määritellyt metadata vaatimukset tulee aina ottaa huomioon tehtäessä dokumenttien metadatamallia. Arkistoinsuunnitelman ja dokumenttien metadatamallilla on suora riippuvuus toisistaan (Arkistolaitos 2015 ja 2013).

Dokumenttien arkistoinnista puhutaan yleensä yhtenä kokonaisuutena, vaikka käytännössä arkistointitasoja ja -tapoja on useita. Arkistointitasot määräytyvät useimmiten tarvittavien säilytysaikojen perusteella. Säilytysajat puolestaan määräytyvät arkistoitavan materiaalin, organisaation tarpeiden ja arkistointimääräysten perusteella. Näin muodostuneet arkistointitasot puolestaan vaikuttavat siihen, mitkä tekniset talletusratkaisut sopivat kullekin arkistointitasolle. Nyrkkisääntönä voidaan pitää, että mitä vanhempi säilytettävä dokumentti on, sen halvemmalla ja hitaammalla talletusmedialla se säilytetään. Toinen nyrkkisääntö on, että asiakirjaa tulee säilyttää niin kauan kuin se on säilyttämisen

arvoinen, mutta ei pidepään kuin vaaditaan (Romppanen 2004). Käytännössä dokumenttien siirto talletusmedialta toiselle toteutetaan dokumenttien metadatassa olevien ikäattribuuttien sisältämän tiedon perusteella automaattisesti. Dokumenttien siirtyessä talletusmediasta toiseen muuttuu myös dokumenttien metadata vastaavasti. Tärkeimpiä metadata-attribuutteja arkistoinnissa ovat dokumentin ikää ja omistajuutta kuvaavat attribuutit. Näiden attribuuttien avulla voidaan dokumentit siirtää arkistointitasolta toiselle. Lisäksi voidaan käyttää yrityksen toimi-alaan liittyviä attribuutteja (esimerkiksi yksityisasiakas tai yritysasiakas), yksityisyydensuojaan liittyviä attribuutteja tai lainsäädäntöön liittyviä attribuutteja.

Dokumenttien metadata pysyy muuttumattomana niin kauan kun ne ovat jollakin arkistointitasolla. Dokumenttien siirtyessä toiselle arkistointitasolle niiden metadata muuttuu uuden arkistointitason mukaiseksi. Arkistoitujen dokumenttien metadata voi myös muuttua, jos ne siirretään elinkaaressaan taaksepäin johtuen liiketoiminnan tarpeista. Tällöin dokumentit siirretään arkistosta liiketoiminnan aktiivisesti käyttämien dokumenttien joukkoon. Siirrettävät dokumentit ovat saaneet uudelleen liiketoiminta-arvoa ja niitä tarvitaan yrityksen liiketoiminnassa. Vakuutusliiketoiminnassa tällainen tilanne syntyy, jos asiakas tai yritys käynnistää oikeuskäsittelyn, jossa tarvitaan todistusaineistona koko vakuutushistoriaa. Kuten edellä olevasta esimerkistä ja kuvasta 1 huomataan dokumentit voivat siirtyä elinkaaressaan molempiin suuntiin, eteenpäin ja taaksepäin, ja dokumenttien metadatan tulee muuttua sen mukaisesti.

Arkistojen käyttötapa ja rooli on muuttunut. Aikaisemmin arkistot olivat paperiarkistoja, joiden tietoihin käsiksi pääsivät ainoastaan arkistonhoitajat. Nykyisin tällaisia arkistoja on, mutta tässä työssä arkistoilla tarkoitetaan sähköisiä arkistoja, jossa dokumentit säilytetään digitaalisessa muodossa. Sähköisiin arkistoihin asiakkaille tarjottavien itsepalvelukanavien myötä arkistojen rooli ja käyttö on muuttunut. Arkistot on avattu internetin välityksellä ulkoisille käyttäjille eli yrityksen palveluja käyttäville asiakkaille. Täysin asiakkaalta suljettuja arkistoja on enää vähän. Vakuutusliiketoiminnassa suljettuja arkistoja ovat esimerkiksi paperiarkistot ja vakuutusetsivien asiakasarkistot. Sen sijaan vakuutusyhtiön keräämät asiakkaiden terveystiedot eivät ole täysin suljettuja vaan osittain suljettuja. Tämä tarkoittaa, että vain asiakkaalla on oikeus ja pääsy tunnistuksen jälkeen omiin terveystietoihin ja että vain vakuutusyhtiön nimetyillä henkilöillä on oikeus käsitellä tätä tietoa.

Lyhyt ja pitkäaikaisen arkistoinnin rajaa ei voida tarkasti määritellä, vaan se on kuin veteen piirretty viiva. Rajanveto riippuu liiketoiminta-alasta ja organisaation tarpeista. Vakuutusliiketoiminnassa dokumenttien lyhytaikainen arkistointi tarkoittaa noin 1-10 vuoden säilytysaikaa. Säilytysaika riippuu kuitenkin dokumenttityypistä. Loppuun käsitellyn yksinkertaisen vahinkoilmoituksen, jota ei ole riitautettu, säilytys lyhytaikaisessa arkistoinnissa voi olla vain kuukausia tai korkeistaan pari vuotta. Sen sijaan mutkikkaan tai riitautetun vahinkoilmoituksen lyhytaikainen arkistointiaika voi olla vähintäänkin vuosia.

Vakuutusliiketoiminnassa asiakkaille lähetetyt mainokset, joissa mainostetaan vakuutuksia ja niihin liittyviä tarjouksia, tulee arkistoida. Syynä tähän on, että mainokset tulkitaan yrityksen ja asiakkaan välisiksi sopimuksiksi. Mainos on liiketoimintakriittinen mainoskampanjan voimassaoloajan, jonka jälkeen se voidaan arkistoida.

Yhteenvetona voidaan todeta, että se kuinka pitkään dokumentti ovat lyhytaikaisessa arkistossa, riippuu dokumenttityypin lisäksi myös liiketoiminta-alueesta ja siihen liittyvästä lainsäädännöstä sekä käytännöistä.

Yrityksen toiminnassa tuotetaan paljon myös yrityksen sisäiseen käyttöön tarkoitettuja dokumentteja. Näitä ovat mm. erilaiset muistiot, suunnitelmat, esitykset, projektidokumentaatio ja pöytäkirjat. Näillä dokumenteilla ei ole liiketoiminta-arvoa tai arvo on pieni. Kuitenkaan yritys ei voisi toimia ilman näitä dokumentteja. Niillä on siis suuri yrityksen sisäinen arvo ja merkitys. Tätä merkitystä ei yleensä arvoteta tai tunneta, mistä johtuen dokumenteille ei määritellä metadatamallia tai huolehdita metadatan oikeellisuudesta. Yrityksen sisäiset dokumentit tulisi kuitenkin käsitellä samoin kuin yrityksen liiketoimintaan liittyvät asiakasdokumentit. Yrityksen sisäiset dokumentit talletetaan useimmiten heti niiden syntyessä arkistoon, joka voi olla lyhytaikainen arkisto tai pitkäaikaisarkisto. Talletuksen yhteydessä luodaan ja talletetaan tarvittava metadata. Arkistoituja yrityksen sisäisiä dokumentteja käsitellään sääntöpohjaisella automatiikalla samaan tapaan kuin asiakasdokumentteja.

3.6 Dokumenttien pitkäaikainen arkistointi

Kohdassa 3.6 tarkastellaan lyhyen ja pitkäaikaisen arkistoinnin eroja sekä sitä millainen on pitkäaikaisarkistointi.

Kuten edellisessä kohdassa 3.5 todettiin, raja pitkäaikaisen ja lyhytaikaisen arkisoinnin välillä ei ole selkeä. Dokumentit voivat pysyä saman teknisen ympäristön sisällä ja ne vain siirretään eri talletusmedialle metadatasta määriteltyjen sääntöjen mukaisesti. Dokumenttien talletukseen käytetty tiedostomuotokaan ei välttämättä muutu. Tarvittaessa dokumenttien tiedostomuoto voidaan muuntaa paremmin pitkäaikaisarkistoitiin soveltuvaan muotoon niiden siirtyessä lyhytaikaisesta arkistoinnista pitkäaikaisarkistointiin. Käytännössä usein dokumentit talletetaan heti niiden syntyessä arkistokelpoiseen tiedostomuotoon. Näin toimien vältytään tiedostomuotojen ja metadatan konversioilta dokumenttien siirtyessä elinkaaren vaiheesta toiseen. Arkistointikelpoisesta talletusmuodosta voidaan tarvittaessa muodostaa (engl. *rendering*) erilaisia kulloiseenkin palvelukanavaan sopivin tiedostomuoto esitettäväksi.

Dokumenttien siirron perusteena pitkäaikaisarkistoon on lähes poikkeuksetta dokumenttien metadatassa oleva ikä ja arkiston ylläpitosuunnitelmassa määritellyt säilytysaikaluokat. Säilytysluokat taas ovat saatu asetuksista, lainsäädännöistä, yrityksen sisäisestäluoki-

tuksesta tai yleisesti käytössä olevista käytännöistä. Vakuutusliiketoiminnassa säilytysluokat ovat 0 - 5 vuotta, 6 - 20 vuotta, 21 - 100 vuotta. Nämä arkistointiportaat eivät ole tarkkoja, vaan antavat yleiskäsityksen vakuutusliiketoiminnan arkistointiajoista ja -säilytysluokista. Esimerkiksi pohjoismaisesti toimivan vakuutusyhtiön säilytysajat vaihtelevat kunkin maan säädösten mukaisesti. Tämä vaihtelu tulee ottaa huomioon dokumenttien metadatassa, jonne tulee arkistointiajan lisäksi tallettaa tieto minkä maan dokumentista on kyse. Maakoodille on siis varattava oma attribuutti dokumenttien metadata mallissa. Maakoodin sisältävää attribuutti tulee myös käyttää automatiikan käyttämissä säännöissä.

Joissakin tapauksissa dokumentti voidaan myös viedä suoraan pitkäaikaisarkistoon. Tällaisia dokumentteja voivat olla esimerkiksi lääkärin kirjoittama diagnoosi, röntgenkuvat tai valokuvat.

Pitkäaikaisarkistoinnissa dokumenttien metadata ei enää muutu. Metadataa voitaisiin kyllä käyttää rekisteröimään, kuinka monta kertaa ja milloin dokumenttia on haettu arkistosta. Tämä auttaisi saamaan käsitystä siitä, mitä dokumenttityyppejä haetaan pitkäaikaisarkistosta ja kuinka usein. Tätä tietoa taas voitaisiin käyttää arkistoinnin käytön seurantaan ja hallintaan.

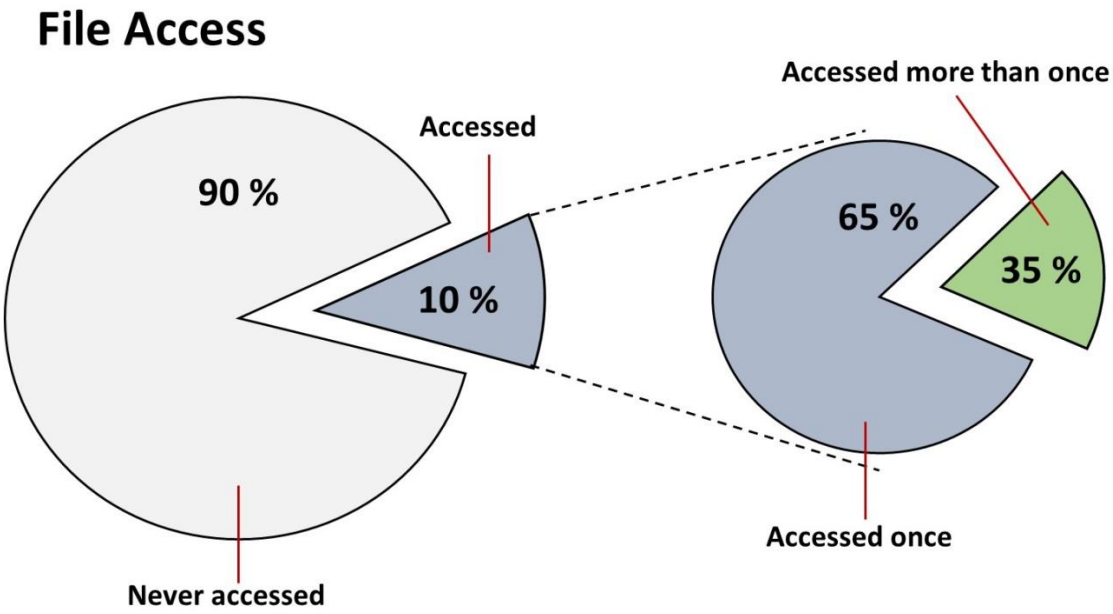
Vakuutusliiketoiminnassa dokumenteille on määritelty useampia säilytysaikoja, joita sovelletaan arkistointiin. Määritellyt portaat vaihtelevat liiketoiminta-aloittain ja maittain. Pohjoismaisessa vakuutusyhtiöissä käytössä olevat arkistointiajat ovat 5 - 10 vuotta, 10 - 20 vuotta, 20 - 50 vuotta ja yli 50 vuotta. Joillekin dokumenteille on lisäksi määritelty ikuinen säilytys, joka on käytännössä tulkattu vakuutusliiketoiminnassa 100 vuodeksi. Pisimmät säilytysajat ovat henkilövahinkoihin liittyvillä dokumenteilla, hyvin vanhoilla ns, ”ikuisilla vakuutuksilla”, vakuutusliiketoimintaan liittyvillä esimerkkitapauksilla sekä tapauksilla, joissa katsotaan olevan riski oikeuskäsittelyyn pitkän ajan kuluessa. Joissakin erikoistapauksissa tai vakuutusyhtiön liiketoiminnan päätöksellä voidaan dokumentteja säilyttää kunnes niiden hävityksestä erikseen sovitaan. Tällöin näiden dokumenttien käytännön arkistointiaika on ikuisuus, koska metadataan ei ole merkitty arkistointiaikaa.

Vakuutusyhtiöiden asiakkailla on nykyään mahdollisuus nähdä koko vakuushistoria netissä olevien palveluiden kautta. Tämä on aiheuttanut sen, että myös pitkäaikaisarkistot on avattu asiakkaille. Yksityisasiakkaat eivät ainakaan vielä ole kovin paljon käyttäneet tätä mahdollisuutta. Yritysasiakkaat ovat olleet enemmän kiinnostuneita tästä palvelusta, mutta yrityksiltä tulevien kyselyiden määrä ei vielä ole suuri. Tulevaisuudessa tämän palvelun käytön ennustetaan kasvavan. Tämä tulee muuttamaan arkistojen käyttöä entisestään sekä niiden roolia asiakkuuden hallinnassa. Asiakkaille luodaan uusia tapoja ja mahdollisuuksia nähdä omia tietojaan vakuutusyhtiöiden arkistoista sekä vaatia tietojen poistamista tai muuttamista. Tämä asettaa uusia vaatimuksia tietojen oikeellisuuteen, metadatan oikeellisuuteen ja tietoturvaan. Vakuutusyhtiöille tulee entistä tärkeämmäksi huolehtia arkistojensa eheydestä ja asiakkaiden tietojen ajan tasaisuudesta. Jo nyt asiakkaat ovat huomanneet vakuutushistoriassaan puutteita ja virheitä, joita vakuutusyhtiöt ovat

korjanneet. Toisaalta asiakkailta tuleva uudenlainen arkistojen käyttö antaa uusia mahdollisuuksia vakuutusyhtiöille. Ne saavat uutta tietoa asiakkaasta heidän kyselyistään ja vierailuistaan arkistoissa. Mitä dokumentteja ja kuinka usein yksittäinen asiakas tai asiakasryhmä hakee arkistoituja dokumentteja sekä kuinka usein nämä kyselyt johtavat jatkotoimenpiteisiin tai tietojen korjauksiin. Tämä on arvokasta tietoa vakuutusyhtiöille. Näiden tietojen perusteella vakuutusyhtiöt voivat parantaa ja kehittää asiakaspalveluaan ja tuotteitaan. Lisäksi asiakkaat auttavat parantamaan yhtiöiden arkistojen laatua korjauttamalla niissä olevia virheitä tai puutteellisia tietoja.

3.7 Dokumenttien pysyvä tuhoaminen

Kohdan 3.7 motivaationa on antaa lukijalle käsitys dokumenttien pysyvään tuhoamiseen liittyvistä tavoitteista, ongelmista ja vaatimuksista. Kohdassa annetaan yksi malli miten dokumenttien pysyvä tuhoaminen voidaan käytännössä toteuttaa.



Kuva 2. *Kuvas esittää talletettujen tiedostojen käyttöä niiden luonnin jälkeen.*

Kuvasta 2 nähdään, että talletetuista ja arkistoiduista tiedostoista vain kymmentä prosenttia haetaan ja käytetään niiden ensimmäisen talletuksen jälkeen. Tuosta kymmenestä prosentista vain pientä osaa (35 %) tiedostoja haetaan ja käytetään enemmän kuin kerran. Tätä tulosta suoraan soveltaen Pohjoismaissa toimivan vakuutusyhtiön 75 000 000 dokumentin arkistoon saadaan tulokseksi, että vain 2 625 000 dokumenttia on käytetty talletuksen jälkeen enemmän kuin kerran. Tämä tulos on oikein, jos oletetaan, että talletettu tiedosto vastaa yhtä talletettua dokumenttia, josta ei ole kopioita tai useita versioita. Näin ei tilanne kuitenkaan ole käytännössä. Todellisuudesta yhdestä dokumentista on talletettuna useita versioita ja jopa kopioita samassa tai toisissa dokumenttivarastoissa. Dokumenttista talletettujen versioiden määrää voidaan rajoittaa ja hallita, jos liiketoiminnan

kanssa on sovittu, montako vanhaa versiota uusimman lisäksi dokumentista säilytetään. Esimerkissä lopputulokseksi saatu määrä useammin kuin kerran haetuista tiedostoista, voi sisältää dokumenttien eri versioita ja kopioita. Toisin sanoen todellinen dokumenttien määrä voi olla esimerkkilaskelman lopputulosta pienempi. Yhteenvetona voidaan todeta, että dokumenttivarastoissa säilytetään paljon dokumentteja, jotka voidaan hävittää turvallisesti ilman, että tietoa asiakkaista tai yrityksen sisäisestä toiminnasta häviää lopullisesti.

Dokumenttien elinkaaren päätepiste on dokumenttien pysyvä tuhoaminen (engl. *permanent disposal*). Dokumenttien tuhoaminen pysyvästi voidaan tehdä metadatasta johdettujen säätöjen perusteella täysin automaattisesti. Automaattinen tuhoaminen ei ole teknisesti vaikea toteuttaa, mutta vaikean asiasta tekee ihminen ja oikeiden sääntöjen rakentaminen.

Ihmiselle jonkin tuhoaminen automaattikan avulla vaikuttaa epäilyttävältä ja epävarmalta. Mieleen tulee paljon kysymyksiä. Miten voin olla varma, että säännöt toimivat oikein? Mitä tapahtuu, jos automatiikka tuhoaa liikaa tai liian vähän dokumentteja? Miten voin olla varma, että en tule enää tulevaisuudessa tarvitsemaan tuhottuja dokumentteja? Kuka valvoo ja miten automatiikkaa ja sen toimivuutta? Onko dokumenttien metadata riittävän laadukasta ja kattavaa oikeiden sääntöjen tekemiseen? Kaikkiin näihin ja moniin muihin kysymyksiin tulee vastata ja riskit minimoida ennen kuin dokumentti voidaan tuhota sääntöpohjaisella automatiikalla. Usein siirtyminen heti kerta rysäytyksellä täysautomatiikalla tehtävään dokumenttien tuhoamiseen tuntuu ja on liian suuri ponnistus organisaatioille ja yrityksille. Tyypillinen käytännön puolitien ratkaisu on tehdä dokumenttien tuhoaminen puoliautomaattisesti. Tässä ratkaisussa dokumenttien standardoidusta metadatasta johdetut säännöt poimivat tuhottavat dokumentit sääntöpohjaisella automatiikalla, mutta niitä ei vielä tässä vaiheessa tuhota. Tuhottavaksi poimituista dokumenteista lähetetään nimetyille vastuuhenkilöille tieto yhteenvedon muodossa. Vastuuhenkilöiden tehtävänä on tarkistaa ja hyväksyä poiminnan tulos ja vahvistaa tuhoaminen tietyn määritellyn ajan kuluessa. Vahvistuksen saatuaan järjestelmä tuhoaa tuhottavaksi poimitut dokumentit. Jos taas vastuuhenkilöt eivät hyväksy ajoissa poiminnan tulosta, niin dokumentteja ei tuhota ja poiminnantulos hylätään ja yritetään uudelleen seuraavassa poimintaajossa.

Riippuen dokumenttien kokonaismäärästä ja poimintasäännöistä tulokseksi voidaan saada hyvinkin suuri joukko dokumentteja. Esimerkiksi Pohjoismaisessa vakuutusyhtiössä on yhdessä dokumenttivarastossa (engl. *document repository*) 75 000 000 liiketoiminnan asiakasdokumenttia. Tästä dokumenttimassasta dokumenttien metadataan perustuvalla automatiikalla poimitaan 0,05 % tuhottavaksi. Lopputuloksena saadaan 3 750 000 dokumenttia, jotka lähetetään vastuuhenkilöille tarkistettavaksi. Selvää on, että tällaisen massan läpikäynti ihmisen toimesta säännöllisin väliajoin vaatii paljon aikaa ja on käytännössä mahdotonta. Tästä voidaan päätellä, että ihmisen tekemällä tarkistuksella ei luoda lisäarvoa tai luotettavuutta poiminnan tulokseen tai dokumenttien tuhoamiseen.

Manuaalitarkistus ainoastaan antaa tarkastajalle tunteen kontrollista. Ihmisen tekemän tarkistuksen sijaan olisi järkevänpää panostaa sääntöjen testaukseen, monitorointiin sekä kehitykseen.

Ihmisen kokemaa epävarmuutta automaattista dokumenttien tuhoamista kohtaan voidaan ehkäistä ja poistaa tekemällä varmuuskopiot tuhottavasta dokumenttijoukosta. Varmuuskopiot voidaan ottaa esimerkiksi nauhalle, jonka kustannukset ovat pienet. Tärkeätä on, että näille varmuuskopiolle määritellään vanhenemisaika, joka jälkeen kopiot hävitetään. Säilytysaika ei saa olla pitkä, koska hävitetyillä dokumenteilla ei ole yrityksen liiketoiminnan kannalta mitään merkitystä ja näin ollen ne ovat saavuttaneet elinkaarensa päätepisteen. Suositeltava säilytysaika on kuukaudesta vuoteen, mutta lyhyempikin säilytysaika on mahdollinen, jos se yrityksen liiketoiminnalle sopii. Yleisesti voidaan todeta, että mitä nopeammin nämä varmistusnauhoilla olevat tiedot saadaan tuhottua niin sen parempi. Pidemmällä tähtäimellä tulisi pyrkiä täysin automaattiseen sääntöpohjaiseen dokumenttien tuhoamiseen, vaikka puoliautomaattinen tuntuisi toimivan hyvin. Ihmisen tekemä näennäinen tuhottavien dokumenttien tarkastus on turha ja jo kertaalleen tuhottujen dokumenttien nauhavarmistusten säilyttäminen on yritykselle turha operatiivinen kustannus.

Vakuutusyhtiöt ovat perinteisesti erittäin hyviä tuottamaan ja säilyttämään dokumentteja. Tämä johtuu liiketoiminnan luonteesta, joka perustuu erilaisiin dokumentteihin kuten vakuutusehtoihin, vahinkoilmoituksiin, terveysselvityksiin, lääkärinlausuntoihin jne. Sen sijaan vakuutusyhtiöt ovat huonoja tuhoamaan vanhoja turhaksi käyneitä elinkaarensa päätepisteeseen ehtineitä dokumentteja, vaikka lainsäädäntö ja erilaiset asetukset velvoittavat yrityksiä tuhomaan vanhaa dokumentaatiota. Yritykset ovat varovaisia tuhomaan dokumentteja, koska usein dokumenteilla ei ole määritelty omistajaa, joka voisi määritellä dokumenteille liiketoiminta-arvon ja sen elinkaaren. Omistajan löytyminen jälkikäteen dokumenteille on vaikeaa ja usein mahdotonta, sillä omistajanvastuu tuntuu liian suurelta. Pelätään vääriä päätöksiä tai epäillään oman tietämyksen ja ammattitaidon suuruutta. Juuri puuttuva dokumenttien omistajuus on suurin yksittäinen syy, miksi vakuutusyhtiöissä vanhoja elinkaarensa päässä olevia dokumentteja ei saada tuhottua.

Vakuutusliiketoiminnassa asiakasdokumentaatiota säädellään monella tavalla ja säännöt ovat selkeitä. Nämä säännöt ja niiden tarvitsemat attribuutit on melko helppo laittaa metadatamalliin ja sääntöihin, joiden perusteella poimitaan dokumenttivarastosta tuhottavat dokumentit. Tuhottavalla materiaalilla ei myöskään ole liiketoiminta-arvoa, mutta tuhoamista ei suoriteta ilman dokumenttien omistajan lupaa. Omistajuus ja omistaja ovat vahvassa roolissa vakuutusliiketoiminnan dokumenttien hallinnassa. Yleistäen voidaan sanoa, että jos dokumenttien omistajuutta ei ole selkeästi ja yksiselitteisesti määritelty, dokumentteja ei saada pysyvästi tuhottua.

Kirjoittajan mielestä toimivin ratkaisu on muodostaa tiimi, jonka tehtävänä on huolehtia yrityksen tarpeettomien dokumenttien tuhoamisesta turvallisesti ja tehokkaasti. Tiimin

asettaa yrityksen johto ja tiimillä on omistajuus kaikkeen yrityksen dokumentaatioon. Tiimissä tulee olla edustajia ainakin IT-organisaatiosta, tietoturvasta, liiketoiminnasta ja yrityksen johtoryhmästä. Tiimin koko riippuu yrityksen koosta, mutta se ei saa olla kovin suuri. Suositus tiimin kooksi on 4 - 10 jäsentä. Tiimi toimii itsenäisesti yhteistyössä liiketoiminnan ja IT:n kanssa. Tiimi on mukana määrittämässä metadata standardia ja siitä johdettuja sääntöjä automaattiseen dokumenttien elinkaaren hallintaan. Tiimi on mukana metadata ja sääntöjen implementoinnissa ja testauksessa. Tuotannon alettua tiimin tehtävänä on monitoroida ja valvoa, että säännöt toimivat sekä edelleen kehittää ja parantaa sääntöpohjaista automatiikkaa.

Yritykset, jotka ovat ottaneet tämän mallin käyttöön, ovat erittäin tyytyväisiä. Syynä ja perusteluna tämän mallin käyttöön on, että kaikki muut yritykset määritellä yrityksen dokumenteille omistajaa tai omistajuutta ovat epäonnistuneet. Kirjoittajan tietojen mukaan tämä malli ei kuitenkaan ole kovin yleisessä käytössä, vaikka mallin edut ovat kiistattomia. Ehkä mallin laajemman käytön esteenä on se, että malli vaatii yrityksen koko organisaation sitoutumista siihen sekä luottamusta vastuutiimin ammattitaitoon ja osaamiseen. Tällainen sitoutuminen saadaan aikaan, kun yrityksen ylin johto aktiivisesti sitoutuu ajamaan mallia yrityksen sisällä.

4. METADATAN HYÖDYNTÄMINEN LIIKETOIMINNASSA

Tässä luvussa käsitellään dokumentteihin liittyvän määrämuotoisen ja yhtenäisen metadatan hyödyntämistä liiketoiminnan apuvälineen. Tämä tapahtuu tutustuttamalla lukija erilaisiin liiketoiminnan näkökulmiin metadatan käytöstä. Lukijalle annetaan myös käytännön esimerkkejä vakuutusliiketoiminnan tavasta käyttää metadataa. Tavoitteena on antaa uusia ideoita ja selvittää lukijalle, miten monipuolisesti dokumenttien metadataa voidaan käyttää.

4.1 Metadatan hyödyntäminen vakuutusliiketoiminnassa

Kohdassa 4.1 käsitellään miten hyvin suunniteltua ja organosoitua metadataa voidaan käyttää eri tavoin hyväksi. Kohdissa 4.2-4.7 kerrotaan yleisesti miten kussakin tapauksessa dokumentteihin liittyvää metadataa voidaan hyödyntää.

Vakuutus on menettely tai toimintapa, jolla tietyn määritellyn riskin alaiset henkilöt tai yritykset sopivat mahdollisten vahinkojen ja siitä aiheutuvien kustannusten tasaamiseen erikoistuneen yrityksen kanssa siitä, että yritys korvaa aiheutuneet vahingot riskin toteutuessa. Näin luotua sopimusta nimitetään vakuutukseksi. Vakuutusta voidaan tarkastella myös liiketalouden näkökulmasta. Tällöin vakuutus voidaan määritellä keinoksi, jolla sattumanvaraiset ennalta aavistamattomasti tapahtuvasta vahingosta aiheutuvat kustannukset voidaan muuntaa vuosittain jatkuvaksi vuosikustannukseksi. Vakuutuksella on ominaisuuksia, joita ovat:

- Sattumanvaraisuus. Vakuutuksen kohteena olevan riskin toteutuminen riippuu ennalta arvaamattomasta ja epävarmasta tapahtumasta.
- Vahingonmahdollisuus, jolle on vakuutuksessa ennalta sovittu vahingonkorvaus.
- Vakuutusmaksun ja vakuutetun riskin tulee olla tasapainossa ottaen huomioon vakuutukseen liittyvät hoitokustannukset. Toisin sanoen vakuutusyhtiölle tulee jäädä voittoa mahdollisten vakuutuskorvausten ja hoitokustannusten jälkeen.
- Yleensä vakuutettava riski on samankaltainen suurella joukolla vakuutuksenottajia, jolloin riski voidaan tasata tuon suuren joukon kesken.
- Vakuutuksenantaja on aina erillinen vakuutuksenottajasta riippumaton taho.

Edellä olevia vakuutuksen ominaisuuksia ei voida suoraan käyttää dokumenttien metamallissa attribuutteina. Sen sijaan erilaista analytiikkaan käyttäen voidaan esimerkiksi tulipalojen vahinkoilmoituksista saada selville samakaltaisuuksia, toistuvia olosuhteita tai vuorokauden aikoja. Näistä tiedoista tehtäviä ennusteita tulipalojen esiintymiselle, voidaan käyttää palovakuutusten hinnoittelussa, tai tehdä asiakasprofiililointii perustuvaa

hinnoittelua. Mahdollisuuksia on paljon, kun analytiikalla saatu tieto yhdistetään dokumenttien metadataan ja muualta saatavaan massatietoon (engl. *big data*), jota saadaan suuria määriä nopeasti jäsentymättömänä (Salo 2015).

Vakuutusliiketoiminta on aina perustunut jonkin tyyppisiin vakioituihin dokumentteihin ja sopimuksiin. Näissä osapuolet sopivat vakuutetusta asiasta tai omaisuudesta, vakuutusehdoista, voimassaoloajasta, turvan hinnasta, korvauksista vakuutetun riskin toteutuessa sekä osapuolten velvollisuuksista. Se mitä vakuutetaan, on määritelty dokumentissa selkeästi siten, että vakuutuksen antaja ja sen ottaja sen ymmärtävät ja hyväksyvät. Ehkä vaikein ja monimutkaisin dokumenteista on vakuutusehdot, jossa kuvataan tarkasti kaikki vakuutukseen ja vakuutusturvaan liittyvät asiat. Vakuutusehdoissa kuvataan mm. vakuutuksen sisältö, vakuutuksenottaja ja -antaja, edunsaajat, vakuutussumma, vakuutussumman jakaminen, korvauksen hakeminen, vakuutussumman maksaminen, vakuutuksen voimaantulo, vakuutuksen voimassaoloaika ja päättyminen sekä muita määräyksiä. Yleisesti voidaan sanoa, että kaikki vakuutusliiketoimintaan liittyvät dokumentit ovat rakenteellisia tai semi-rakenteellisia dokumentteja, joille voidaan määritellä standardoitu metadatamalli ja tyypitys. Rakenteellisia dokumentteja ovat esimerkiksi lomakkeet, jotka on toteutettu html-kielellä ja valintalistoilla tai automatiikalla toteutetut dokumentit, joissa käytetään ennalta määriteltyä sijoittelua (engl. *layout, document logic*) ja vakioituja tekstikappaleita tietokannasta. Esimerkki semi-rakenteellisista dokumenteista ovat automatiikalla tehtävät dokumentit, joissa käytetään ennalta määriteltyä sijoittelua ja vakioituja tekstikappaleita vain osassa dokumentin sisältöä, ja jotka lopuksi viimeistellään yrityksen henkilökunnan toimesta kohderyhmälle sopivaksi. Kohderyhmä voi olla esimerkiksi kaikki Ford-merkkisen auton omistajat.

Seuraavissa kappalaissa annetaan lukijalle tietoa ja näkökulmia dokumenttivarastojen dokumenttien metadatan hyödyntämisen mahdollisuuksista vakuutusliiketoiminnassa.

4.2 Raportointi

Kohdan 4.2 tarkoituksena on antaa lukijalle käsitys dokumenttien metadatalla toteutetusta raportoinnista. Kohdassa kerrotaan vakuutusliiketoimintaan liittyvistä erilaisista dokumenttivarastoista ja niiden sisältämistä dokumenttimääristä. Lisäksi annetaan esimerkkejä näistä varastoista saatavista raporteista.

Dokumenttivarastot sisältävät suuren määrän dokumentteja. Dokumentteja varastossa voi olla useita kymmeniä tai satoja miljoonia kappaleita. Kappalemäärä kertoo, kuinka monta dokumenttia varastossa on, mutta se ei kerro, ovatko kaikki dokumentit yksittäisiä uniikkeja dokumentteja. Käytännössä dokumenttivarastossa on sama dokumentti useaan kertaan talletettuna eri nimillä ja yksittäisestä dokumentista on talletettuna useita versioita. Edellä olevasta johtuen uniikkien dokumenttien kokkonaismäärä dokumenttivarastossa on pienempi kuin dokumenttien yhteenlaskettu kappalemäärä. Kuvassa 3 on

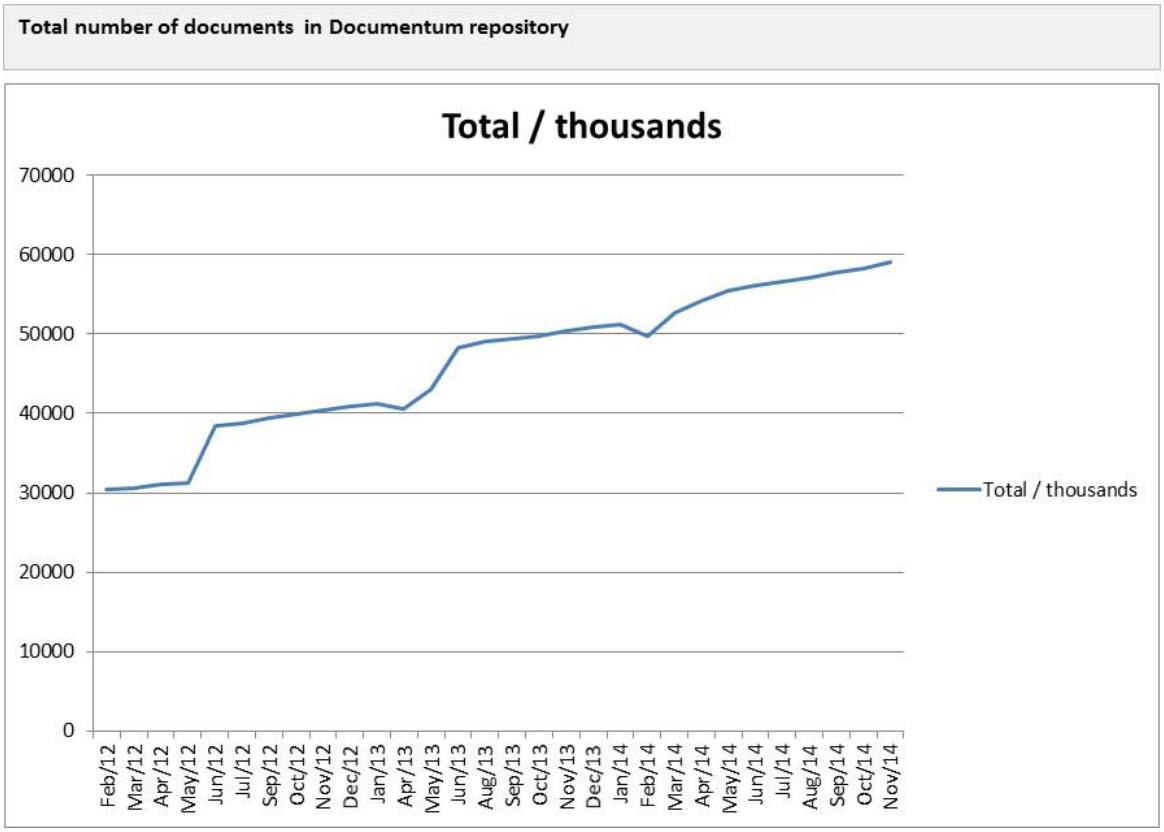
esitettynä yhden Pohjoismaissa toimivan vakuutusyhtiön yhden tietovaraston dokumenttien kokonaiskappalemäärän kehitys kolmen vuoden ajalta. Tämä on vain yksi vakuutusyhtiön useista dokumenttivarastoista. Tämä dokumenttivarasto sisältää asiakkaalle lähetettyjä ja asiakkaalta saatuja dokumentteja kuten vahinkoilmoituksia, kuvia, lääkärinlausuntoja ja sähköposteja. Dokumentit ovat useassa tiedostoformaatissa riippuen lähteestä esimerkiksi tiff-muotisia kuvia, PDF- ja DOC-muotoisia dokumentteja. Tähän dokumenttivarastoon voidaan myös tallentaa ääni- ja videotiedostoja, mutta niitä on toistaiseksi vähän varastoituna.

Taulukko 1. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön yhden dokumenttivaraston dokumenttimääristä ja kasvusta.*

Month / Year	Total number of documents / thousands
February / 2012	30 414
March / 2012	30644
April / 2012	31005
May / 2012	31323
June / 2012	38504
July / 2012	38756
September / 2012	39479
October / 2012	39878
November / 2012	40339
December / 2012	40888
January / 2013	41232
April / 2013	40495
May / 2013	43065
June / 2013	48328
August / 2013	49001
September / 2013	49391
October / 2013	49810
November / 2013	50395
December / 2013	50882
January / 2014	51272
February / 2014	49667
March / 2014	52688
April / 2014	54157
May / 2014	55492
June / 2014	56065
July / 2014	56602
August / 2014	57144
September / 2014	57824
October / 2014	58309
November / 2014	58989

Taulukon 1 ensimmäisessä sarakkeessa Month / Year on tieto kuukaudesta (engl. *month*) ja vuodesta (engl. *year*), jolta sarakkeessa Total number of documents dokumenttien kappalemäärä on laskettu. Dokumenttien kappalemäärät on taulukossa esitetty

tuhansina. Esimerkiksi September / 2013 tarkoittaa 2013 syyskuuta, jolloin dokumenttien kappalemäärä dokumenttivarastossa oli 49 391 000 (49 391 tuhatta). Kuvassa 3 taulukon 1 tiedot on esitetty graafisessa muodossa.



Kuva 3. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön EMC Documemtum infrastruktuurissa olevan dokumenttivaraston dokumenttien kokonaismäärän kehitys.*

Kuvasta 3 nähdään, että tässä varastossa on hieman alle 60 000 000 dokumenttia marraskuussa 2014. Tietovarasto näyttää kasvavan tasaisesti vuosittain. Käyrässä olevat jyrkät hypyt ylöspäin dokumenttien kokonaiskappalemäärissä johtuvat dokumenttivarastojen yhdistämisistä. Yhdistämisen yhteydessä suljettavan dokumenttivaraston kaikki dokumentit on siirretty toiseen varastoon kertaluonteisella siirrolla. Yleisesti voidaan sanoa, että tietovaraston koolle ylärajan määrittäminen on hankalaa, koska se riippuu käytettävästä teknisestä ratkaisusta ja valituista tuotteista. Yleisohjeena voidaan pitää, että dokumenttien maksimimäärän tietovarastossa määrää valitun ratkaisun tekniset ominaisuudet, dokumenttivarastolle asetetut vaatimukset kuten vasteaikavaatimukset ja tietoturvavaatimukset sekä liiketoimintavaatimukset. Esimerkiksi asiakaspalvelutilanteessa on erittäin tärkeätä, että asiakkaaseen liittyvät dokumentit saadaan haettua dokumenttivarastosta nopeasti ilman kohtuuttomia hakuaikoja, jotta voidaan tarjota laadukasta asiakaspalvelua. Hyvin isosta dokumenttivarastosta hakuaika voi olla liian pitkä asiakaspalvelun tarpeisiin, jolloin iso dokumenttivarasto tulee jakaa useampaa pienempään dokumenttivarastoon.

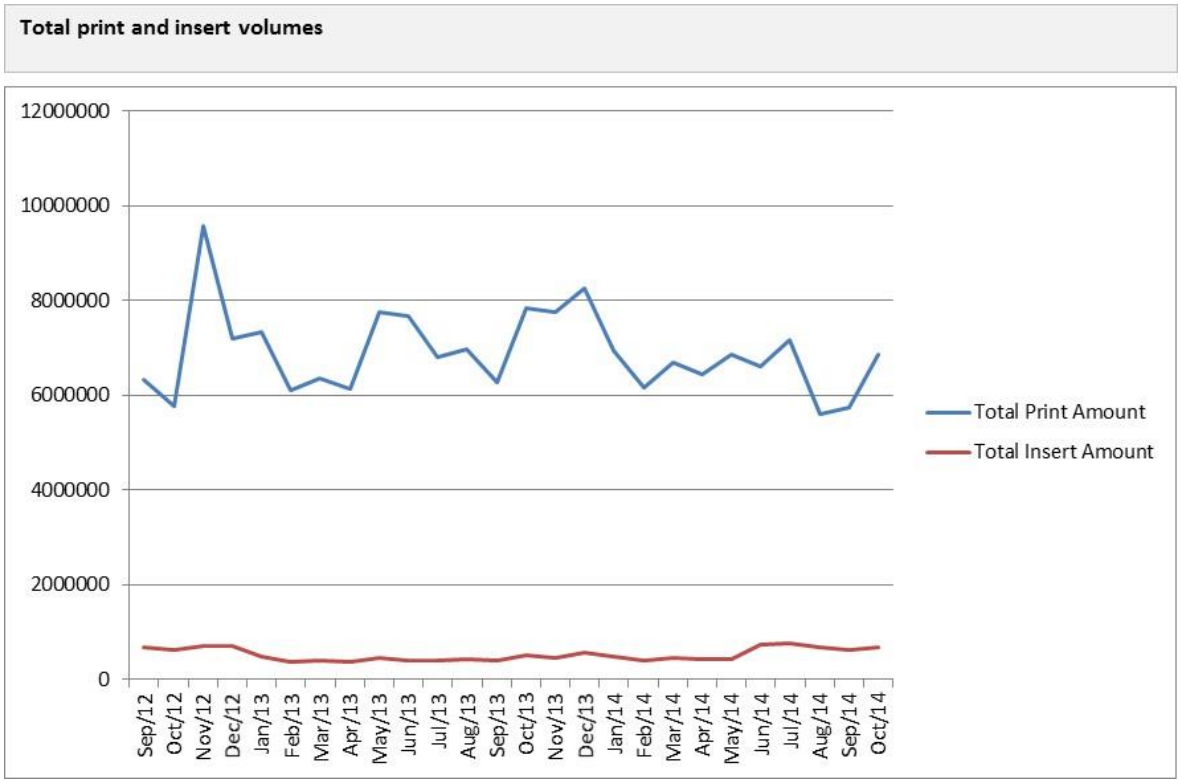
Vakuutusyhtiöt pyrkivät pankkien tavoin kasvattamaan paperittomien asiakkaiden määrää. Pankkisektorilla valtaosa asiakkaista on paperittomia, mutta vakuutuspuolella paperittomien asiakkaiden määrä on huomattavasti pienempi. Syitä tähän ovat vakuutusliiketoimintaa säätelevät määräykset, pankkeja myöhäisempi liikkeellelähtö paperittömuuten sekä vähäinen hyöty asiakkaalle, jos ei lasketa mukaan ekologisen ajattelun mukanaan tuomaa imagoetua. Vakuutusyhtiöt siis edelleen tuottavat ja lähettävät paljon paperisia dokumentteja asiakkailleen. Edellä kuvassa 3 esimerkkinä olevalla vakuutusyhtiössä on kaksi AFP-dokumentteja (engl. *advanced fucntion printing*) sisältävää dokumenttivarastoa, jonne asiakkaille paperilla lähetettävät dokumentit talletetaan AFP-muodossa.

Taulukko 2. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön kuukausittaisista massatulostus määristä ja postitukseen käytettyjen kirjekuorien määristä.*

Month / Year	Total Print Amount	Total Insert Amount
August / 2012	7223886	727024
September / 2012	6329537	682331
October / 2012	5759398	614324
November / 2012	9583289	700262
December / 2012	7189273	711178
January / 2013	7342394	496065
February / 2013	6101118	389430
March / 2013	6352151	406890
April / 2013	6134827	379024
May / 2013	7754941	462614
June / 2013	7659774	409862
July / 2013	6796581	398610
August / 2013	6976160	426911
September / 2013	6278233	412626
October / 2013	7846235	504085
November / 2013	7753639	461260
December / 2013	8249059	574616
January / 2014	6930687	491813
February / 2014	6169621	403352
March / 2014	6694737	472363
April / 2014	6445745	424593
May / 2014	6849791	435430
June / 2014	6610461	745197
July / 2014	7168968	761761
August / 2014	5602611	686115
Septembet / 2014	5729186	616585
October / 2014	6846467	689148

Taulukon 2 ensimmäisessä sarakkeessa Month /Year on tieto kuukaudesta (engl. *month*) ja vuodesta (engl. *year*), jonka tulostettujen A4-sivujen kokonaistulostusmäärät esitetään

sarakkeessa Total Print Amount ja käytettyjen kirjekuorien kokonaismäärät esitetään sarakkeessa Total Insert Amount Esimerkiksi September / 2013 tarkoittaa 2013 syyskuuta, jolloin tulostettiin 6 278 233 kappaletta A4-sivuja ja niiden kirjekuoritukseen käytettiin 412 911 kirjekuorta. Kuvassa 4 on taulukon 2 tiedot esitetty graafisessa muodossa.



Kuva 4. *Pohjoismaissa toimivan vakuutusyhtiön kuukausittaiset massatulostusmäärät kolmen vuoden ajalta.*

Kuvasta 4 nähdään, että kuukausikohtaiset tulostusmäärät vaihtelevat vajaasta 6 000 000:sta aina yli 9 500 000:aan asti. Nämä ovat A4-paperille tehtyjä tulosteita. Keskimääräinen kuukausitulostemäärä on noin 6 900 000. Kuvassa oleva punainen käyrä kuvaa kirjekuorien (engl. *insert*) kuukausikulutusta, kun tulostetut dokumentit postitetaan vakuutusyhtiön asiakkaille. Massatulostuksen suuruuteen tai sen käyttöön ei dokumenttien metadatalla tai sen pohjalta tehdyllä raportoinnilla voida vaikuttaa. Sen sijaan dokumenttien metadatalla voidaan saada selville, mitä AFP-dokumentteja ja kuinka usein asiakkaat hakevat dokumenttivarastoista. Tämä raportointi edellyttää, että dokumenttien metadatassa on attribuutit dokumenttityypille (esim. lasku, vakuutusehto tai vakuutuskirja) sekä hakukerroille. Näihin kahteen attribuuttiin perustuvat kuukausittaiset raportit antavat yleiskuvan asiakkaita kiinnostavista dokumenteista ja dokumenttien hakuaktiivisuudesta. Toistuessaan kuukaudesta toiseen näiden raporttien pohjalta voidaan muodostaa trendi, jonka kehitystä voidaan seurata ja jonka perusteella voidaan tehdä ennusteita. Tarkempia raportteja, trendejä ja ennusteita saadaan, kun kahden metadatan attribuutin sijasta käytetään useampaa attribuuttia.

Edellä olevasta huomataan, että dokumenttivarastoja tulee ylläpitää, valvoa ja kehittää jatkuvasti, jotta ne pysyvät hyvässä kunnossa. Dokumenttien metadataa hyväksikäyttäen voidaan saada tietoa mm. tietovarastojen kasvuvauhdista, koosta, vasteajoista, talletusmedian täyttöasteesta, virhetilanteista, mitä dokumentteja on haettu eniten, varmuuskopiosta. Tiedon saanti edellyttää määrämuotoisten raporttien ajamista sääntöpohjaisella automatiikalla säännöllisin väliajoin. Dokumenttien metadataa voidaan käyttää sellaisenaan tai yhdistettynä muualta saatavaan tietoon määriteltäessä raportointisääntöjä. Metadatasta saatava tieto voidaan yhdistää esimerkiksi järjestelmälokeista saatavaan tietoon.

Tulevissa kohdissa käsitellään tarkemmin, miten näitä raportteja voidaan käyttää ohjaamaan yrityksen liiketoimintaa, infrastruktuurin suunnittelua ja ylläpitoa ja tietoturvaa.

4.3 Tiedolla ohjaaminen ja johtaminen

Kohdassa 4.3 selvitetään lukijalle erilaisten tietolähteiden käyttöä vakuutusliiketoiminnassa sekä eri lähteistä saatavan tiedon yhdistämistä asiakkuuden hoidossa.

Maailmassa arvioidaan nykyisin olevan 8 zettatavun (1 ZB = 10^21 tavua) verran digitaalista tietoa (Takala 2015). Digitaalisen tiedon määrän kasvuvauhti on kiihtyvä ja erilaisten digitapahtumien kirjo kasvaa edelleen. Tiedon määrän kasvaessa entistä tärkeämmäksi tulee se, miten sitä voidaan hyödyntää ja käsitellä siten, että siitä saadaan oikeata tietoa yrityksen ohjaukseen ja päätöksen tekoon. Automatiikalla voidaan tuottaa raportointia antamaan tilannekuvan ja kehitysennusteen. Näiden perusteella voidaan ohjata ja päättää, miten pitää seuraavaksi toimia. Dokumenttivarastot ovat yksi tietolähde muiden joukossa, mistä saatavaa tietoa voidaan käyttää ohjaukseen ja päätöksen tekoon. Yksinään se ei anna tarvittavaa kokonaiskuvaa, mutta yhdistettynä muihin tietolähteisiin kuva tarkentuu. Muita tietolähteitä, joiden kanssa dokumenttivarastoista saatavaa tietoa yhdistellään, ovat erilaiset asiakkuuden hoitojärjestelmät (engl. *customer relationship management*), sosiaalisen median kanavat (Facebook, Twitter, Pinterest, Vimeo jne.) ja asiakkaan käyttäytymistä seuraavat järjestelmät yrityksen nettisivuilla ja itsepalvelukanavilla.

Vakuutusliiketoiminnassa dokumenttivarastoista metadatan avulla saatavaa tietoa yhdistetään asiakkaasta olevaan ydintietoon (engl. *master data*). Ydintieto käsittää tietoa mm. asiakkaan terveydestä, perheestä, omaisuudesta, henkilötiedot, mahdolliset maksuhäiriötiedot, varallisuudesta. Tämä ydintieto asiakkaasta on pohjatietoa, jota täydennetään dokumenttien metadatasta saatavalla tiedolla ja asiakuuden hoitojärjestelmästä saatavalla tiedolla. Näiden tietojen yhdistelmän avulla vakuutusyhtiöt saavat melko hyvän kuvan asiakkaasta ja voivat tehdä esimerkiksi tietylle asiakassektorille kohdennettua markkinointia. Hyvä esimerkki tällaisesta markkinoinnista on tietyn automerkin omistajalle suunnattu lisäturvakampanja.

Vakuutusyhtiöt haluaisivat seurata asiakasta entistä tarkemmin ja perusteellisemmin. Erityisesti tietoa asiakkaan nettikäyttäytymisestä haluttaisiin tallentaa. Asiakkaan sosiaalisen median käyttöä haluttaisiin analysoida, palvelupuheluja tallettaa muuhunkin tarkoitukseen kuin vain määräaikaisesti koulutuskäyttöön, seurata miten asiakas käyttää yrityksen nettisivuja ja miksi asiakas ei tehnyt ostotapahtumaa loppuun. Erityisen mielenkiintoinen mahdollisuus vakuutusyhtiöille olisi päästä käsiksi asiakkaan paikkatietoihin. Jos vakuutusyhtiöt voisivat seurata asiakkaan liikeitä ja saada selville missä hän kulloinkin on niin, se avaisi monia uusia myynti- ja markkinointimahdollisuuksia. Toinen houkutteleva mahdollisuus vakuutusyhtiöille olisi päästä käsiksi asiakkaan netin hyvinvointipalveluissa olevaan elämäntapa ja kuntoilutietoon, jota hyväksikäyttäen voitaisiin esimerkiksi antaa kohdennettuja alennuksia sairaskuluvakuutuksien maksuista. Vakuutusyhtiöt eivät kuitenkaan voi toteuttaa kaikkia näitä asiakkuuteen liittyviä uusia tiedonkeruutapoja ja suunnitelmia, koska asiakkaan yksityisyydensuojaa suojellaan monin eri tavoin.

Kirjoittajan mielestä dokumenttien metadatasta saatavaa tietoa ei ole täysin haluttu tai osattu hyödyntää vakuutusyhtiöissä. Jos tuota tietoa hyödynnettäisiin paremmin, vakuutusyhtiöt voisivat tehostaa omaa sisäistä toimintaansa, säästää operatiivisissa kustannuksista ja tehostaa vakuutusten myyntiä. Dokumenttivarastossa olevien dokumenttien metadata yhdistettynä big dataan mahdollistaisi uuden tyyppisiä vakuutustuotteita ja -palveluita asiakkaille. Yksinkertaistaen sanottuna vakuutusyhtiöiden tulevaisuuden tavoitteena tulisi olla nykyisten vakuutustuotteiden ja -palveluiden ”tappaminen” uusilla paremmilla tuotteilla ja palveluilla käyttäen saavilla olevaa tietoa entistä paremmin.

4.4 Ulkoinen ja sisäinen asiakas ja asiakkuus

Kodassa 4.4 esitellään, mitä ulkoisella ja sisäisellä asiakuudella tarkoitetaan sekä miten ne poikkeavat toisistaan. Lisäksi kerrotaan, miten dokumentteihin liittyvää metadataa voidaan hyödyntää.

Vakuutusliiketoiminnassa puhutaan kahdesta eri asiakkaasta; ulkoisesta asiakkaasta ja sisäisestä asiakkaasta. Ulkoisella asiakkaalla tarkoitetaan vakuutusyhtiöltä vakuutuksia ostaneita asiakkaita. Sisäisellä asiakkaalla tarkoitetaan vakuutusyhtiön työtekijöitä, jotka hoitavat vakuutusyhtiössä erilaisia tehtäviä sekä palvelevat ulkoisia asiakkaita. Näillä kahdella asiakuudella on hieman erilaiset dokumenttivarastojen käyttötavat ja käyttötarpeet, mutta myös yhtäläisyyksiä.

Ulkoiset asiakkaat voidaan vielä jakaa yksityisasiakkaisiin ja yritysasiakkaisiin. Molemmat ulkoiset asiakkaat haluavat tyypillisesti tietoa vakuutuksiensa turvasta, vakuutuskorvauksista, vakuutusmaksuista, yhteenvedon kaikista vakuutuksista ja uusista tuotteista. Tätä tietoa haetaan käyttäen netissä olevia itsepalvelukanavia, jotka mahdollistavat asiakuuden tunnistuksen jälkeen dokumenttien hakemisen vakuutusyhtiön dokumenttivarastoista ja esittämisen oikeassa muodossa asiakkaan käyttämässä päätelaitteessa. Dokumentit poimitaan dokumenttivarastosta asiakasnumeroa ja vakuutusnumeroa käyttäen.

Näiden kahden metatiedon lisäksi käytetään vahinkotunnusta, jos haetaan korvaustietoa tai vahinkoilmoitusta. Nämä tiedot ovat talletettuna dokumentin metadataan jokainen omaan attribuuttiinsa. Näiden yleisimmin käytettyjen metadata-attribuuttien lisäksi voidaan käyttää muitakin attribuutteja. Esimerkiksi voidaan käyttää dokumentin version kertovaa attribuuttia, jos halutaan jokin tietty vanhempi versio dokumentista tai aikaleima-attribuuttia, jos halutaan kaikki tietyllä päivämäärällä tehdyt asiakkaan dokumentit.

Sisäistä asiakasta ei voida jakaa osiin järkevästi. Yksi mahdollinen tapa olisi jakaa sisäinen asiakas vakuutusyhtiön organisaation mukaisesti, mutta jako ei olisi oikein ja kovinkaan käytännöllinen. Yritysten organisaation rakenteet vaihtuvat usein aiheuttaen jakoperusteiden muutoksia, vaikka työtehtävät ja tarpeet käyttää dokumenttivarastoja pysyisivät samoina. Riippumatta vakuutusyhtiön organisaatiosta yrityksen henkilökunnan tarvitsee käyttää dokumenttivarastoja yli organisatoristen rajojen. Esimerkiksi yksityisasiakkaita palvelevan organisaation työntekijöillä on pääsy myös yritysasiakkaiden dokumentteihin ja päinvastoin. Vakuutusyhtiön henkilökunta käyttää dokumenttivarastoja samaan tapaan kuin ulkoiset asiakkaat. He hakevat dokumenttivarastosta asiakkaan dokumentteja saadakseen kuvan asiakkaan vakuutushistoriasta ja mahdollisuuksista lisämyyntiin ja markkinoitiin. Esimerkiksi tietoa asiakkaan koko vakuutusportfoliosta voidaan käyttää, keskittämisalennusta laskettaessa tai perusteena tarjota asiakkaalle uusia täydentäviä vakuutustuotteita. Dokumenttien metadatasta tyypillisesti käytetään asiakasnumeroa, jolla saadaan lista asiakkaan kaikista vakuutuksista. Asiakasnumeron lisäksi usein käytetään aikaleima-attribuuttia rajoittamaan listan pituutta tai tarkentamaan haluttua tarkasteluaikaa. Asiakasrajapinnassa työskentelevät käyttävät paljon vahinkotunnus-attribuuttia, koska asiakkaat ovat usein kiinnostuneita sattuneiden vahinkojen korvauksista ja korvausperusteista.

Yhteenvetona voidaan todeta, että ulkoisen ja sisäisen asiakkaan tarpeet päästä käsiksi vakuutusyhtiön dokumenttivarastoihin ovat samankaltaisia, mutta näkökulma ja käyttötarkoitus ovat erilaisia. Yksittäinen ulkoinen asiakas haluaa yleensä tietoa jostakin yksittäisestä vakuutuksesta tai vahingosta. Sen sijaan sisäinen asiakas haluaa kokonaisnäkymän asiakkaan koko vakuutushistoriaan ja -portfolioon. Dokumenttien metadatasta tärkeimpiä ja käytetyimpiä attribuutteja ovat asiakasnumero, vakuutusnumero ja aikaleimat sekä vahinkotunnus.

4.5 Uudet palvelut ja tuotteet

Tämän kohdan tavoitteena on antaa lukijalle kuva miten dokumenttien metadataa voidaan hyödyntää liiketoiminnassa uusien palveluiden ja tuotteiden suunnittelussa ja kehittämisessä.

Kohdassa 4.3 todettiin, että dokumenttien metadatasta saatavaa tietoa ja raportointia voidaan käyttää yhdessä muualta saatavan tiedon kanssa lisäämään vakuutusmyyntiä ja

markkinointia. Näin saadaan nykyiset vakuutustuotteet tehokkaasti hyödynnettyä ja tuotettua asiakkaille lisäarvoa. Vakuutusyhtiöt eivät kuitenkaan voi ikuisesti käyttää nykyisiä tuotteita ja palveluita, vaan niiden tulee kehittää uusia vakuutustuotteita ja -palveluita nykyisille ja uusille asiakkaille. Kuten aikaisemmin todettiin, vakuutusliiketoimintaa säädellään säädöksillä ja laeilla, jotka asettavat uusille vakuutustuotteille tietyt vaatimukset ja rajat. Lisäksi tulee muistaa, että sääntelyn taso vaihtelee maasta toiseen ja että Euroopan komissio säätelee tietosuojaa koko Euroopan tasolla (Euroopan komissio 2015). Se mikä on sallittu yhdessä maassa, ei olekaan sallittua toisessa maassa. Esimerkki käyttämässämme Pohjoismaissa toimivassa vakuutusyhtiössä tämä on otettu huomioon siten, että liiketoiminta tekee läheistä yhteistyötä juristien kanssa ennen kuin uusia tuotteita julkaistaan Pohjoismaisesti. Tästä yhteistyöstä huolimatta liiketoiminta usein unohtaa tarkistaa kaikkien maiden lainsäädännön ja käytännöt. Seuraavassa esitetään uusia palveluita ja tuoteideoita ottamatta huomioon vakuutusliiketoiminnan sääntelyä. Esimerkit ovat oikeita ideoita, joita on tutkittu ja työstetty Pohjoismaassa toimivassa vakuutusyhtiössä.

Paikkatiedon saaminen ja hyödyntäminen avaa monia mahdollisuuksia uusille vakuutustuotteille ja -palveluille. Paikkatietoa hyväksikäyttämällä voitaisiin toteuttaa vaihtelevia makuutusmaksuja ja vakuutusturvia. Ehkä eniten tutkittu esimerkki tästä on autovakuutuksen turvatasojen ja maksujen muuttuminen auton sijaintitietojen perusteella. Auton liikkuessa sen sijaintitiedot välitettäisiin reaaliajassa vakuutusyhtiölle, joka siten näiden tietojen perusteella päättelisi sopivat vakuutusmaksut ja turvat. Reaaliaikaiset sijaintitiedot mahdollistaisivat myös arviot kuljettajan ajotavasta. Ajaako kuljettaja ylinopeutta aiheuttaen näin vahinkoriskin kasvua, joka puolestaan lisää perittävien vakuutusmaksujen suuruutta. Paljon käytetyssä esimerkissä auton liikkeitä seurataan ja sen ylittäessä maan rajan auton vakuutusturvaa muutetaan kyseisen maan riskitasonmukaisesti. Esimerkiksi auton siirtyessä maahan, jossa autovarkauksien riski on suurempi kuin lähtömaassa, auton vakuutuksen varkausturvaa liittyvää vakuutusmaksua korotetaan ja tarjotaan vakuutuksenottajalle mahdollisuutta ostaa lisäturvaa varkauden varalle. Asiakkaan hyväksyessä ehdotetun lisäturvan aiheuttaa se uuden vakuutussopimuksen luomisen kattamaan muuttuneen tilanteen. Dokumenttivarastoon luodaan siis uusi dokumentti siihen kuuluvin metatiedoin. Toisin sanoen aina, kun vakuutusturvassa ja -vakuutusmaksuissa tapahtuu muutoksia, aiheuttaa se muutoksia asiakkaan dokumentteihin vakuutusyhtiön dokumenttivarastoissa. Erityisen tärkeiksi metadatan attribuuteiksi muodostuvat vakuutusturvaan liittyvät aikaleimat ja niihin mahdollisesti liitettävät sijaintitietoattribuutit. Myöskin muuttuneen vakuutussopimuksen hyväksymisestä tulee merkintä dokumenttien metadataan. Näiden metadata tietojen perusteella voidaan nähdä, miten asiakkaan vakuutukset ovat muuttuneet asiakkaan sijaintitietojen muuttuessa.

Edellä olevaa sijaintitietojen perusteella muuttuvan vakuutusturvan ja -maksujen periaatetta voidaan soveltaa myös henkilövakuutuksiin. Tässä tapauksessa henkilövakuutuksen

tapaturma- ja sairaskuluvakuutuksen ehtoja ja maksuja muutetaan asiakkaan sijaintitietojen perusteella. Perusteena voi olla esimerkiksi asiakkaan siirtyminen toisen maahan, jossa henkirikoksia tehdään paljon tai jossa on suuri sairastumisiksi. Näillä muutoksilla on samanlainen vaikutus asiakkaan dokumentteihin ja niiden metadatan sisältöön kuin autovakuutusesimerkissä. Tällainen henkivakuutuksen sijaintitietoihin perustuvat vakuutusmuutoksen edellyttävät asiakaan varmaa tunnistusta. Vaikka kännykästä saataisiin tarvittavat sijaintitiedot, ei se vielä ole asiakkaan yksilöivä varmatunnistus. Kännykkä voi olla jonkin muun kuin vakuutuksen ottajan käytössä.

Molemmissa esimerkeissä asiakkaan vakuutushistoria tulee selvittää, jotta voidaan arvioida kuinka riskialtis asiakas on. Tämä taas omalta osaltaan vaikuttaa asiakkaan sijaintitietojen mukaan muuttuviin turviin ja maksuihin. Asiakkaan vakuutushistoria selvityksessä dokumenttivarastoissa olevat dokumentit ja niiden metadata ovat hyödyksi joko erikseen tai yhdessä muulta saatavan tiedon kanssa. Dokumenttien metadatan hyväksikäyttäen voidaan esimerkiksi hakea, montako tapaturmaan liittyvää vahinkoilmoitusta asiakas on tehnyt viimeisen kahden vuoden aikana tai montako vahinkoilmoitusta asiakas on tehnyt liittyen autovakuutukseen. Dokumenttien metadataan perustuvan dokumenttihaun jälkeen voidaan saatua hakutulosta analysoida eteenpäin käyttäen tekstianalytiikkaa. Tekstianalytiikalla saadaan hakutuloksesta esimerkiksi selville, missä asiakkaan tapaturmat ovat tapahtuneet tai missä autoa on kolaroitu. Näin saadaan tarkempaa tietoa asiakkaan käyttäytymisestä, nähdään trendejä ja voidaan tehdä ennusteita, joiden perusteella voidaan asiakkaalle tarjota uusia turvia, muuttaa turvaa ja maksuja. Asiakkaalle voitaisiin tarjota hyvinkin henkilökohtaisesti räätälöityjä vakuutuksia ja palveluita, jos käytössä olisi dokumenttien metadatalla saatu vakuutushistoria yhdistettynä netin palveluista löytyvään asiakkaan itsensä tallentamaan elämäntapa- ja kuntoilutietoon.

Vakuutusyhtiöt tarjoavat asiakkailleen erilaisia itsepalvelukanavia. Erityisen kiinnostuksen kohteena ovat olleet ja ovat edelleen asiakkaan käyttäytyminen vakuutusyhtiöiden nettisivustoilla. Mitä asiakas klikkaa, kuinka kauan asiakas viipyy kullakin sivulla ja miksi asiakas jättää ostokset kesken. Tätä klikkaushistoriaa halutaan tallettaa ja analysoida, jotta voidaan parantaa asiakaspalvelua, helpottaa vakuutusten uusmyyntiä sekä tarjota asiakkaille apua kesken jääneeseen ostoon. Ihanne tilanteessa asiakas on tunnistettu, jolloin dokumenttien metadatan avulla voidaan kartoittaa vakuutushistoriaa ja yhdistää tämä tieto asiakkaan talletettuun klikkaushistoriaan. Kirjoittajan kokemusten mukaan lähes aina asiakasta ei voida tunnistaa vaan asiakas selaa nettisivuja ja niillä olevia palveluita anonyymisti. Tällä hetkellä vakuutusyhtiöllä on paljon erilaisia suunnitelmia siitä miten asiakkaan käyttäytymistä yhtiöiden nettisivuilla voitaisiin tallettaa, analysoida ja hyväksikäyttää erikseen tai yhdistettynä muualta saatavaan tietoon.

Digitalisaatio (engl. *digitalization*) ja robotisaatio (engl. *robotization*) tulevat vaikuttamaan vakuutusliiketoimintaan sekä siihen liittyvään dokumentointiin ja dokumenttivarastoihin. Osa nykyisistä vakuutustuotteista tulee poistumaan ja myös niihin liittyvät dokumentit tulevat poistumaan. Dokumentit ja niihin liittyvät metadatan tuhoaminen tulee

tapahtumaan viiveellä määrätyn arkistointiajan jälkeen. Edellä oleva autovakuutusesimerkki on myös hyvä kuvaamaan tätä muutosta. Kuljettamattomien robottiautojen korvatessa nykyisen kaltaiset autot aiheuttaa se kaikkein positiivisempien ennusteiden mukaan kuljettajan virheistä johtuvien onnettomuuksien, kolareiden ja kuolemien katoamisen kokonaan. Jos näin käy, tarkoittaa se, että nykyisten kaltaisia autovakuutuksia ei enää tarvita. Kotitalouksien robottiautot olisivat rinnastettavissa kodin irtaimistoon samaan tapaan, kuin kahvinkeitin tai pölynimuri, jolloin ne voitaisiin liittää kotivakuutuksen piiriin. Tällöin muita vakuutuksia ei robottiauton omistajan tarvitsisi ottaa. Vakuutusdokumenttien näkökulmasta erillisiä robottiautoihin liittyviä vakuutusdokumentteja ei tarvita vaan ainoaksi dokumentiksi jää kotivakuutus ja siihen liittyvä metadata. Lisäksi voi olla, että digitalisaation myötä vakuutusdokumentaation tarve katoaa tai muuttuu niin, että nykyinen dokumentaatiota metadatoineen kutistuu muutamaan dokumenttiin. Esimerkiksi jo nyt yksinkertaisten vahinkojen vahinkoilmoituksien elinkaarien on lyhyt, koska ilmoitukset ja korvaukset käsitellään koneellisesti ilman viivettä. Voi siis olla hyvin mahdollista, että digitalisaation myötä vahinkoihin liittyvät dokumentit metadatoineen tulevat tarpeettomiksi ainakin yksinkertaisissa vahinkotapauksissa. Digitalisaation ja robotisaation vaikutuksien syvempi tutkiminen ja analysointi ovat tämän työn ulkopuolella ja ne vaatisivat oman jatkotutkimuksen.

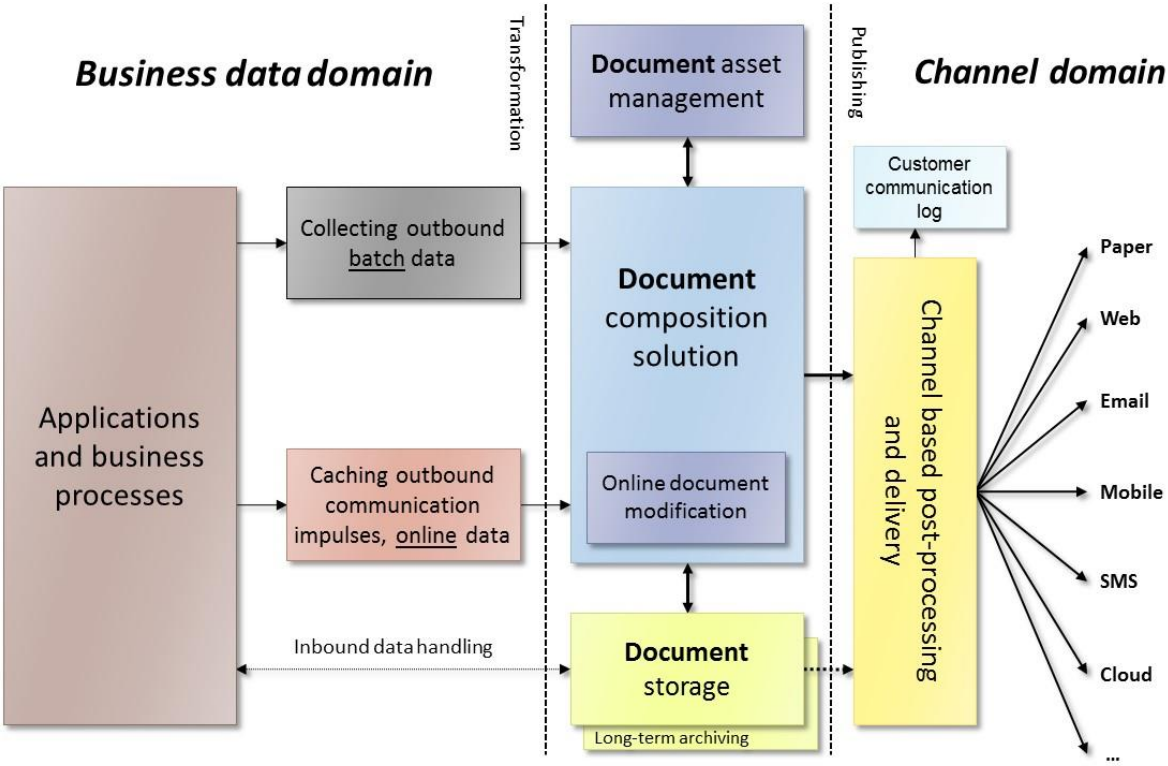
Yhteenvetona voidaan todeta, että dokumenttien metadata ei itsessään itsenäisesti mahdollista uusien vakuutustuotteiden ja palveluiden kehittämistä. Sen sijaan dokumenttien metadataa hyväksikäyttäen saadaan asiakkaan vakuushistoriasta esille pohjatiedot, joita voidaan jatkojalostaa eteenpäin. Jatkojalostuksessa mukaan tuodaan muualta saatua tietoa, jotka yhdistetään pohjatietoihin. Tämän jatkojalostuksen tavoitteena on saada mahdollisimman tarkka kuva asiakkaasta, jotta vakuutusyhtiöt voisivat ennustaa asiakaan käyttäytymistä etukäteen, kehittää uusia tuotteita ja palveluita asiakkailleen sekä samalla parantaa yrityksen tulosta.

4.6 Liiketoimintaprosessit

Kohdassa 4.6 esitellään, miten liiketoimintaprosessit ja dokumenttien metadata nivoutuvat yhteen. Kohdan motivaationa on esitellä, miten dokumenttien eri osien elementtien käsittely nivoutuu liiketoimintaprosesseihin ja dokumenttien metadataan.

Liiketoimintayksiköt ovat määritelleet ja hioneet omia työ- ja liiketoimintaprosesseja jo pitkään. Tämän työn tuloksena käytössä olevia prosesseja on pystytty yksikertaistamaan, tehostamaan ja nopeuttamaan. Työt voidaan tehdä vähemmällä henkilökunnalla ja samalla asiakkaan saama palvelu on parantunut. Prosesseihin liittyy paljon kommunikointia asiakkaan kanssa käyttäen erilaisia kommunikointitapoja. Prosessin eri vaiheissa käsitellään monia erilaisia dokumentteja, jotka sitten siirtyvät prosessin vaiheesta toiseen. Dokumenttien siirtyessä prosessin vaiheesta toiseen dokumentteja käsittelevät organisaatiot ja ihmiset vaihtuvat. Jokaisessa prosessin vaiheessa dokumenttien sisältö, metadata tai

molemmat muuttuvat ja täydentyvät. Tämä dokumenttien käsittely vie paljon aikaa ja vakuutusyhtiöt pyrkivätkin automaatiolla ja uudelleen organisoinnilla tehostamaan ja lyhentämään dokumenttien käsittelyyn käytettävää aikaa. Vakuutusyhtiöt tavoittelevat kuvan 5 kaltaista tavoitetilaa.



Kuva 5. *Kuvassa esitetään tavoitetila, jossa asiakas saa saman informaation ja palvelukokemuksen riippumatta palvelukanavasta tai asiakkaan päätelaitteesta.(Laine 2012)*

Kuvan 5 tavoitetila edellyttää, että dokumenttien luonti, muotoilu ja muunnokset kuhunkin jakelukanavaan (kuvassa 5 Channel domain) tehdään automaatiolla. Henkilökuntaa tarvitaan vain erikoistapauksissa esimerkiksi lisäämään asiakaspersonointia tai luomaan henkilökohtaiset asiakaskirjeet. Tavoitetilassa lähes kaikki metadata dokumentteihin luodaan automaattisesti. Automaation ulkopuolelle jäävät mahdolliset manuaalisesti tehtävät korjaukset tai käsittelijän tietojen muutokset. Toimiakseen dokumenttiautomaatio vaatii dokumenttien tyypityksen, tyypitykseen perustuvat vakioidut dokumenttimallit sekä vakioidut tekstikappaleet ja fraasit. Dokumenttiautomaatio ja vakioitu dokumenttien metadatamalli mahdollistavat laadullisesti korkeatasoisen ja tasalaatuisen dokumenttien metadatan.

Dokumenttiautomaation avulla liiketoimintaprosessien eri vaiheissa asiakasdokumentteja käsitelevien ei enää tarvitse luoda dokumentteja tai niihin liittyvää metadataa. Dokumenttiautomaatiossa dokumentit koostetaan ennalta määriteltyjen sääntöjen eli dokumenttilogiikan avulla käyttäen hyväksi vakuutusyhtiön asiakastietokantoja ja -sovelluksia (kuvassa 5 Business data domain) sekä vakioituja dokumenttimalleja. Käsittelijän tehtäväksi

jää siis vain hakea asiakkaan dokumentti tai dokumentit käyttäen hyväksi automatiikalla luotua metadataa. Dokumenttien löydyttyä käsittelijän tulee vielä tarkistaa ne ja tarvittaessa tehdä dokumentteihin asiakaspersonointia sekä valita, mihin palvelukanavaan dokumentit muotoillaan ja lähetetään. Dokumenttien manuaalinen tarkistus on poikkeustapaus. Yleensä dokumenttiautomaatiolla tuotetut dokumentit lähetään sellaisenaan suoraan asiakkaalle. Jakelukanava valinta perustuu puolestaan asiakkaan ilmoitukseen hänelle parhaiten sopivasta palvelukanavasta. Kirjoittajan saamien tietojen mukaan vakuutusyhtiöt ovat eri tasoilla edellä kuvatun tavoitetilan toteutuksessa. Jotkin vakuutusyhtiöt ovat hyvin pitkällä dokumenttiautomaatiossa, kun taas toiset ottavat vasta ensi askeliaan. Kaikille yhtiöille pulmallisia ovat massatulostukseen liittyvät AFP-muotoiset dokumentit ja niiden käsittely. Hankaluudet johtuvat AFP-tiedostojen rakenteesta. AFP-tiedosto muodostuu dokumenttimallista, asiakkaiden tiedoista sekä massatulostuksen ohjaustiedoista. Usein myös AFP-tiedostoissa oleva dokumenttien metadata on suppeaa verrattuna muihin tiedostoformaatteihin. Tämä johtuu massatulostuksen luonteesta, jossa metadatalla ei ole kovin suurta merkitystä. Tärkeitä on löytää oikeat dokumentit tulostettavaksi. Tähän tarkoitukseen riittää asiakasnumero ja aikaleima. Tarvittaessa metadataa voidaan generoida automatiikalla lisää muunnettaessa AFP-tiedostoja toisiin formaatteihin. Yleensä muunnos tehdään PDF-tiedostomuotoon, joka sitten lähetetään asiakkaan valitsemaan jakelukanavaan. Toinen vaikeasti hallittava asia on paperiton asiakas, joka haluaa kaiken tiedon mobiililaitteeseen eikä halua käyttää sähköpostia. Dokumentit tulee löytää nopeasti ja ne pitää optimoida kullekin mobiililaitteelle sopiviksi tunnistamalla asiakkaan mobiililaite. Tällaisia asiakkaita ei kuitenkaan vielä tiettävästi ole, mutta vakuutusyhtiöt pyrkivät varautumaan heihin.

Tiivistettynä voidaan todeta, että dokumenttien metadata liittyy liiketoimintaprosesseihin dokumenttiautomaation ja dokumenttihakujen välityksellä. Dokumenttiautomaatiolla luodaan laadukasta standardin mukaista metadataa, jota sitten käytetään oikeiden dokumenttien hakuun ja löytämiseen. Metadatalla haetut dokumentit sitten tarvittaessa muunnetaan ja lähetetään asiakkaan valitsemaan jakelukanavaan. Yleisimpiä haussa käytettyjä metadata-attribuutteja ovat vakuutusnumero, vahinkotunnus, terveystiedot, ajoneuvotiedot esimerkiksi rekisterinumero, henkilötunnus ja asiakasnumero

4.7 Petosten havaitseminen ja ennalta ehkäisy

Kohdan 4.7 tarkoituksena on antaa lukijalle käsitys, miten dokumenttien metadataa voidaan hyödyntää yhtenä osana petosten havaitsemisessa ja ennaltaehkäisyssä yhdessä muualta saatavien tietojen kanssa.

Vakuutusyhtiöissä on vakuutusetsiviä, jotka kuuluvat tutkijatoimeen. Heidän tehtävänään on selvittää ja ennaltaehkäistä vakuuksiin liittyviä rikoksia ja väärinkäytöksiä. Ne ovat vakuutuspetoksia, joita tekevät asiakkaat ja vakuutusyhtiöiden oma henkilökunta. Lähes kaikki petokset tai niiden yritykset ovat vakuutusyhtiöiden asiakkaiden tekemiä. Monet

tätä tutkintatyötä tekevät henkilöt ovat saaneet poliisikoulutuksen tai ovat entisiä poliiseja. He tutkivat mahdollisia väärinkäytöksiä perinteisin poliisin menetelmin, mutta entistä enemmän on myös tullut tarve käyttää tietotekniikan mahdollisuuksia hyväksi. Näitä mahdollisuuksia ovat tekstianalytiikan käyttö sekä dokumenttien ja tiedostojen metadatan käyttö. Tässä työssä keskitytään metadatan mahdollisuuksiin väärinkäytösten selvityksessä.

Kuten aikaisemmin on jo todettu, dokumenttien metadatan avulla voidaan löytää asiakkaaseen liittyvät kaikki dokumentit. Näin saadaan selville asiakkaan koko vakuutushistoria. Tutkimalla vakuutushistoriaa ja dokumenttien metadataa saadaan selville, millaisia vahinkoja ja kuinka usein asiakkaalla on ollut. Metadatan aikaleimoja tarkastelemalla nähdään, milloin vahingot ovat tapahtuneet sekä onko vahinkojen ja vahinkoilmoitusten aikajana uskottava.

Asiakkaiden tekemiin vahinkoilmoituksiin liitetään oheismateriaalina enenevässä määrin valokuvia ja videoita tapahtuneesta vahingosta. Tämä materiaali tallennetaan yhdessä vahinkoilmoituksen kanssa vakuutusyhtiön tietovarastoihin. Usein ne tallennetaan samaan dokumenttitietovarastoon vahinkoilmoituksen kanssa. Valokuvat ja videot ovat dokumenttien hallintajärjestelmien kannalla myös sähköisiä dokumentteja, joilla on dokumenttityyppi ja metadata-attribuutit. Tällä hetkellä asiakkaat lähettävät pääsääntöisesti valokuvia tapahtuneesta vahingosta, mutta videoiden osuuden ennustetaan kasvavan lähitulevaisuudessa. Vakuutusyhtiöiden tutkijatoimi saa jo nyt melko paljon teräväpiirtomateriaali poliisilta, jonka kanssa se tekee yhteistyötä.

Valokuviin liittyy kahden tyyppistä metadataa; exchangeable image file format (lyh. EXIF) metadataa ja international press telecommications council (lyh. IPTC) metadataa. EXIF-metadata liittyy tekniikkaan, ei itse kuvaan tai sen sisältöön. Lisäksi tätä metadataa ei voida jälkikäteen muokata, mutta se kuitenkin voidaan poistaa kokonaan. IPTC-metadata on taas muokattavaa itse kuvaan liittyvää metadataa. ITPC-metadata on alun perin kehitetty uutistoimistojen käyttöön helpottamaan kuvien arkistointia ja julkaisua. Monet kuvankäsittelyohjelmat tukevat ITPC-metadataa (ITPC 2010 ja Metadata working group 2015).

Väärinkäytösten tutkinnan kannalta EXIF on mielenkiintoisempi, koska sitä ei voi jälkikäteen editoida. Valokuviin EXIF-metadata syntyy, jos kamerassa on standardoidun metadatan (JEITA 2002) luonnin mahdollistava ominaisuus otettu käyttöön. EXIF-metadata sisältää seuraavat tiedot:

- Päivämäärä ja aika
- Kameran asetukset, esimerkiksi valotusaika, ISO-lukema ja objektiivin aukko
- Kuvatietoja, kuten esimerkiksi kuvan resoluutio ja kuvaformaatti
- Pikkukuva (engl. *thumbnail picture*) varsinaisesta valokuvasta
- Tekijänoikeustiedot (engl. *copyright information*)
- Paikkatiedot (engl. *location information*)

EXIF-metadata sisältää teknistä tietoa kuvasta ja käytetystä laitteistosta. Väärinkäytösten selvityksen kannalta mielenkiintoisimpia näistä tiedosta ovat paikkatiedot, päivämäärätiedot ja aikaleimat (Kansallinen kirjasto-hanke 2009 ja Photometadata.org 2015).

Väärinkäytösten tutkinnassa kuvien paikkatietoja ja vahinkoilmoituksen tietoja vertailemalla voidaan selvittää, onko kuvat otettu sieltä minne vahinko on vahinkoilmoituksessa kerrottu kohdistuneen. Päivämäärä- ja aikaleimoja voidaan taas selvittää, täsmäävätkö kuvat ja vahinkoilmoitukset ajallisesti. Asiakkaiden lähettämät kuvat voivat olla eri kohteesta ja otettu eri aikaan, mitä asiakas on ilmoittanut vahinkoilmoituksessa. Tällaisia petosyrityksiä on valtaosa tutkittavista tapauksista. Hankalammin selvitettäviä tapauksia ovat epäilyt vahinkojen ketjutuksesta tai tapaukset, joissa poliisi epäilee rikosta. Erityisen hankalia ovat myös vakuutusyhtiöiden henkilökunnan tekemien väärinkäytösten selvitys. Yleensä ne liittyvät myyntiin ja markkinoitiin, joissa on otettu erivapauksia liittyen vakuutusten ehtoihin ja hintoihin. Lähes aina näihin tapauksiin liittyy myös tietoturvan väärinkäytöksiä, joissa on käytetty, jonkun toisen tunnusta ja salasanaa. Tällaiset tapaukset ovat harvinaisia, mutta sitäkin harmillisimpia.

Yhteenvetoja voidaan todeta, että dokumenttien metadata on yksi tietolähde petosten selvityksessä ja antaa tutkijoille kokonaiskäsityksen tutkittavan kohteen vakuutushistoriasta sekä asiakkaan toimintatavoista ja käyttäytymisestä. Erityisesti asiakkaan vahinkoilmoitukseen liittämien valokuvien metadatat ovat vakuutusetsiville tärkeä tietolähde. Niiden sisältämän tiedon perusteella voidaan selvittä tapahtumien ajallinen eheys sekä vahingonkohteen sijainti ja kohteen oikeellisuus. Yhdistämällä näin saatu tieto muista lähteistä saatavaan tietoon, voidaan petokset ja niiden yritykset tunnistaa. Lisäksi voidaan tehdä asiakkaista riskianalyyseja, joiden avulla voidaan varautua mahdollisiin väärinkäytöksiin ja ennalta ehkäistä uusia väärinkäytöksiä.

Vakuutusyhtiöiden oman henkilökunnan väärinkäytöksissä dokumenttien metadatan avulla saadaan selville väärinkäytöstä epäillyn henkilön dokumenttihistoria. Tästä tutkijat näkevät epäilyn henkilön asiakaskunnan ja mitä sopimuksia tehty. Dokumenteissa olevia aikaleimoja ja käyttäjätietoja voidaan puolestaan verrata tietoturvalokeihin ja tietoturvamäärityksiin, jotta voidaan selvittää mahdolliset käyttäjätunnusvarkaudet ja väärinkäytökset. Dokumenttien metadatan avulla voidaan myös selvittää yksitäisen dokumentin versiohistoria. Versiohistoriasta tutkijat voivat selvittää, kuka on luonut dokumentin, kuka sitä on muuttanut ja millaisia käyttäjäoikeuksia muutokset ovat edellyttäneet. Esimerkiksi vain erikseen nimetty osa vakuutusyhtiön henkilökunnasta voi käsitellä asiakkaiden terveystietoja.

Kirjoittajan tietojen mukaan asiakkaalta tulevien valokuva-, video- ja äänitiedostojen metadataa ei kattavasti talleteta dokumenttivarastoihin. Vakuutusyhtiöillä on suunnitelmia ja ajatuksia, miten ne voisivat paremmin tallentaa ja hyödyntää tätä metadataa. Erityisesti

vakuutusyhtiöiden tutkijatoimi pyrkii aktiivisesti edistämään tätä kehitystä. Ensimmäisenä he haluaisivat paremmin hyödyntää valukuvien metadataa ja heti seuraavana olisi vuorossa videoiden metadatan hyödyntäminen.

4.8 Tietotekniikan operatiiviset kustannukset

Tässä kohdassa käsitellään dokumenttien metadataa tietotekniikan operatiivisten kustannusten näkökulmasta. Erityisesti keskitytään, miten operatiivisia kustannuksia voidaan hallita käyttäen dokumenttien metadatan mahdollisuuksia hyväksi.

Dokumenttivaraston sisällä dokumenttien koko, dokumenttityypit, tiedostomuodot ja sisällöt vaihtelevat. Vain tekstiä sisältävät dokumentit eivät vaadi paljon talletustilaa ja ne voidaan tehokkaasti pakata pienempään tilaan käyttäen pakkausalgoritmeja. Videoita, kuvia ja ääntä sisältävät tiedostot vaativat huomattavasti enemmän talletustilaa ja niitä ei voida enää talletusvaiheessa pienentää pakkausalgoritmeja käyttäen, koska ne ovat jo itsessään pakatussa muodossa. Tekstimuotoisia dokumentteja voidaan tallettaa yhteen tietovarastoon enemmän kuin dokumentteja, jotka sisältävät kuvaa, videota tai ääntä ilman, että niiden hakuajat kasvavat liian suuriksi liiketoimintaprosessi näkökulmasta katsottuna.

Talletetuista dokumenteista tulee operatiivisia kuluja, joita ovat:

- levytilakustannukset
- varmuuskopiointikulut
- palvelinkustannukset
- tietoliikennekustannukset
- konesaliin liittyvät kustannukset, esimerkiksi sähkö, jäähdytys jne.
- monitorointikustannukset
- varusohjelmistokustannukset

Kertyneet operatiiviset kustannukset jaetaan vakuutusyhtiön sisällä eri liiketoimintayksiköille. Kustannusten jakoperusteena käytetään liiketoimintayksikön eri resurssien kulutusta. Tämä on hyvä operatiivisten kustannusten jakoperuste. Toinen vakuutusyhtiöissä käytetty kustannusten jakoperuste on liiketoimintayksiköiden kanssa sovittu kiinteä kustannusten jako. Joissakin tapauksissa voidaan myös käyttää maakohtaista kustannustenjakoa.

Operatiivisia kustannuksia pyritään hallitsemaan ja ennustamaan mahdollisimman tarkasti. Hallinnassa tavoitteena on päästä mahdollisimman hyvään talletusmedian kokonaiskäyttöön. Tähän pyritään tarkkailemalla talletusmedian, yleensä erilaisten levyjen ja levyalueiden, käyttöä ja täyttöastetta. Tavoitteena on välttää ylimitoitettuja levyalueita, jotka eivät vastaa käyttötarvetta ja - ennusteita. Lisäksi halutaan minimoida orvot levyallokoinnit eli levyalueiden varaukset. Orvot levyallokoinnit ovat levyalueita, jotka on mää-

ritelty ja varattu, johonkin käyttöön, mutta jostain syystä niitä ei ole käytetty. Käyttämättöminä orvot levyaluevaraukset kerryttävät operatiivisia kustannuksia turhaan ja ne tulee poistaa. Tällainen hallinta edellyttää säännöllistä automaattista raportointia ja mahdollisimman tarkkoja ennusteita talletusmedian käytön trendeistä ja kasvuvauhdista.

Dokumenttivarastoissa talletettujen dokumenttien metadataa voidaan käyttää hyväksi talletusmedian käytön hallinnassa sekä kustannusjaossa. Vakuutusyhtiöiden käytössä olevissa dokumenttivarastoissa käytetään hieman toisistaan poikkeavia metadatamalleja riippuen dokumenttivaraston käyttötarkoituksesta ja sisällöstä. Massatulostuksen dokumentteja säilyttävässä dokumenttivarastossa metadata-attribuutteja on vähemmän käytössä kuin dokumenttivarastossa, jossa säilytetään asiakkaan kanssa käytyä kommunikaatiota. Kuitenkin kaikissa näissä erityyppisissä dokumenttivarastoissa on seuraavat dokumenttien metadata-attribuutit:

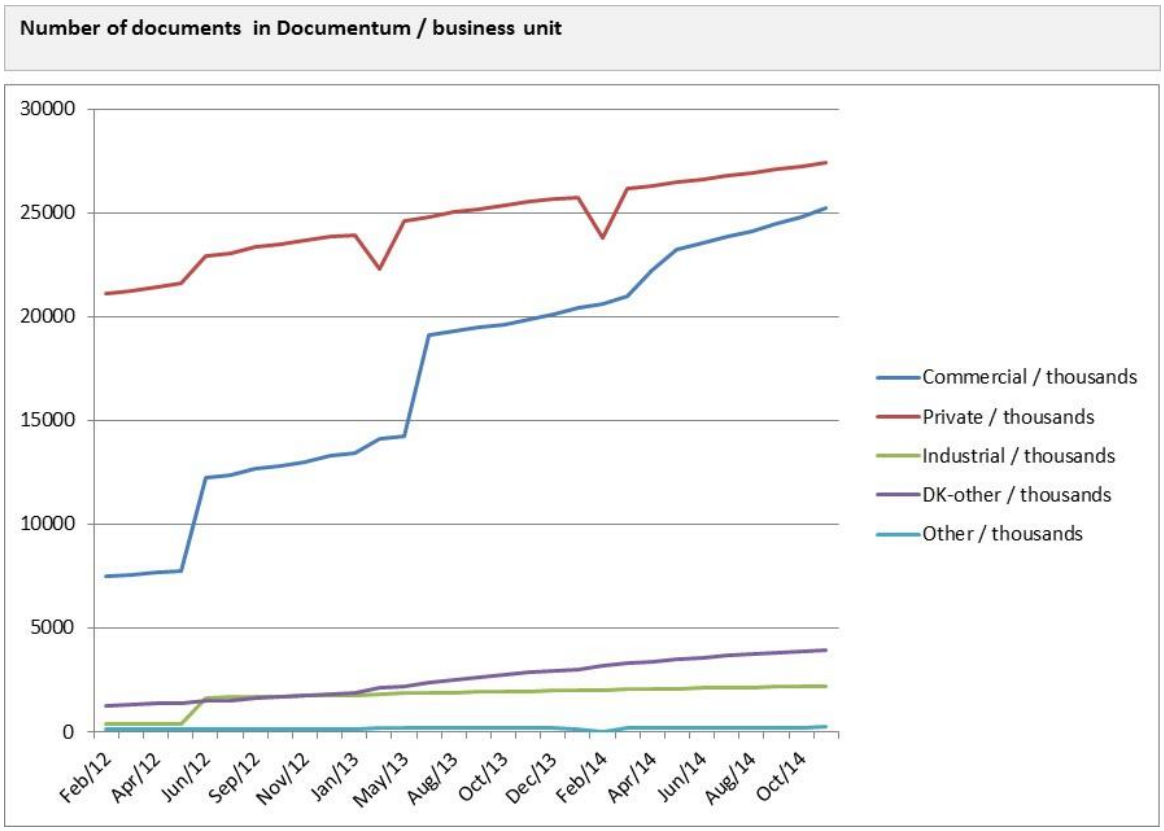
- liiketoimintayksikkö (yksityis, yritys tai teollisuus)
- aikaleima dokumentin luonnille
- maakoodi
- dokumenttityyppi, jonka perustuu vakuutusyhtiön dokumenttityypitykselle (esimerkiksi vakuutuskirja, vahinkoilmoitus, korvauspäätös jne.)
- sovelluskoodi, joka kertoo dokumentin luoneen sovelluksen

Kohdassa 4.2 esiteltiin esimerkkivakuutusyhtiön dokumenttivarastoista saatavaa raportointia. Esitelty raportointi on kokonaisraportointi, josta saadaan rakennettua ennusteita ja voidaan laskea koko dokumenttivaraston keskimääräinen kuukausittainen kasvuvauhti. Näitä hyväksikäyttäen voidaan jo seurata operatiivisia kustannuksia ja ennustaa niiden kehitystä. Kohdan 4.2 raportit eivät kuitenkaan anna tarvittavaan tietoa, jolla kertyneitä kustannuksia jakaa eri liiketoimintayksiköille tai dokumentteja luoneille sovelluksilla. Seuraavassa esimerkeissä kohdan 4.2 raporteissa käytettyihin dokumenttien metadata-attribuutteihin on lisätty liiketoimintayksikön kertova attribuutti.

Taulukko 3. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön yhden dokumenttivaraston dokumenttimääristä ja kasvusta liiketoimintayksiköittäin.*

Month / Year	Commercial / thousands	Private / thousands	Industrial / thousands	DK-other / thousands	Other / thousands
February / 2012	7507	21129	354	1265	159
March / 2012	7576	21241	359	1309	159
April / 2012	7676	21429	369	1372	159
May / 2012	7771	21613	370	1410	159
June / 2012	12274	22917	1654	1500	159
July / 2012	12373	23022	1666	1536	159
September / 2012	12649	23328	1695	1646	161
October / 2012	12812	23488	1712	1705	161
November / 2012	13008	23662	1730	1777	162
December / 2012	13302	23824	1755	1843	164
January / 2013	13456	23940	1771	1900	165
April / 2013	14093	22265	1836	2128	173
May / 2013	14260	24585	1853	2192	175
June / 2013	19081	24819	1882	2366	180
August / 2013	19318	25068	1908	2522	185
September / 2013	19454	25194	1922	2634	187
October / 2013	19616	25332	1939	2734	189
November / 2013	19863	25518	1961	2859	194
December / 2013	20096	25646	1981	2961	198
January / 2014	20396	25755	1999	3013	109
February / 2014	20585	23816	2033	3196	37
March / 2014	20982	26143	2052	3304	207
April / 2014	22205	26270	2072	3400	210
May / 2014	23213	26443	2095	3527	214
June / 2014	23538	26604	2113	3594	216
July / 2014	23826	26757	2129	3671	219
August / 2014	24115	26917	2144	3747	221
September / 2014	24505	27104	2164	3827	224
October / 2014	24785	27231	2178	3889	226
November / 2014	25203	27392	2198	3967	229

Taulukon ensimmäisessä sarakkeessa Month / Year on tieto kuukaudesta (engl. *month*) ja vuodesta (engl. *year*), jolta dokumenttien kappalemäärät ovat. Seuraavien sarakkeiden otsikossa kerrotaan minkä liiketoimintayksikön, yritys (engl. *commercial*), yksityis (engl. *private*) ja teollisuus (engl. *industrial*), kuukausittaisista dokumenttimääristä on kyse. Kappalemäärät on taulukossa ilmoitettu tuhansina (engl. *thousands*). Esimerkiksi helmikuussa 2012 (February / 2012) yritysliiketoimintayksiköllä (Commercial) oli dokumenttivarastossa 7 507 000 kappaletta dokumentteja. Taulukon kaksi viimeistä saraketta kertovat niiden dokumenttien määrän dokumenttivarastossa, joilla ei ole liiketoimintayksikön kertovassa attribuutissa mainintaa liiketoimintayksiköstä. Toiseksi viimeisessä sarakkeessa on niiden dokumenttien kappalemäärät, joilla on merkintä DK liiketoimintayksikköattribuutissa. Viimeisessä sarakkeessa on niiden dokumenttien kappalemäärät, joiden liiketoiminta-attribuutissa ei ole mitään merkintää eli attribuutti on tyhjä.



Kuva 6. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön EMC Documemtum infrastruktuurissa olevan dokumenttivaraston dokumenttien kokonaismäärän kehityksestä liiketoimintayksiköittäin.*

Liiketoimintayksiköstä kertovan metadata-attribuutin lisääminen raportointisääntöihin toi mukaan paljon lisää informaatiota eri liiketoimintayksiköiden dokumenttivarastossa olevien dokumenttien kokonaiskappalemääristä ja miten niiden määrät ovat kehittyneet kuukausittain. Nämä tiedot antavat jo riittävästi tietoa jakaa dokumenttivaraston operatiiviset kustannukset kulutusperustaisesti eri liiketoimintayksiköille, sekä tehdä tarkempia ennusteita dokumenttivaraston kehityksestä että liiketoimintayksiköiden erilaisista tilantarpeista. Vielä tarkempiin ennusteisiin ja kustannusjakoihin päästään, kun raportointisääntöihin otetaan mukaan lisää attribuutteja. Esimerkiksi sovelluksen kertovan metadata-attribuutin mukaan ottaminen mahdollistaa dokumenttivaraston operatiivisten kustannusten jakamisen yhden liiketoimintayksikön sisällä sovelluskohtaisesti niitä käyttäville organisaation osille; esimerkiksi kustannusten jakamisen osastoittain.

Dokumenttivarastoissa, joissa säilytetään massatulostettuja dokumentteja, ei ole käytettävissä yhtä monipuolista dokumenttien metadata-attribuutteja kuin muissa dokumenttivarastoissa. Massatulostettujen dokumenttien operatiivisia kustannuksia ei usein ole tarpeenkaan jakaa tarkasti eri liiketoimintayksiköille tai sovelluksilla. Tyypillisesti massatulostuksen tekee ulkopuolinen painotalo, jonka kanssa on tehty massatulostuksesta sopimus. Sopimuksessa on tyypillisesti määritelty massatulostukselle sivuhinta, joka kattaa kaikki tulostuksesta tulevat operatiiviset kustannukset. Sopimus kattaa kuukaudesta toiseen toistuvat vakiotulostukset kuten laskut, vakuutuskirjat ja vakuutusehdot. Muille

massatulosteille kuten mainos- ja markkinointikampanja materiaaleille voidaan tehdä eri sopimus tai ne voidaan hinnoitella tapauskohtaisesti. Vakiotulostuksien sivuhintaan sisältyy yleensä materiaalikustannukset, kuoritus, mahdolliset niittaukset ja tietty ennalta määritelty väripeitto.

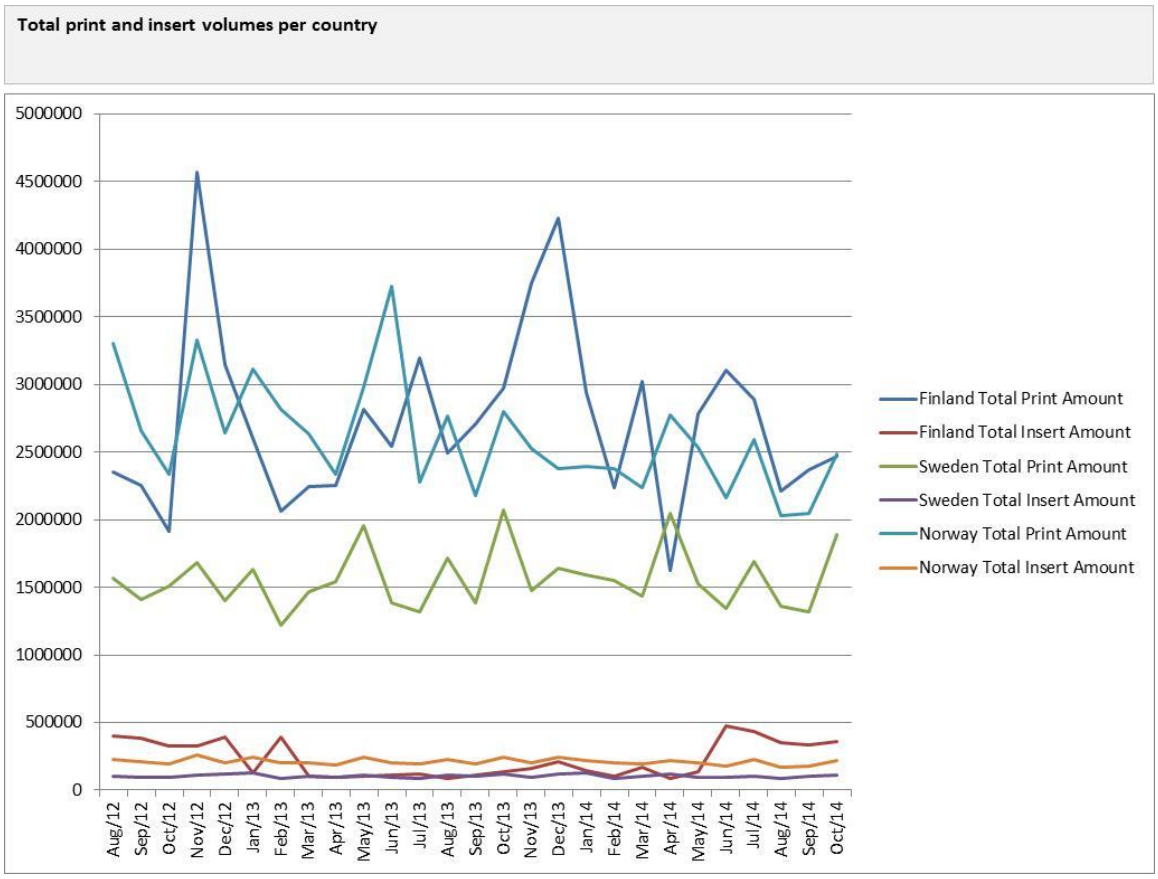
Kohdassa 4.2 massatulostuksen määrät ovat vakiotulosteiden kokonaismääriä ja sopimusmalli kustannuksille on sivuhinta sisältäen kuorituskustannukset. Kohdassa 4.2 raportointi ei anna riittävän tarkkaa kuvaa kustannusjaolle tai massatulostuksen ennustettavuudelle. Seuraavassa esimerkissä kohdan 4.2 raportointisääntöihin on lisätty metadata-attribuutti, joka kertoo maakoodin, jolla saadaan massatulostuksen määrät jaoteltuina maittain.

Taulukko 4. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön kuukausittaisista massatulostus määristä ja postitukseen käytettyjen kirjekuorien määristä jaoteltuina maittain.*

Month / Year	Finland Total Print Amount	Finland Total Insert Amount	Sweden Total Print Amount	Sweden Total Insert Amount	Norway Total Print Amount	Norway Total Insert Amount
August / 2012	2351793	400422	1570535	100714	3301558	225888
September / 2012	2257162	384820	1412090	95871	2660285	210640
October / 2012	1912600	326451	1507569	94714	2339229	193159
November / 2012	4569985	326636	1681202	114651	3332102	258975
December / 2012	3146465	391516	1398507	116717	2644301	202945
January / 2013	2601573	128419	1629804	124218	3111017	243428
February / 2013	2066046	389430	1218537	87993	2816535	204586
March / 2013	2249383	103331	1469797	100608	2632971	202951
April / 2013	2250950	98087	1544214	95960	2339663	184977
May / 13	2813365	105975	1956165	115268	2985411	241371
June / 2013	2545639	109910	1386125	94228	3728010	205724
July / 2013	3198255	119643	1322110	86506	2276216	192461
August / 2013	2490796	89396	1715715	111038	2769649	226477
September / 2013	2711673	111602	1388237	106111	2178323	194913
October / 2013	2975176	139845	2074491	122835	2796568	241405
November / 2013	3753331	157688	1477445	97299	2522863	206273
December / 2013	4227723	211782	1645785	120013	2375551	242821
January / 2014	2943947	144207	1589805	124822	2396935	222784
February / 2014	2241440	106338	1552183	90437	2375998	206577
March / 2014	3023668	172777	1436305	105368	2234764	194218
April / 2014	1625811	89448	2047760	115786	2772174	219359
May / 2014	2786391	137805	1526881	97057	2536519	200568
June / 2014	3102030	474202	1347756	93362	2160675	177633
July / 2014	2889913	432044	1690119	103317	2588936	226400
August / 2014	2208495	353247	1361531	82591	2032585	171480
September / 2014	2369269	333311	1316194	102189	2043723	181085
October / 2014	2471275	359653	1892767	111108	2482425	218387

Taulukon ensimmäisessä sarakkeessa Month /Year on tieto kuukaudesta (engl. *month*) ja vuodesta (engl. *year*), jonka tulostettujen A4-sivujen tulostus- (engl. *total print amount*)

ja kirjekuorimäärät (engl. *total insert amount*) esitetään seuraavissa sarakkeissa maittain Suomi (engl. *Finland*), Ruotsi (engl. *Sweden*) ja Norja (engl. *Norway*). Esimerkiksi syyskuussa 2012 (September / 2012) Suomeen (Finland) A4-sivuja tulostettiin 2 257 162 sivua. Kirjakuoria tulostettujen sivujen postittamiseen asiakkaille tarvittiin 384 820 kappaletta.



Kuva 7. *Pohjoismaissa toimivan vakuutusyhtiön kuukausittaiset massatulostusmäärät ja postitukseen käytettyjen kirjekuorien määrät kolmen vuoden ajalta maittain.*

Maakoodiattribuutin lisääminen raportointisääntöihin toi esille tietoa maiden välisistä eroista massatulostuksessa. Kuvasta 7 nähdään, että Suomessa massatulostetaan keskimäärin eniten ja Ruotsissa vähiten. Esimerkkinä käytetyssä vakuutusyhtiössä Tanskan massatulostus hoidetaan muista maista poikkeavasti vakuutusyhtiön sisällä, mistä syystä Tanskan luvut puuttuvat. Maatiedolla täydennetyistä raporteista nähdään, että tulostus määrät vaihtelevat kuukausittain paljon kaikissa Pohjoismaissa. Näin saadut tiedot auttavat tekemään tarkempia ennusteita massatulostuksesta sekä kohdistamaan massatulostusmäärien optimointi ja pienennystoimenpiteet sinne, mistä voidaan saavuttaa suurimmat hyödyt massatulostuskustannuksien optimoinnilla. Optimointia voidaan tehdä esimerkiksi uudistamalla tulostettavat dokumentit siten, että niiden sivumäärä pienenee, mutta tietosisältö pysyy samana.

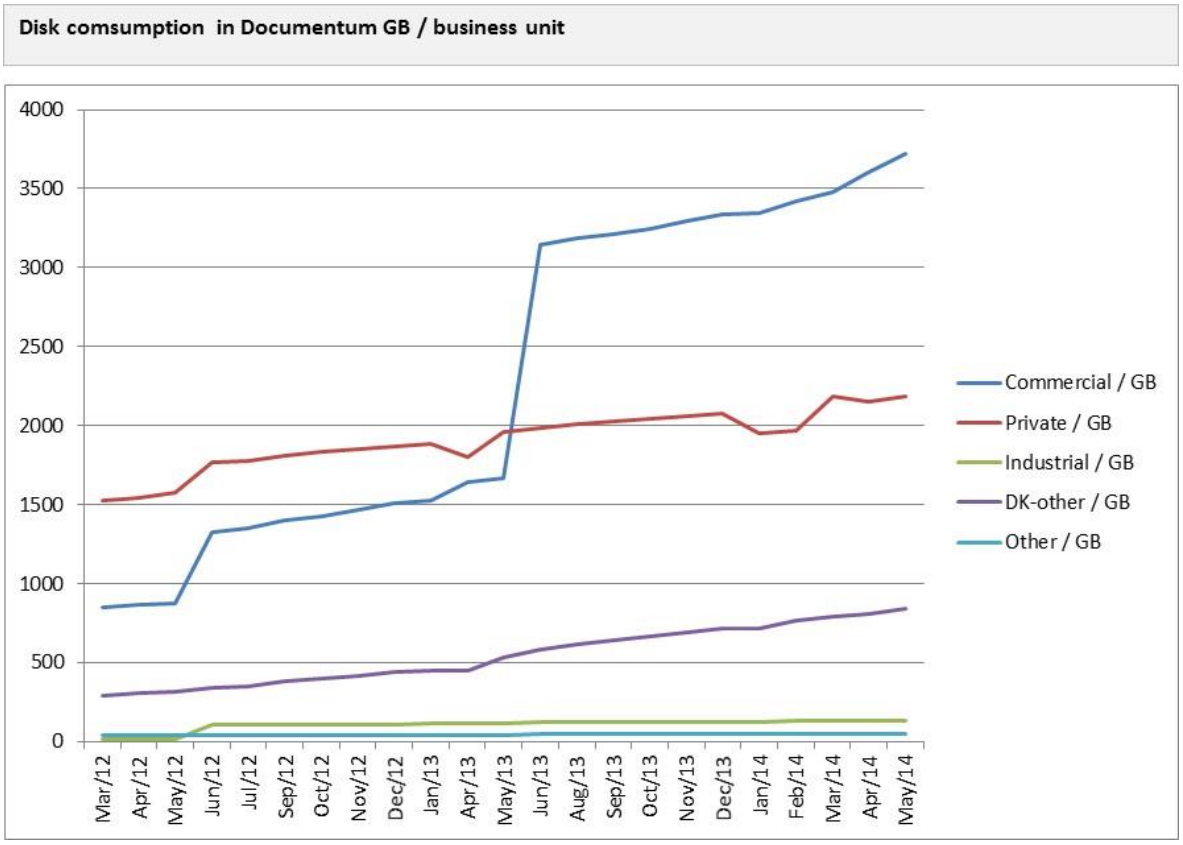
Ainoastaan dokumenttien metadataa hyväksikäyttäen ei saada levytilan käytöstä raportointia. Yhdistettynä infrastruktuurin monitoritietoihin tai dokumenttien hallintaohjelmistosta saataviin tietoihin levytilan käytöstä saadaan hyvä kuva. Tätä tietoa taas voidaan käyttää kohdistamaan levyihin liittyvät operatiiviset kustannukset kulutusperusteisesti eri liiketoimintayksiköille sekä levytilan hallintaan. Suuret dokumenttivarastot kuluttavat paljon levytilaan, joten niiden hallinta on vakuutusyhtiöille tärkeätä. Hyvä hallinta mahdollistaa myös automaattisen dokumenttien siirron erityyppisten levyjen välillä (engl. *auto -tiering tai automated tiered storage*). Seuraavassa on esimerkki levytilaraportista, jossa on yhdistetty dokumenttien metadatasta saatava liiketoimintayksikön tunnus ja dokumenttien hallintaohjelmistosta saatava levytilan kulutustieto.

Taulukko 5. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön yhden dokumenttivaraston levytilan käytöstä ja kasvusta liiketoimintayksiköittäin*

Month / Year	Commercial / GB	Private / GB	Industrial / GB	DK-other / GB	Other / GB
February / 2012	832	1513	17	278	42
March / 2012	848	1529	18	291	42
April / 2012	866	1547	18	307	42
May / 2012	876	1575	19	320	42
June / 2012	1329	1766	105	342	43
July / 2012	1348	1778	106	354	43
September / 2012	1398	1813	108	384	43
October / 2012	1429	1832	110	400	43
November / 2012	1465	1852	111	419	43
December / 2012	1506	1872	112	439	44
January / 2013	1529	1884	114	454	44
April / 2013	1641	1802	114	454	44
May / 2013	1670	1961	119	536	45
June / 2013	3147	1989	121	586	46
August / 2013	3189	2014	123	619	47
September / 2013	3214	2028	124	645	47
October / 2013	3246	2043	125	667	47
November / 2013	3292	2063	127	694	48
December / 2013	3333	2078	128	717	48
January / 2014	3348	1956	129	719	48
February / 2014	3419	1972	131	767	48
March / 2014	3475	2186	133	791	51
April / 2014	3607	2151	134	813	51
May / 2014	3722	2186	136	840	51

Taulukon ensimmäisessä sarakkeessa Month / Year on tieto kuukaudesta (engl. *month*) ja vuodesta (engl. *year*), jolta levytilan kulutukset ovat. Seuraavien sarakkeiden otsikossa kerrotaan minkä liiketoimintayksikön, yritys (engl. *commercial*), yksityis (engl. *private*) ja teollisuus (engl. *industrial*), kuukausittaisista levytilan kulutuksesta on kyse. Taulukossa levytila on ilmoitettu gigatavuina (GB). Esimerkiksi helmikuussa 2012 (February / 2012) yritysliiketoimintayksikön (Commercial) dokumentit veivät 832 GB levytilaa dokumenttivarastosta. Taulukon kaksi viimeistä saraketta kertovat niiden dokumenttien levynkulutuksen, joilla ei ole liiketoimintayksikön kertovassa attribuutissa mainintaa liiketoimintayksiköstä. Toiseksi viimeisessä sarakkeessa on niiden dokumenttien levynkulutus, joilla on merkintä DK liiketoimintayksikköattribuutissa. Viimeisessä sarakkeessa on

niiden dokumenttien levynkulutus, joiden liiketoiminta-attribuutissa ei ole mitään merkintää eli attribuutti on tyhjä



Kuva 8. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön EMC Documemtum infrastruktuurissa olevan dokumenttivaraston levyn käytöstä ja kasvusta liiketoimintayksiköittäin kolmelta vuodelta.*

Kuvassa 8 on esitetty taulukon 5 tiedot graafisessa muodossa. Kuvasta nähdään, että toukokuuhun 2013 asti yksityisliiketoimintayksikkö käytti dokumenttivaraston levytilaa enemmän kuin yritysliiketoimintayksikkö. Toukokuun 2013 tilanne muuttui ja yritysliiketoimintayksikön levytilan käyttö hyppäsi ylös rajusti ja ohitti selvästi yksityispuolen levynkulutuksen. Tätä tietoa jatkotutkimalla voidaan selvittää, mistä suuri hyppy yritysliiketoimintayksikön levytilakulutuksessa johtui ja miten entistä suurempaan levytilan kuukausikulutukseen voitaisiin vaikuttaa.

Yhteenvetona voidaan todeta, että dokumenttien metadataa hyödyntämällä sellaisenaan tai yhdistettynä muualta saatavaan tietoon, mahdollistaa dokumenttivarastojen seurannan ja raportoinnin ja operatiivisten kustannusten seurannan, hallinnan sekä jyvityksen kulutusperusteisesti dokumenttivaraston käyttäjille. Näin saadaan pohjatietoa, jota voidaan edelleen jatkojalostaa operatiivisten kustannusten ennustamiseen ja strategiseen liiketoiminnan ja IT-infrastruktuurin suunnitteluun.

5. METADATAN HYÖDYNTÄMINEN YRITYKSEN SISÄISESSÄ PÄÄTÖKSENTEOSSA JA SUUNNITTELUSSA

Tämän luvun tavoitteena on käsitellä dokumentteihin liittyvän standardoidun metadatan käyttömahdollisuuksia yrityksen sisäisessä toiminnassa. Käyttökohteet vaihtelevat tietoteknisestä infrastruktuurista yrityksen strategiseen suunnitteluun. Jokaisessa kohdassa käytetään esimerkkinä vakuutusliiketoimintaa ja vakuutusyrityksiä.

5.1 Liiketoimintastrategian suunnittelu ja seuranta

Kohdassa 5.1 esitellään, miten liiketoimintastrategian suunnitteluun voidaan dokumenttien metadatalla antaa tarkentavaa tietoa sekä miten suunnitelmien seurantaa voidaan toteuttaa metadatapohjaisella raportoinnilla.

Vakuutusyhtiöt tuottavat ja lähettävät asiakkailleen paljon erilaisia dokumentteja. Edelleen suuri määrä näistä asiakasdokumenteista tulostetaan paperille ja lähetään perinteisellä postilla asiakkaalle. Dokumenttien tuottamiseen yhtiöillä on erilaisia välineitä. Kokonaistavoitteena on tarjota asiakkaille mahdollisimman paljon tietoa mahdollisimman helposti ja mahdollisimman kustannustehokkaasti. Tämä edellyttää automaattisia toimintaprosesseja, tietotekniikkakulujen optimointia sekä uusia helppokäyttöisiä itsepalvelukanavia.

Kohdassa 4.2 käsiteltiin dokumenttien metadataan pohjautuvaa raportointia ja kohdassa 4.8 kerrottiin, miten dokumenttien metadataa hyväksikäyttämällä saadaan raportointia operatiivisista tietotekniikkakuluista. Kohdissa myöskin kerrottiin, miten tietotekniikkakustannuksia voidaan jakaa ja kohdistaa kustannuksien aiheuttajille käyttäen hyväksi säännöllistä raportointia. Tämä raportointi antaa hyvän lähtökohdan liiketoimintastrategian suunnittelulle ja seurannalle. Vakuutusyhtiöt haluavat siirtää asiointia yhä enemmän verkossa tarjottaviin itsepalvelukanaviin. Asiakkaat eivät juurikaan enää käy vakuutusyhtiöiden konttoreissa asioimassa vaan vakuutusten myynnit, turvamuutokset ja irtisanomiset sekä vahinkoilmoitukset hoituvat internetin kautta. Tämä ei kuitenkaan tarkoita, että asiakkaat, jotka käyttävät verkossa olevia itsepalvelukanavia olisivat paperittomia asiakkaita, joilla dokumentit lähettäisiin sähköisessä muodossa. Tällä hetkellä paperittomien asiakkaiden määrä kaikista vakuutusyhtiöiden asiakkaista on pieni. Esimerkiksi tässä työssä käytettävän Pohjoismaissa toimivan vakuutusyhtiön paperittomien asiakkaiden määrä on niin pieni, että sillä ei ole vielä ollut merkittävää vaikutusta massatulostuksen kuukausittaisiin tulostusmääriin tai käytettyihin kirjakuorimääriin. Tästä syystä yhti-

össä on yhdistetty dokumenttien metadatalla tuotettujen raporttien tiedot muualta saatavaan tietoon ja mietitty, miten paperittomien asiakkaiden määrää saataisiin kasvatettua ja vastaavasti pienennettyä tulostusmääriä sekä kustannuksia. Tietojen analyysin jälkeen on määritelty kaksi liiketoimintastrategiaa; tulostusmääriä tulee pienentää yksinkertaistamalla käytettäviä määrämuotoisia dokumentteja sekä tarjottava ja rakennettava asiakkaille parempia palveluita verkon kautta käytettäväksi. Näin uskotaan päästävän tilanteeseen, jossa valtaosa vakuutusyhtiön asiakkaista olisi paperittomia ja paperillisille asiakkaille lähettäisiin ymmärrettävämpiä selkokielisiä dokumentteja. Tavoitteena on myös alentaa liiketoiminnan tietotekniikkakuluja ja saavuttaa kilpailuetua markkinoilla. Tämän strategian seurantaan voidaan käyttää samaa dokumenttien metadataan pohjautuvaa raportointia, jota liiketoiminta käytti määritellessään strategiansa. Seurantaan voidaan käyttää raportointia kuukausittaisista tulostusmäärien ja talletustilan käytöstä ja niihin liittyvistä ennusteista.

Kirjoittajan tietojen mukaan kaikilla vakuutusyhtiöillä on edellä kuvatun kaltaisia suunnitelmia. Paperittomien asiakkaiden määrän kasvatuksen esteenä voi olla erilaiset maakohtaiset säädökset. Esimerkiksi maassa voi olla säädös, joka edellyttää vakuutusehtojen lähettämistä asiakkaille paperilla. Kirjoittajan mielestä suurin este paperittomien asiakkaiden määrän kasvulle on kannustimien puute. Siirtyessään paperittomaan palveluun asiakas ei saa siitä hyötyä pienempinä palvelumaksuina tai vakuutusmaksuina.

5.2 Operatiivisten kustannusten suunnittelu ja seuranta

Tässä kohdassa 5.2 käydään läpi, miten dokumenttien metadataa hyväksi käyttäen voidaan suunnitella, seurata ja hallita tietotekniikan operatiivisia kustannuksia.

Tyypillisesti yrityksen ovat rakentaneet omia konesaleja tietovarastoineen tai ulkoistaneet konesalipalvelut kotimaisille konesalipalveluita tuottaville yrityksille. Viime vuosina erilaiset pilvessä (engl. *cloud*) olevien talletuspaikkojen käyttö on voimakkaasti yleistynyt. Useimmiten yritykset ovat käyttäneet yksityisiä pilvipalveluita (engl. *private cloud*), joita aikaisemmin vain konesalipalveluita tarjonneet yritykset ovat rakentaneet. Nämä palvelut ovat rakennettu olemassa olevien konesalien varaan ja ne on rakennettu ensisijassa yrityskäyttöön ottaen huomioon yritysten vaatimukset käytölle ja tietoturvalle. Julkiset pilvipalvelut (engl. *public cloud*) on rakennettu mahdollisimman kustannustehokkaiksi, jotta kuka tahansa voisi niitä käyttää. Tästä johtuen niissä on usein tingitty tietoturvan tasosta, käyttäjähallinnasta ja varmistuspalveluista (engl. *backup services*).

Yrityksien kannalta pilvipalveluiden käyttö edellyttää riittävää tietoturvatasoa palveluilta sekä hyvää käytönhallintaa. Usein yritykset myös vaativat, että tiedot talletetaan tiettyyn maatieteelliseen paikkaan. Tyypillinen vaatimus on, että pilvipalvelun toimittajan konesalit sijaitsevat Euroopassa tai että yrityksen tietoja ei saa tallettaa pilvipalvelun toimittajan Yhdysvalloissa sijaitseviin konesaleihin. Sopimusehtoja talletuspaikasta yrityksen on

kuitenkin vaikea seurata, koska seuranta edellyttää tietojärjestelmien monitorointityökalujen käyttöä omalta henkilökunnalta tai monitorointipalveluiden ostamista ulkopuoliselta toimittajalta. Usein monitorointi tehdään kerran, jotta saadaan varmistus, että sopimuksen pykälät täyttyvät. Tämä ei kuitenkaan riitä vaan monitorointi tulisi olla jatkuvaa tai määritellyn väliajoin toistuvaa, jotta voidaan varmistaa, että pilvipalveluntoimittaja ei riko sopimusehtoja vahingossa tai tahallaan.

Dokumentteja, jotka ovat elinkaarensa lopussa, ei kannata säilyttää nopeilla ja korkeiden kustannusten talletusmedioilla. Tällaiset dokumentit kannattaa arkistoida mahdollisimman halvalle talletusmedialle, josta dokumentit saadaan haettua tarvittaessa, mutta dokumenttien haulle ei ole asetettu vasteaikavaatimuksia. Jos vasteaika- tai hakuaikavaateita on määritelty, niin ne ovat pitkiä. Joissakin tapauksissa voidaan hyväksyä jopa päivien pituisia hakuaikoja. Yleensä tällaisissa tapauksissa pitää ennen varsinaista dokumenttien hakua täyttää hakupyyntölomake, joka sitten lähetetään arkiston pitäjälle. Hakupyyntölomakkeen saatuaan arkiston ylläpitäjä tarkistaa lomakkeen ja tekee varsinaisen dokumenttihaun käyttäen tarvittavia dokumentin metadata-attribuutteja. Haussa löytyneet dokumentit sitten lähettää pyytäjälle. Tyypillisesti tätä prosessia käytetään pitkäaikaisarkistojen yhteydessä.

Operatiivisten kustannusten suunnittelussa ja seurannassa voidaan käyttää dokumenttien metadatasta ajettuja raportteja kertomaan dokumenttivarastojen koosta ja dokumenttien kokonaismäärän kasvuvauhdista. Yhdistämällä näiden raporttien tiedot kertyneisiin kustannuksiin ja veloituksiin voidaan asettaa operatiivisille kustannuksille tavoitteita ja seurata niiden toteutumista. Dokumenttivarastojen kustannukset muodostuvat tyypillisesti talletustilaveloituksista ja infrastruktuurikustannuksista kuten konesali- ja palvelinkustannukset. Näitä kustannuksia pyritään minimoimaan ja optimoimaan operatiivisten kustannusten suunnittelulla ja seurannalla.

Kuva 9. *Esimerkki Pohjoismaissa toimivan vakuutusyhtiön talletustilan käytöstä ja siihen liittyvistä kustannuksista.*

Kuvassa 9 on esitetty, miten vakuutusyhtiön dokumenttivarastojen talletustila ja kustannukset ovat kehittyneet viidessä vuodessa. Pystyakselilla vasemmalla on kokonaistalletustila esitetty gigatavuina (GB). Pystyakselilla oikealla on esitetty kokonaistalletustilasta aiheutuneet kustannukset miljoonina Ruotsin kruunuina (SEK). Kustannukset eivät ole euroissa, koska esimerkki vakuutusyhtiössä kaikki kustannukset on päätetty esittää Ruotsin kruunuina yhtenäisyyden vuoksi. Kuvasta nähdään, että tallennuskustannukset ovat alentuneet samaan aikaan, kun talletettavien dokumenttien kokonaismäärä on kasvanut.

Kuvasta käy myös ilmi, että loppuvuonna 2012 oli taitepiste, jonka jälkeen talletuskustannukset ovat pysyneet huomattavasti alemmalla tasolla kuin ennen vuotta 2012. Syynä tähän on operatiivisten kustannusten parempi suunnittelu ja seuranta. Työkaluina on käytetty dokumenttivarastoista saatavaa raportointia yhdistettynä IT-infrastruktuurista saatavaan monitorointitietoon sekä aloitettu tietojenluokitteluun perustuva tietojen elinkaaren hallinta.

Yhteenvetona voidaan todeta, että dokumenttien metadatan avulla tuotetulla raportoinnilla yhdistettynä IT-infrastruktuurista saatavaan tietoon ja kustannustietoon voidaan nähdä painopistealueita ja ongelmakohtia. Näihin voidaan puolestaan vaikuttaa liiketoiminnan ja tietotekniikkainfrastruktuurin suunnittelulla.

5.3 IT-strategian suunnittelu ja seuranta

Kohdassa 5.3 esitellään lukijalle, miten ja mitä vaikutuksia dokumenttien metadatalla on IT-strategiaan ja -infrastruktuuriin.

Liiketoiminta näkee useimmiten tietotekniikan välttämättömänä pahana, jonka kustannukset ovat korkeita. Liiketoiminta haluaa hyviä tietotekniikkapalveluita halpaan hintaan. Tämä liiketoiminnanvaatimus asettaa reunaehdot IT-strategian suunnittelulle ja myöskin seurannalle. Jotta nämä vaatimukset saavutettaisiin, tietotekniikassa tulee hakea uusia kustannustehokkaampia ratkaisuja sekä infrastruktuurissa että sovelluksissa. Dokumenttien metadatan avulla saatavasta raportoinnista nähdään dokumenttien määrän kasvuvauhti ja kasvunopeus. Näiden tietojen pohjalta voidaan arvioida tulevaa talletustilan tarvetta ja niihin liittyviä kustannuksia. Nämä tiedot yhdistettynä liiketoiminnan kustannusvaatimuksiin ja muihin, vaatimukset antavat hyvät pohjatiedot IT-strategian suunnitteluun sekä infrastruktuurin tavoitetilan määritykseen. Esimerkkinä käyttämässämme vakuutusyhtiössä tämä on tarkoittanut IT-strategiaa, jossa tietotekniikkainfrastruktuurista on poistettu päällekkäisiä dokumenttivarastoja yhdistämällä ja poistettu päällekkäisiä sovelluksia. Näin toimien yhtiö on voinut pienentää dokumenttivarastoihin ja sovelluksiin liittyviä operatiivisia kustannuksia, niihin liittyviä laite- ja lisenssikustannuksia sekä vuosittaisia ylläpitomaksuja jo useana peräkkäisenä vuotena.

Dokumenttien metadatalla tuotettu raportointi antaa siis IT-strategian tekijöille pohjatietoa, jota hyväksikäyttämällä voidaan laatia ennusteita ja suunnitella muutoksia sekä sovellus-, tietovarasto- että laiteinfrastruktuuriin.

5.4 Projektisalkun suunnittelu ja seuranta

Kohdassa käydään läpi niitä mahdollisuuksia, joita dokumenttien metadatan käyttö antaa projektisalkun suunnitteluun ja hallintaan.

Dokumenttien metadatalla ei ole suoraa yhteyttä projektisalkun suunnitteluun, mutta selvä välillinen yhteys on olemassa. Tämä välillinen yhteys syntyy liiketoiminta ja IT-strategian kautta. Näiden strategioiden yhtenä tietolähteenä ovat olleet dokumenttien metadata pohjaiset raportit. Projektisalkku taas puolestaan suunnitellaan siten, että nuo määritellyt strategiat toteutuisivat suunnitellusti. Projektisalkun ja sen sisältämien projektien tavoitteiden toteutumista voidaan seurata käyttäen joko yleisiä tai projektille räätälöityjä raportteja dokumenttivarastoista.

5.5 Metadatan hyödyntäminen osana tietoturvaa

Kohdan tavoitteena on esitellä lukijalle, miten dokumentteihin liittyvää metadataa voidaan käyttää tietoturvan näkökulmasta. Haluttu tietoturvan taso asettaa vaatimuksia metadatalle. Tässä kohdassa esitellään näitä vaatimuksia. Lopuksi annetaan esimerkkejä vakuutusliiketoiminnan tietoturvasta.

Tietoturvan tehtävänä on pitää huolta, että vakuutusyhtiöt täyttävät vakuutusliiketoiminnalle määritetyt tietoturvavaatimukset ja -määräykset. Lisäksi tietoturvan tehtävänä on testata toteutettujen ratkaisujen toimivuus suorittamalla erilaisia tietoturvaan liittyviä testejä. Tällaisia testejä ovat mm. erilaiset murtotestit, identiteettivarkaustestit, palvelunestohyökkäysten testit ja disaster recovery -testit. Tietoturvan tehtävänä on myös kehittää ja ylläpitää ohjeistoa siitä, miten toimia tietoturvallisesti ja miten toipua erilaisista odottamattomista tapahtumista kuten esimerkiksi kovalevyjen rikkoutumisesta tai asiakastietokannan korruptoitumisesta.

Tietoturvan vaatimukset dokumenttivarastoille ja dokumenttien metadatalle liittyvät käyttäjätunnistukseen, tietojenluokitteluun, omistajuuteen ja jäljitettävyyteen. Seuraavissa kohdissa käydään tarkemmin läpi käyttäjätunnuksiin liittyviä asioita ja millaisia attribuutteja voi dokumenttien metadatassa olla liittyen tietoturvaan.

5.5.1 Käyttäjien tunnistus ja oikeudet

Alakohdan 5.5.1 tavoitteena on antaa lukijalle käsitys dokumenttien metadatasta saatavasta lisäarvosta käyttäjähallintaan. Sen sijaan varsinaisia teknisiä tietoturvaratkaisuja tai tuotteita ei käsitellä.

Varsinainen käyttäjätunnistus (engl. *authentication*) ja valtuutus (engl. *authorization*) tehdään ennen kuin tunnistettu käyttäjä pääsee käsiksi dokumenttivarastojen tietoihin. Tämä käyttäjätunnistus tehdään sekä sisäisille käyttäjille ja ulkoisille käyttäjille. Tässä työssä sisäisillä käyttäjillä tarkoitetaan vakuutusyhtiöiden työntekijöitä ja ulkoisilla käyttäjillä tarkoitetaan vakuutusyhtiöiden asiakkaita, jotka käyttävät erilaisia itsepalvelukanavia. Dokumenttivaraston näkökulmasta katsottuna nämä kaksi käyttäjäryhmää näkyvät eri tavalla ja myöskin vaikutukset dokumenttien metadatan sisältöön ovat erilaiset.

Sen jälkeen, kun ulkoinen käyttäjä on tunnistettu ja todettu vakuutusyhtiön asiakkaaksi, varsinainen dokumenttien haku ja katselu vakuutusyhtiön sisäverkossa olevasta dokumenttivarastosta tapahtuu käyttäen yhtä teknistä käyttäjätunnusta. Näin vältetään tilanne, jossa vakuutusyhtiön asiakkaille tulisi rakentaa oma käyttäjähallinta. Teknisten käyttäjätunnusten käyttö aiheuttaa sen, että dokumenttien metadataa ei päivitetä dokumentteja haettaessa, koska käyttäjäksi metadataan tulisi aina sama tekninen tunnus. Lisäksi lähes aina ulkoisilla käyttäjillä on vain haku- ja lukuoikeudet. He eivät siis voi muokata varastossa olevia dokumentteja ja näin aiheuttaa metadatan päivitystarpeita. Sen sijaan kaikki teknisen käyttäjätunnuksen käyttökerrat kirjataan dokumenttivaraston sisäisiin lokitiedostoihin sekä myös käyttöjärjestelmän lokitietoihin. Näin täytetään tietoturvan vaatimus jäljitettävyydestä.

Dokumenttien metadatan kannalta mielenkiintoisempi käyttäjäryhmä ovat sisäiset käyttäjät. Sisäiset käyttäjät jakautuvat osaryhmiin sen mukaan, millä liiketoiminta-alueella (yksityis-, yritys- ja teollisuusala) he toimivat sekä millaisiin tietoihin heillä on pääsy- ja käsittelyvaltuudet. Sisäisistä käyttäjistä ylläpidetään käyttäjähallintaa, jossa on käyttäjätunnuksia, käyttöoikeustietoja ja käyttäjäryhmätietoja. Sisäinen käyttäjä tunnistetaan ja hänelle annetaan käyttöoikeudet käyttäen tätä käyttäjähallinnan käyttäjärekisteriä. Näitä tietoja käytetään hyväksi haettaessa ja talletettaessa dokumentteja dokumenttivarastoista. Myöskin sisäiset käyttäjät keskustelevat dokumenttivaraston kanssa käyttäen teknisiä käyttäjätunnuksia, mutta sisäiset käyttäjät myös muokkaavat ja täydentävät jaettuja dokumentteja. Tallennettaessa näitä dokumentteja niiden metadataa muokataan ja päivitetään. Metadataan lisätään uusia aikaleimoja ja versiotiedot päivittyvät, uusia versioita syntyy uusilla metadatoilla, käsittelijän tunnukset tallentuvat dokumentteja käsiteltäessä sekä dokumentin omistajatieto voi vaihtua. Sisäisestä käyttäjästä ja tehdyistä toimenpiteistä jää siis paljon jälkiä dokumenttien sisältämään metadataan.

Yhteenvetona voidaan sanoa, että ulkoiset käyttäjät eivät jätä jälkiä dokumenttivarastojen dokumenttien metadataan, koska he käyttävät teknisiä käyttäjätunnuksia yhdistettynä rajoittuneisiin käyttövaltuuksiin. Nämä käyttövaltuudet ovat tyypillisesti haku- ja katseluoikeuksia ilman muokkausmahdollisuutta. Tietoturvan kannalta enemmän tietoa dokumenttien metadata-attribuutteihin jää sisäisistä käyttäjistä, joista pidetään yhtiön sisäistä käyttäjä- ja valtuusrekisteriä, minkä tietoja käytetään dokumenttien metadatoja päivitettäessä. Sisäiset käyttäjät voivat valtuuksiensa rajoissa hakea, lukea ja muokata dokumenttivarastojen dokumentteja. Nämä tiedot tuovat lisäarvoa tietoturvaan lisäten jäljitettävyyttä sekä voivat ennaltaehkäistä henkilökunnan tekemiä väärinkäytöksiä ja auttaa tehtyjen väärinkäytösten selvittämiseen.

5.5.2 Tietoturvan vaatimukset dokumenttien metadatalle

Alakohdan 5.5.2 tavoitteena on antaa lukijalle käsitys tietotuvan vaatimuksista dokumenttien metadatan sisältöön ja attribuutteihin. Lisäksi kerrotaan, miten vakuutusyhtiön tietoturvassa on käytetty hyväksi dokumenttien metadataa.

Tietoturvan tehtävänä on varmistaa tietojen turvallinen säilytys, käsittely sekä jäljitettävyys että väärinkäytösten ehkäisy. Näistä tehtävistä johtuen dokumenttien metadatalle tulee vaatimuksia. Vaatimukset liittyvät kiinteästi tietojen ja dokumenttien luokitteluun sekä dokumenttien tyypitykseen. Tietojen ja dokumenttien luokittelusta saadaan suoraan attribuutteja ja niiden arvoja dokumenttien metadataan. Yleisimmin käytettyjä tietoturva attribuutteja dokumenttien metadatassa ovat saatavuus (engl. *availability*) ja luottamuksellisuus (engl. *confidentiality*). Nämä metadata-attribuutit saavat erilaisia arvoja, jotka tulevat dokumenttien ja niiden sisältämän tiedon luokittelusta. Dokumenttien saatavuudelle usein käytettyjä arvoja ovat seuraavat:

- yleinen (engl. *common*)
- rajoitettu (engl. *restricted*)
- salainen (engl. *secret*)

Dokumenttien luottamuksellisuudelle usein käytettyjä arvoja ovat seuraavat:

- julkinen (engl. *public*)
- luottamuksellinen (engl. *confidential*)
- ehdottoman luottamuksellinen (eng *strictly confidential*)

Edellä olevilla kahdella dokumentin metadata-attribuutilla ja niiden yhdistelmillä päästään jo hyvään tietoturvatasoon. Kolmas metadatan attribuutti, jota käytetään yhdessä saatavuuden ja luottamuksellisuuden kanssa on omistajuuden kertova attribuutti. Omistajuuden määrittäminen on usein vaikeata, mutta välttämätöntä, jotta dokumentit voidaan luokitella oikein ja määritellä niille oikeat arvot dokumenttien metadataan.

Vakuutusyhtiöille riittävään tietoturvatasoon dokumenttien metadatassa päästää käyttämällä edellä esitettyjä kolmea attribuuttia (omistajuus, saatavuus ja luottamuksellisuus). Käytännössä näiden attribuuttien määritys on ensimäistä kertaa tehtäessä työlästä. Tämä johtuu siitä, että usein ensimmäisellä kerralla joudutaan tekemään koko materiaalin läpikäynti kokonaan tai osittain manuaalisesti. Tämä työ on kuitenkin välttämätöntä pohjatyötä, jonka perusteella sitten voidaan laatia automatiikan vaatimat käsittelysäännöt. Juuri tämä manuaalisen pohjatyön suuruus on rajoittanut vakuutusyhtiöissä laajamittaista tietoturva-attribuuttien käyttöönottoa dokumenttivarastoissa. Kirjoittajan tietojen mukaan attribuutit on otettu tai ollaan ottamassa käyttöön asiakasdokumentteja sisältävissä dokumenttivarastoissa. Sen sijaan dokumenttivarastoissa, jotka sisältävät vakuutusyhtiöiden sisäiseen toimintaa liittyviä dokumentteja, ei tietoturva-attribuutteja ole dokumenttien metadatassa otettu käyttöön eikä suunnitelmia tilanteen nopeaan muuttamiseen ole.

Yhteenvetona voidaan todeta, että tietoturvan vaatimukset dokumenttien metadataan liittyvät saatavuuteen, luottamuksellisuuteen ja omistajuuteen. Nämä kolme attribuuttia riittävät vakuutusyhtiöille. Saatavuus ja luottamuksellisuus attribuuteille on määritelty erilaisia arvoja perustuen dokumenttityyppiin ja sen sisältämään tietoon. Käytännössä näillä

kahdella attribuutilla ja niiden erilaisilla arvoilla päästään hyvään tietoturvatason. Huomionarvoista on myös se, että saatavuus-, luottamuksellisuus- ja omistajuusattribuuttien käyttö ei ainoastaan rajoitu vain dokumenttien metadataan, vaan näitä attribuutteja voidaan käyttää muunkin tyyppisen datan metadatassa luokitteluun ja tietoturvaan.

6. YHTEENVETO

6.1 Yhteenveto työstä ja työn tuloksista

Työssä käsitellään dokumenttien metadataa ja sen hyödyntämistä antamalla käytännön esimerkkejä vakuutusliiketoiminnasta, josta kirjoittajalla on usean vuoden kokemus.

Työssä käydään läpi, mitä dokumenttien metadatalla tarkoitetaan ja miten laadukasta metadataa saadaan. Laadukkaan metadatan saaminen edellyttää kerättävän tiedon standardointia. Tämä standardointi tehdään määrittelemällä metadatamalli, jossa määritellään metadatalle rakenne ja eri attribuuttien väliset suhteet. Metadatamalli on siis hierarkkinen esitys dokumenttien metadatasta. Metadata ja metadatamalli ovat aina sidoksissa liiketoimintaan ja sen tarpeisiin. Tästä syystä määritelty metadatamalli tulee yhtenäistää. Yhtenäistämisessä metadatamallista poistetaan päällekkäiset samaa tietoa sisältävät attribuutit sekä yhtenäistetään attribuutteihin laitettavan tiedon rakenne ja sisältö. Kuten työssä todetaan, päällekkäisten attribuuttien poistaminen on melko helppo tehtävä, mutta attribuutteihin laitettavan tiedon yhtenäistäminen on vaikeaa, joten sitä ei useinkaan tehdä. Työssä selvitetään erilaisia tapoja kerätä dokumentteihin liittyvää metadataa. Työssä nähdään, että paras lopputulos saadaan, kun metadata luodaan mahdollisimman pitkälti automaattisesti ilman, että metadataa kysytään käyttäjiltä. Näin metadatan laatu ja eheys pysyvät korkealla tasolla. Työssä myös todetaan, että käyttäjiltä voidaan tarvittaessa kysyä joitakin metadata attribuuttien tietoja.

Työssä analysoidaan miten dokumenttien metadataa voidaan käyttää dokumenttien elinkaaren hallintaa, joka on ehkä eniten käytetty ja tunnetuin käyttökohde dokumenttien metadatalle. Työssä selvitetään dokumentin elinkaaren vaiheet ja niihin liittyvät metadatat. Huomiota kiinnitetään dokumenttien arvottamiseen liiketoiminnan kannalta ja siihen, miten liiketoiminnan tarpeet ohjaavat dokumenttien siirtymistä eri elinkaaren vaiheiden välillä. Työssä huomiota kiinnitetään dokumenttien elinkaaren loppupäähän. Erityisesti huomioita kiinnitetään dokumenttien pysyvään tuhoamiseen, joka usein laiminlyödään. Työssä painotetaan dokumenttien tuhoamista metadataan perustuvalla automaatiolla sekä annettaan esimerkki toimivaksi ratkaisuksi.

Työssä analysoidaan miten dokumenttien metadataa voidaan käyttää sellaisenaan tai yhdistettynä muualta saatavaan tietoon liiketoiminnan suunnittelussa, tietotekniikan ohjauksessa ja suunnittelussa sekä tietoturvassa. Työssä todetaan, että dokumenttien metadatalla saatavalla säännöllisellä raportoinnilla voidaan saada selville mm. kuinka dokumenttivarastot kasvavat ja tehdä ennusteita lähitulevaisuuden talletustilan tarpeesta ja kustannusten kehittymisestä. Näitä tietoja taas puolestaan voidaan käyttää hyväksi suunniteltaessa tietotekniikan operatiivisten kustannusten jyvittämistä käyttöperusteisesti liiketoiminta-

organisaatiolla tai suunniteltaessa muutoksia tietotekniseen infrastruktuuriin. Työssä käsitellään miten dokumenttien ja erityisesti valokuvien metadataan voidaan käyttää erilaisten väärinkäytösten ja petosten selvityksessä vakuutusliiketoiminnassa. Dokumenttien metadata yhdistettynä muualta saatavaan tietoon antaa liiketoiminnalla hyvää taustietoa sen suunnitellessa uusia tuotteita ja palveluita asiakkailleen.

Työssä käydään läpi erilaisia tiedostoformaatteja ja niiden käyttökohteita. Työssä todetaan, että yrityksien henkilökunnan osaaminen ja yrityksen panostukset toimisto-ohjelmistoihin ohjaavat voimakkaasti mitä formaatteja käytännössä yritysmaailmassa käytetään. Tämä puolestaan hidastaa uusien tiedostoformaattien käyttöönottoa vaikka niistä saatavat edut olisivat nähtävissä ja oleellisia yritykselle pidemmällä tähtäimellä.

6.2 Tulevaisuus

Yrityksillä on useita dokumenttivarastoja, joissa on talletettuna asiakasdokumentteja ja yrityksen sisäiseen toimintaan liittyviä dokumentteja. Nämä dokumenttivarastot kasvavat ja niihin talletetut dokumentit monipuolistuvat ja muuttuvat tulevaisuudessa. Tämä tarkoittaa, että näiden varastojen tehokas hallittavuus ja kustannustehokashoito vaativat, että yritysten tulee tietää, mitä varastoissa on, mikä on talletettujen dokumenttien arvo yrityksen liiketoiminnalle ja miten niitä voidaan hyödyntää. Nämä tiedot yrityksien on mahdollista saada dokumenttien metadataa hyödyntämällä. Lisäksi yrityksien tavoitteet yhdistää muualta saatava tieto yrityksen sisällä olevaan tietoon lisäävät dokumenttien metadatan tärkeyttä. Kirjoittajan mielestä yritykset tulevat tulevaisuudessa hyödyntämään entistä enemmän dokumenttivarastojaan. Niiden hyödyntämisen edellytyksenä on, että dokumenttien metadata on kunnossa ja sitä voidaan käyttää antamaan pohjatietoja yrityksien päätöksentekoon ja liiketoiminnalle. Dokumenttivarastoissa olevien dokumenttien tiedostoformaatit ja rakenteet tulevat muuttumaan tulevaisuudessa. Tulevaisuudessa dokumenttivarastoihin talletetaan enenevässä määrin mm. videota, ääntä ja sosiaalisenmedia sisältöjä. Näihin liittyvät metadatat tulee ottaa haltuun. Tässä työssä olevat asiat eivät ole ainoastaan dokumentteihin liittyviä vaan niitä voidaan soveltaa ja käyttää minkä tahansa tyyppisen tiedon metadataan.

6.3 Jatkotutkimuskohteet

Metadatalla itsellään on ominaisuuksia, joista joitakin luetellaan työssä, mutta joita ei sen paremmin käsitelty. Nämä ominaisuudet sekä niiden vaikutus ja hyväksikäyttö dokumenttien metadatassa vaatisi jatkotutkimuksen.

Metadataan perustuvan sääntöpohjaisen automaation käyttö dokumenttien hallinnassa ja kokoamisessa on mielenkiintoinen aihe, jolla voisi olla monia käytännön soveltamisalueita. Esimerkiksi automaatiota voitaisiin käyttää entistä paremmin ja enemmän dokumenttien personoinnissa tai työssäkin mainitussa dokumenttien pysyvässä tuhoamisessa. Tutkimuksessa voisi paljastua uusia ennalta arvaamattomia soveltamiskohteita.

Yksi mielenkiintoinen jatkotutkimuskohde voisi olla tutkia digitalisaation ja robotisaation vaikutus dokumentteihin ja dokumentaatioon. Poistaako robotisaatio tarpeen nykyisen kaltaiseen ja laajuiseen dokumentointiin vai tuleeko tilalle tarve uudentyyppiseen dokumentaatioon metadatoineen.

LÄHTEET

[1] Adobe (2015), PostScript Language Reference Manual third edition. Saatavissa https://www.adobe.com/products/postscript/pdfs/PLRM.pdf . Viitattu 4.7.2015.

[2] AFP Consortium (2015). Saatavissa http://afpcinc.org/ . Viitattu 4.7.2015.

[3] Alliance Group (2015), PDF file format. Saatavissa http://www.alliancegroup.co.uk/pdf.htm . Viitattu 6.8.2015.

[4] Arkistolaitos (2013), Arkistolaitoksen Arkistowiki. Saatavissa http://wiki.narc.fi/arkistowiki/index.php/Etusivu . Viitattu 10.5.2015.

[5] Arkistolaitos (2006), Asiakirjahallinnon ja arkistotoimen oppikirja. Saatavissa http://www.arkisto.fi/uploads/Palvelut/Julkaisut/asiakirjahallinnon_oppikirja.pdf . Viitattu 10.5.2015.

[6] Arkistolaitos (2015), Keskeisiä käsitteitä. Saatavissa http://www.arkisto.fi/normit/arkistolaitoksen-suositus-arkistonmuodostussuunnitelmaan-termit/ . Viitattu 10.5.2015.

[7] Baldeón, Jorge (2010), Data preservation management: Long-term data preservation in a Nordic insurance company. Turun kauppakorkeakoulu, Information Systems Science. Viitattu 5.5.2015.

[8] Digital Preservation Org (2002), Exif Exchangeable Image File Format Version 2.2. Saatavissa http://www.digitalpreservation.gov/formats/fdd/fdd000146.shtml . Viitattu 5.8.2015.

[9] Euroopan komissio (2015), Oikeus, Tietosuoja, Henkilötietosuoja. Saatavissa http://ec.europa.eu/justice/data-protection/index_fi.htm . Viitattu 3.6.2015.

[10] Fermilab (2015), PostScript lauguage reference manual, http://www-cdf.fnal.gov/offline/PostScript/PLRM2.pdf . Viitattu 4.7.2015.

[11] Gartner (2015), Digitalization, IT Glossary. Saatavissa http://www.gartner.com/it-glossary/digitalization . Viitattu 10.6.2015.

[12] Gartner (2015), IT Glossary, Big Data. Saatavissa http://www.gartner.com/it-glossary/big-data/ . Viitattu 30.8.2015.

[13] Hoppu Esko, ja Hemmo Mika, (2006), Vakuutusoikeuden opintomoniste. Helsingin yliopisto. Saavissa http://www.helsinki.fi/oikeustiede/oppiaineet/velvoiteoikeus/materiaali/Hoppu_Hemmo2006_Vakuutusoikeuden-perusteet.pdf . Viitattu 5.5.2015.

[14] International Press Telecommunications Council (2010), IPTC – core standard. Saatavissa https://www.iptc.org/std/photometadata/documentation/GenericGuidelines/index.htm#!Documents/iptccore.htm . Viitattu 10.7.2015.

[15] ISO (2006), Open Document Format for Office Application, ISO/IEC 26300:2006. Saatavissa http://www.iso.org/iso/catalogue_detail?csnumber=43485 . Viitattu 5.8.2015.

[16] ISO (2008), PDF format becomes ISO standard http://www.iso.org/iso/news.htm?refid=Ref1141 . Viitattu 4.7.2015.

[17] Kansallinen digitaalinen kirjasto –hanke, (2009), Digiwiki. Saatavissa http://www.digiwiki.fi/fi/index.php?title=EXIF . Viitattu 10.7.2015.

[18] Laine, Kari, Outbound customer communication – target view, P&T Privete Strategy Unit, If. Viitattu 27.6.2015.

[19] Lyytikäinen, Virpi (2004), Contextual and structural metadata in enterprise docu-ment management, PhD Thesis, University of Jyväskylä. Viitattu p.k.v.

[20] Metadata working group (2015), Specifications. Saatavissa http://www.metadataworkinggroup.org/specs . Varattu 10.7.2015.

[21] OASIS (2007), Metadata Model, Open Document 1.2 File Format Change Proposal, OASIS. Saatavissa https://www.oasis-open.org/committees/download.php/24327/ODF-Metadata-Proposal.pdf . Viitattu 5.8.2015.

[22] OASIS (2015), Open document format. Saatavissa http://www.opendocumentformat.org/aboutODF/ . Viitattu 5.8.2015.

[23] Photometadata.org (2015), Meta 101, Exif. Saatavissa http://www.photometadata.org/meta-resources-metadata-types-standards-exif . Viitattu 10.7.2015.

[24] Pirttimäki, Ville (2012), Yhteisöpalveluiden tägien hyödyntäminen ICT-alan muutosten ennakoinnissa ja havainnoinnissa. Saattavissa

https://jyx.jyu.fi/dspace/bitstream/handle/123456789/37703/URN:NBN:fi:jyu-201204171545.pdf?sequence=1 . Viitattu 29.7.2015.

[25] Romppanen, Mirja (2004), Digitaalisten asiakirjojen pitkäaikaissäilytys: tutkimus- ja kehittämisprojektien lähestymistapojen vertailu. Saatavissa https://tampub.uta.fi/bitstream/handle/10024/92246/gradu00393.pdf?sequence=1 . Viitattu 10.5.2015.

[26] Salminen Airi, Hajautettu hypermedia, Tietojenkäsittelytieteiden laitos, Jyväskylän yliopisto (1997). Saatavissa http://users.jyu.fi/~airi/opetus/tkod54/kasitteet.html . Viitattu 30.6.2015.

[27] Salminen Airi (2005), Metatiedot organisaatioiden sisällönhallinnassa. Saatavissa: http://users.jyu.fi/~airi/papers/Metatietoartikkeli-2005.pdf . Viitattu 15.7.2015.

[28] Salo, Immo (2015),Big Data, Big Data -määritelmä. Saatavissa http://www.bigdata.fi/big-data-maaritelma . Viitattu 26.6.2015.

[29] Standard of Japan Electronics and Information Technology Industries Association (2002), JEITA CP-3451, Excahngeable image file format for digital still cameras: Exif Version 2.2. Saatavissa http://www.exiv2.org/Exif2-2.PDF . Viitattu 10.7.2015.

[30] Takala, Petri (2015), Tiedolla Johtaminen – datan kaaoksesta aktiiviseen tuloksen johtamiseen, gofore. Saavissa https://gofore.com/it-johtaminen/tiedolla-johtaminen-datan-kaaoksesta-aktiiviseen-tuloksen-johtamiseen/ . Viitattu 27.6.2015.

[31] W3C (2015), Extensiple Markup Language (XML), http://www.w3.org/XML/ . Viitattu 7.7.2015.