

Matti Järvenpää

**TIEDON PITKÄAIKAISSÄILYTYS DIGITALISOITUVASSA
ORGANISAATIOSSA – TAPAUS JYVÄSKYLÄN
YLIOPISTON INFORMAATIOTEKNOLOGIAN
TIEDEKUNTA**

Tietojärjestelmätieteen
pro gradu -tutkielma
1.4.2003

Jyväskylän yliopisto
Tietojenkäsittelytieteen laitos
Digitaalinen media

TIIVISTELMÄ

Järvenpää, Matti Veikko

Tiedon pitkäaikaissäilytys digitalisoituvassa organisaatiossa – Tapaus Jyväskylän

yliopiston Informaatioteknologian tiedekunta / Matti Järvenpää

Jyväskylä: Jyväskylän yliopisto, 2003.

86 s.

Tutkielma

Tutkimus käsittelee digitaalisen tiedon pitkäaikaissäilytystä teknisestä ja organisatorisesta näkökulmasta kirjallisuuskatsauksen ja empiirisen tapaustutkimuksen avulla. Tutkielman keskeinen tutkimusongelma on kartoittaa tiedon pitkäaikaissäilytyksen tilaa kohdeorganisaatiossa ja verrata saatuja tuloksia kirjallisuudessa esitettyihin näkemyksiin. Tutkimuksessa tarkastellaan myös genre-pohjaisen tietoresurssien analysointimenetelmän soveltuvuutta edellä mainittuun tarkoitukseen.

Kirjallisuudessa digitaalisen pitkäaikaissäilytykseen liittyvien ongelmien ratkaisemiseksi on esitetty toimivia teknisiä ratkaisuja, mutta niiden soveltaminen on hankalaa jatkuvan teknistaloudellisen muutoksen hallitsemisessa organisaatioissa. Tapaustutkimuksen keskeinen johtopäätös puolestaan oli, että tiedon pitkäaikaissäilytys kohdeorganisaatiossa toimii ongelmitta analogisesti, vaikka merkittävä osa organisatorisesta kommunikaatiosta on luontvaiheessa digitaalista. Kirjallisuudessa esitetyt tekniset lähestymistavat eivät siis ole kohdeorganisaation kannalta vielä ajankohtaisia. Niiden soveltaminen voi kuitenkin olla tarpeellista lähitulevaisuudessa, sillä tutkimuksessa tehtyjen havaintojen perusteella digitaalisten tiedon käyttötavat monipuolistuvat kohdeorganisaatiossa nopeasti. Samanaikaisesti kohdeorganisaation informaationhallinnasta löytyi monia kirjallisuudessa kuvattuja piirteitä, jotka ovat digitaalisen tiedon pitkäaikaissäilytyksen toteutumisen kannalta ongelmallisia.

AVAINSANAT: digitaalinen tiedon pitkäaikaissäilytys, migraatio, emulointi, genreteoria, organisatorisen kommunikaation mittaus

ABSTRACT

Järvenpää, Matti Veikko

Long-term Preservation of Digital Information in Organization Undergoing Digitalization – Case Faculty of Information Technology at University of Jyväskylä
Jyväskylä: University of Jyväskylä, 2003.

86 pages

Master's Thesis

In this thesis long-term preservation of digital information is examined from technological and organizational viewpoint. Main research objective is to analyze the state of information preservation in the target organization and compare achieved results with views expressed in scientific literature. Other aim of the thesis is to study the applicability of genre-based analysis method for evaluating long-term preservation of digital information in organizations.

On basis of the literature review there are feasible solutions to technological problems of digital preservation but implementing them in organizations under constant transition proves to be difficult. In the target organization information with long-term value is preserved in a centralized archive using analogue storage media, although most of the organizational communication is created in digital form. Currently, this is an effective approach to long-term preservation as preserved digital material is mostly text-based and can be converted to analogue form for preservation purposes. Thus, discussion on technical solutions to problems of digital preservation is currently not very relevant from target organization's point of view. Use of technical methods may still be necessary in the near future as findings of the case study indicate that amount of digital materials is likely to increase quickly. At the same time information resources management of the target organization has many features mentioned in literature that are problematic for long-term preservation of digital information.

KEYWORDS: long-term preservation of digital information, migration, emulation, genre theory, measuring organizational communication

SISÄLLYSLUETTELO

1. JOHDANTO	6
1.1 Tutkimuksen tausta	6
1.2 Tutkimusongelma ja rajaukset	7
1.3 Tutkimusmenetelmät ja tutkimuksen kulku	9
2. ORGANISAATION DIGITAALINEN TIETO PITKÄAIKAISSÄILYTYKSEN NÄKÖKULMASTA	11
2.1 Tiedon pitkäaikaissäilytyksen yleiset tavoitteet ja merkitys	11
2.1.1 Digitaalinen tieto, dokumentti ja säilytys	12
2.1.2 Syyt tiedon pitkäaikaissäilytykseen organisaatiossa	14
2.1.3 Digitaalisen tiedon merkitys organisaatioissa	15
2.2 Digitaalisen tiedon säilymisen ongelmat	16
2.2.1 Eheys, tulkinta ja digitaalinen objekti	17
2.2.2 Epävarmat tallennusmediat	22
2.3 Organisatoriset ongelmat	24
2.4 Yhteenveto	25
3. LÄHESTYMISTAPOJA DIGITAALISEN TIEDON PITKÄAIKAISSÄILYTYKSEEN	27
3.1 Tekniset lähestymistavat	27
3.1.1 Esitettyjä teknisiä lähestymistapoja	27
3.1.2 Migraatio lähestymistapana	29
3.1.3 Standardit ja tallennusmuodot	31
3.1.4 Metatieto	35
3.2 Organisatoriset lähestymistavat	36
3.2.1 Informaation elinkaarimalli ja jatkumomalli	36
3.2.2 Vastuu ja pitkäaikaissäilytyksen organisointi	40
3.3 Yhteenveto	43
4. TAPAUSTUTKIMUS – KOHDEORGANISAATIO JA TUTKIMUSPROSESSI	45
4.1 Kohdeorganisaation valinta ja tutkimusongelma	45
4.2 Tapaustutkimuksen tausta	46
4.2.1 Kohdeorganisaatio	46
4.2.2 Tutkimusmenetelmä	47
4.2.3 Aineiston keräys	49
4.2.4 Tulosten kattavuus	54
5. TAPAUSTUTKIMUKSEN TULOKSET	57
5.1 Tapaustutkimuksen tulokset	57
5.1.1 Yleiskuva kohdeorganisaation tietovarannosta	57
5.1.2 Tiedon tallennusmediat ja säilytysajat	59
5.1.3 Tiedon raekoko	62
5.1.4 Organisaation pitkäaikaissäilytettävä tieto	64

6. TAPAUSTUTKIMUKSEN JOHTOPÄÄTÖKSET	67
6.1 Keskustelua tapaustutkimuksessa käytetystä menetelmästä	67
6.2 Keskustelua tapaustutkimuksen tuloksista	69
7. YHTEENVETO	73
7.1 Tulokset.....	73
7.2 Johtopäätökset.....	78
7.3 Jatkotutkimusaiheita.....	79
LÄHDELUETTELO.....	81

1. JOHDANTO

1.1 Tutkimuksen tausta

Digitaalisen tiedon pitkäaikaissäilytys muodostaa merkittävän, monitahoisen ja huonosti ymmärretyn ongelma-alueen. Tiedon määrä kasvaa nykyisin erittäin nopeasti kaikilla sovellusalueilla ja yhä kasvava osa siitä tallennetaan digitaaliseen muotoon. Samalla on selvää, että analogisen tiedon tavoin myös digitaaliseen muotoon tallennetun tiedon pitää olla saatavilla tulevaisuudessa. Asiaan liittyy suuria yhteiskunnallisia, taloudellisia sekä kulttuurillisia kysymyksiä ja siksi tiedon säilyttämismäärä on myös osa lainsäädäntöä.

Analogisten medioiden kuten paperin tai mikrofilmin kohdalla pitkäaikaissäilytys on pyritty ratkaisemaan arkistoinnilla ja siinä käytetyt toimintatavat ovat vakiintuneet satojen vuosien kuluessa. Digitaalisen tiedon kohdalla tilanne on huomattavasti ongelmallisempi. Sen käsittelyyn tarkoitettu teknologia on kehittynyt ja kehitty edelleen erittäin nopeasti. Tämä herättää kysymyksen siitä, miten tiedon saatavuus ja käyttökelpoisuus voidaan varmistaa tulevaisuudessa ohjelmistojen, laitteiden ja toimintatapojen vanhetessa jo muutamassa vuodessa. Digitaalinen tieto myös poikkeaa arkistoinnin kannalta ratkaisevasti analogisesta tiedosta, eikä siihen voida suoranaisesti soveltaa analogisen tiedon pitkäaikaissäilytykseen tarkoitettuja menetelmiä. Digitaalisen tiedon käsittely- ja muokkausmahdollisuudet ovat huomattavasti monipuolisemmat. Tämä on digitaalisen tiedon etu, mutta myös uhka ajateltaessa tiedon säilyvyyttä pitkällä tähtäimellä. Tietoa on helppo käsitellä tavoilla, jotka uhkaavat sen lyhytaikaistakin säilyvyyttä.

Digitaalisen tiedon pitkäaikaisen arkistoinnin toteuttamiseen liittyvien ongelmien ratkaiseminen on kuitenkin vasta aluillaan. Etenkin teknisestä näkökulmasta ongelmaa on tutkittu jo vuosikymmeniä, mutta yleiseen tietoisuuteen se on noussut vasta aivan viime vuosina. Tähän on vaikuttanut se, että digitaalista teknologiaa sovelletaan yhä useammilla inhimillisen toiminnan alueilla ja ongelmat ovat käymässä ilmeisiksi.

Tiedon pitkäaikaissäilytys ei ole ainoastaan siihen erikoistuneiden organisaatioiden kuten arkistojen ja kirjastojen ongelma, vaan sen kohtaavat yhä useammat yritykset ja julkiset yhteisöt.

Tavallinen ratkaisu digitaalisen tiedon pitkäaikaissäilytyksen ongelmiin on tähän asti ollut digitaalisessa muodossa olevan tiedon muuttaminen analogiseen muotoon arkistointia varten. Näin toimitaan etenkin organisaatioissa, joiden toiminta ei suoranaisesti liity arkistointiin tai suurien tietomäärien säilyttämiseen. Pitkäaikaissäilytyksen toteutuminen analogisessa muodossa kuitenkin hidastaa digitaalisen teknologian ja tallennusmuotojen käyttöön siirtymistä monella sovellusalueella, vaikka siitä muuten olisikin merkittävää hyötyä (Levy, 1998, 152). Toisaalta muunnoksessa menetetään myös kaikki digitaalisen tallennusmuodon tarjoamat edut ja monet arkistoinnin kannalta merkittävät ominaisuudet. Paperille tulostettua tietoa ei voida enää käsitellä koneellisesti, jos sitä ei digitalisoida uudelleen. On myös paljon digitaalista tietoa, jota syystä tai toisesta ei ole edes mahdollista muuttaa analogiseen muotoon. (Rothenberg, 1999a, 9 - 10)

1.2 Tutkimusongelma ja rajaukset

Digitaalisen tiedon pitkäaikaissäilytystä on käsitelty runsaasti tieteellisessä kirjallisuudessa ainakin teoreettisella tasolla (esim. Garrett & Waters, 1996; Rothenberg, 1999a; Bullock 1999). Empiiristä tutkimusta aihealueella on kuitenkin tehty vähän ja pääosin vain suoraan aihepiirin parissa toimivien organisaatioiden kuten arkistojen ja kirjastojen näkökulmasta (esim. Carrol & Hodge, 1999). Pitkäaikaissäilytykseen liittyviä malleja ja viitekehyksiä on siis olemassa, mutta tutkimustietoa niiden soveltuvuudesta toista päätehtävää varten muodostuneen organisaation tarpeisiin huomattavasti vähemmän (esim. Beagrie & al. 2001).

Tutkielman tavoitteena on osaltaan korjata aiemman tutkimuksen puutteita tapaustutkimuksen avulla käyttäen kohdeorganisaationa Jyväskylän yliopiston informaatioteknologian tiedekuntaa. Kohdeorganisaation valintaan vaikuttivat sen läheisyyden lisäksi muutkin tekijät. Kyseessä on julkishallinnollinen organisaatio, jonka tiedon pitkäaikaissäilytystä säätelevät monet ulkopuoliset sidosryhmät.

Kohdeorganisaatio on myös monessa suhteessa hyvin tyypillinen 2000-luvun alun organisaatio kuten myöhemmin tutkimuksessa käy ilmi.

Tutkielman keskeinen tutkimusongelma on selvittää kohdeorganisaation tila digitaalisen tiedon pitkäaikaissäilytyksen näkökulmasta ja peilata saatuja tuloksia aihepiiriä käsittelevässä kirjallisuudessa esitettyihin näkemyksiin. Pääongelma on tietoisesti jätetty suhteellisen avoimeksi, jotta saadaan liikkumavaraa tapausosuudessa mahdollisesti esiin nousevien yllättävienkin tekijöiden käsittelyä varten. Tutkielman osatavoitteita on kartoittaa digitaalisen tiedon pitkäaikaissäilytystä käsittelevää kirjallisuutta sekä selvittää viestinnällisiin genreihin pohjautuvan mittausmenetelmän soveltuvuutta tiedon pitkäaikaissäilytyksen kartoittamiseen organisaatioissa. Tutkimusongelman pohjalta luotu tutkimuskysymys ja sen jakautuminen osakysymyksiin esitetään tarkemmin seuraavassa kuvassa (KUVIO 1).

Tutkimuskysymys	Osakysymykset
<p>Saavatko kirjallisuudessa esitetyt digitaalisen tiedon pitkäaikaissäilytystä koskevat näkemykset tukea tarkasteltaessa kohdeorganisaation tilaa aihepiirin näkökulmasta?</p>	<p>Mikä digitaalisen tiedon pitkäaikaissäilytyksessä on kirjallisuuden valossa ongelmallista ja miten ongelmia on yritetty ratkaista?</p>
	<p>Soveltuuko kommunikatiivisiin genreihin perustuva mittausmenetelmä organisaation tietoresurssien tarkasteluun tiedon pitkäaikaissäilytyksen näkökulmasta?</p>

KUVIO 1. Tutkimuskysymys ja sen jakautuminen osakysymyksiin.

Tarkemman rajauksen kannalta on tärkeä ymmärtää, mitä tiedon pitkäaikaissäilytyksellä tarkoitetaan. Stenvall (2000, 13) määrittelee säilytyksen hallinnollisiksi, taloudellisiksi ja teknisiksi toimenpiteiksi, joiden avulla varmistetaan arkistoaineiston sekä sen sisältämän informaation säilyminen. Pitkäaikaisuuden Stenvall puolestaan määrittelee ”riittävän pitkäksi aikaväliksi, jotta säilytykseen joudutaan kiinnittämään huomiota”.

Genrellä tarkoitetaan tässä työssä organisaatiossa yhteisesti ymmärrettyä toistuvaa kommunikatiivista ilmaisua tai tilannetta, joka voidaan tunnistaa samanlaisen sisällön ja

muodon perusteella (Yates & Orlikowski, 1992). Tyypillisiä genrejä opetusorganisaatioissa ovat esimerkiksi tenttivastaus, luento ja todistus tutkinnon suorittamisesta. Tietojärjestelmätieteessä genrejä on sovellettu tietojärjestelmien suunnittelussa (Päivärinta, Halttunen ja Tyrväinen, 2001) sekä arvioitaessa organisaatioissa tapahtuvan kommunikaation määrää ja laatua (Tyrväinen & Päivärinta, 2003). Tämän työn kohdalla genrejä sovelletaan ennen kaikkea viimeksi mainitussa tarkoituksessa. Tapaustutkimuksessa käytettävää genre-pohjaista tutkimusmenetelmää esitellään tarkemmin luvussa 4.

Tutkimus keskittyy pääosin edellä mainittuihin hallinnollisiin ja teknisiin kysymyksiin. Tiedon pitkäaikaissäilytykseen liittyvät taloudelliset ja lainsäädännölliset kysymykset pyritään rajaamaan pääosin tutkielman ulkopuolelle, vaikka niitä joudutaankin esimerkin vuoksi käsittelemään. Metatietoa tutkielmassa ei aiheen laajuuden vuoksi voida käsitellä kovinkaan syvällisesti ja siten esimerkiksi metatietostandardit jäävät myös tutkimuksen ulkopuolelle. Genreihin perustuvaa analyysimenetelmää käytetään tapaustutkimuksessa keskeisenä työkaluna, sitä ei kuitenkaan esitellä enempää kuin tulosten ymmärtämiseksi on välttämätöntä, eikä sitä verrata muihin mahdollisiin menetelmiin.

1.3 Tutkimusmenetelmät ja tutkimuksen kulku

Kirjallisuuskatsaus käynnistyy arkistoinnin yleisten lähtökohtien ja tavoitteiden esittelyllä. Tästä edetään pohtimaan, miten digitaalinen tallennusmuoto vaikuttaa näiden tavoitteiden saavuttamiseen. Näin saadaan käytyä läpi digitaalisen tiedon pitkäaikaissäilytyksen keskeinen ongelmakenttä. Kirjallisuuskatsauksen toisessa vaiheessa kartoitetaan, mitä ongelmien ja erityiskysymysten ratkaisemiseksi on esitetty. Kiinnostuksen kohteena ovat tällöin tiedon pitkäaikaissäilytykseen liittyvät lähestymistavat, viitekehykset ja käytännöt.

Tapaustutkimusosuuden pohjamateriaalina toimii kohdeorganisaatioon kohdistunut aiempi tutkimusprojekti, PROMI. Kyseisessä projektissa kohdeorganisaation prosesseja ja niissä kulkevaa tietoa kartoitettiin genre-tasolla. Tapaustutkimuksessa tässä kartoituksessa nimetyistä genreistä kerätään pitkäaikaissäilytyksen kannalta

merkityksellistä lisätietoa. Tällaista tietoa ovat esimerkiksi tiedot genrejen määristä, säilytysajoista ja tallennusmedioista. Viimeisessä vaiheessa pyritään tämän, sekä muun kohdeorganisaatiosta kerätyn tiedon perusteella arvioimaan kohdeorganisaation tiedonhallinnan tilaa pitkäaikaissäilytyksen näkökulmasta sekä hakemaan vastauksia tutkimuskysymyksiin.

2. ORGANISAATION DIGITAALINEN TIETO PITKÄAIKAISSÄILYTYKSEN NÄKÖKULMASTA

Luvussa lähdetään liikkeelle esittelemällä peruskäsitteet ja pohtimalla lyhyesti tiedon pitkäaikaissäilytyksen merkitystä organisaatioissa. Seuraavaksi käsitellään digitaalisen tiedon luonnetta ja sen aikaansaamaa organisatorista muutosta. Lopuksi pohditaan, mikä ongelmia digitaalinen tallennusmuoto aiheuttaa pitkäaikaissäilytykselle.

2.1 Tiedon pitkäaikaissäilytyksen yleiset tavoitteet ja merkitys

Ihmiskunnan kehityksen kannalta tiedon ja kulttuuriomaisuuden pitkäaikainen säilyminen on ollut keskeistä. Aluksi tieto siirtyi eteenpäin suullisina kertomuksina, mutta kirjoitustaidon keksimisen jälkeen tietoa alettiin tallentaa pysyvämmiin erilaisille medioille. Ensimmäiset todisteet kirjoittamisesta löytyvätkin jo noin 6000 vuoden takaa, jolloin tietoa alettiin kirjata savitauluille (Conway, 1996).

Yhteiskunnan kehittyessä tarve tiedon säilyttämiseen on korostunut tietomäärän kasvaessa räjähdysmäisesti. Tietomäärän kiihtyvästi kasvavaa määrää kuvaa Berkeleyn yliopistossa tehty tutkimus, jossa arvioitiin maailmassa tuotetun tiedon määrää. Sen mukaan ihmiskunnan syntymisen ja vuoden 1999 välillä tuotettiin tietoa karkeasti arvioiden 12 exatavua¹. Tutkimuksessa määrän myös ennustettiin kaksinkertaistuvan vuoden 2002 puoliväliin mennessä. (Lyman & Varian, 2000) Arvio on luonnollisesti erittäin karkea, mutta kehityksen suunnasta ei liene epäilystä. Emme puhu turhaan tietoyhteiskunnasta ja informaation aikakaudesta. Erityisen suuri vaikutus tähän kehitykseen on ollut digitaalisella teknologialla, joka on monin tavoin mullistanut tapamme käsitellä, välittää ja tallentaa tietoa viimeisten vuosikymmenien aikana. Historiantutkija Morris (1998, 32) vertaakin informaatioteknologian ja digitaalisten tallennusmuotojen yleistymisen vaikutusta kirjapainon tai jopa kirjoitustaidon keksimiseen.

¹ 1 exatavu = miljardi gigatavua

2.1.1 Digitaalinen tieto, dokumentti ja säilytys

Mitä tiedon säilyttämällä sitten tarkoitetaan ja milloin se on pitkäaikaista? Hedstrom (1995) määrittelee *säilyttämisen* tarkoittavan kaikkia niitä toimenpiteitä, joiden tavoitteena on pitää jatkuvaa arvoa omaava tieto käyttökelpoisena. Perinteisen arkistoinnin näkökulmasta pitkäaikainen on ymmärretty yli 10 vuoden pituiseksi aikajaksoksi (Arkistolaitos, 2001). Tämän työn yhteydessä ”pitkäaikainen” ei kuitenkaan viittaa mihinkään tiettyyn ajanjaksoon. Enemmänkin kyseessä on aikaväli, jonka jälkeen toimintaympäristössä tapahtuvat muutokset pakottavat kiinnittämään huomiota säilytykseen. Ympäristöstä riippuen tämä aikaväli on yleensä 2-10 vuotta. (Carrol & Hodge, 1999; Stenvall, 2000, 13)

Tiedon ja digitaalisen tiedon määrittely onkin ongelmallisempaa. Suomen kielessä *tieto* on väljä yleiskäsite, joka voidaan jakaa kolmeksi alakäsitteeksi. Termillä *data* tarkoitetaan symboleja, joilla on tietty merkitys. Data sisältää *informaatiota*, jos yksikin ihminen kykenee tulkitsemaan dataa ja kasvattamaan sen avulla tietämystään. (Päivärinta, 2000, 8 - 9) *Digitaalinen* tarkoittaa nykykäsityksen mukaan teknologiaa, joka luo, varastoi ja käsittelee tietoa kahden toisistaan poikkeavan tilan eli bitin avulla (Rothenberg 1999a, 3). *Digitaalinen tieto* on siis tämän työn puitteissa biteiksi tallennettua informaatiota. Tallennusmuoto ei rajoita mitenkään digitaalisen informaation varsinaista sisältöä. Bittivirrat voivat kuvata tekstiä, ääntä, liikkuvaa kuvaa, tietokantataulukoita, hypertekstiä sekä näiden yhdistelmiä.

Käytännössä digitaalinen tieto muodostaa jonkinlaisia kokonaisuuksia ja näiden rajaamista varten joudutaan hakemaan käsitteitä analogisen tiedon maailmasta. Eri tieteenalojen edustajat käyttävät erilaisia käsitteitä viitatessaan digitaalisen tiedon muodostamiin tietokokonaisuuksiin. Tällaisia termejä ovat esimerkiksi digitaalinen dokumentti (Rothenberg, 1999a), sähköinen tallenne (Dollar, 1992) ja elektroninen julkaisu (Mackenzie ja Walle, 1996). Tutkijat ovat nähneet paljon vaivaa pohtiessaan näitä termejä ja niiden merkityksistä on keskusteltu paljon. Kunyn (1997) mielestä pohdinnasta saatava hyöty on kuitenkin tiedon pitkäaikaissäilytyksen kannalta kyseenalaista ja hän ehdottaakin käytettäväksi yläkäsitettä *digitaalinen objekti*. Tätä hän

perustelee digitaalisten tietokokonaisuuksien kasvavalla monimutkaisuudella ja multimediatiedon yleistymisellä.

Digitaalisen objektin käsite onkin toimiva puhuttaessa digitaalisen tiedon pitkäaikaissäilytyksestä pelkästään teoreettisella tasolla. Organisaatioiden tieto on kuitenkin usein asiakirjallista eli tieto on tallennettu digitaalisiin tai ei-digitaalisiin dokumentteihin. Tietojärjestelmätieteen näkökulmasta *dokumentti* voidaan määritellä loogiseksi, esitettävissä olevaksi informatiiviseksi kokonaisuudeksi tallennettua tietoa, jolla on potentiaalista merkitystä vähintään yhdelle ihmiselle jossakin organisationaalisessa kontekstissa ja joka on esitettävissä uudelleen käyttäjän niin halutessa (Päivärinta, 2000, 7). Arkistotieteessä dokumentti puolestaan ymmärretään asiakirjalliseksi tiedoksi, jota syntyy yhteisön hoitaessa tehtäväänsä tai henkilön toiminnan tuloksena (Arkistolaitos, 2001). Tällainen asiakirjallinen tieto toimii myös todisteena organisaatiossa tapahtuneesta toiminnasta ja siten sillä on myös juridista merkitystä.

Tässä tutkielmassa käytetään sekä digitaalisen dokumentin että digitaalisen objektin käsitettä viitattaessa digitaalisen tiedon muodostamaan kokonaisuuteen. Dokumentti pyritään ymmärtämään mahdollisimman laajasti ja siten molemmat edellä mainitut määritelmät sisältyvät siihen. Asiakirjojen muodostamia kokonaisuuksia ja tiedon pitkäaikaissäilytystä käsiteltäessä esiin nousee myös arkiston käsite. Sana *arkisto* voidaan määritellä kahdella tavalla. Ensinnäkin arkisto voidaan määritellä tiettyjen, toisiinsa liittyvien asiakirjojen kokoelmaksi (Arkistolaitos, 2001). Toiseksi termillä viitataan rakennukseen tai laitokseen, jossa arkistoinnin teorian ja metodologian hallitsevat ammattilaiset huolehtivat aktiivisen vaiheen ulkopuolella olevan tiedon säilymisestä (Dollar, 1992, 53). Arkisto voi olla myös digitaalinen, jos säilytettävä tieto on digitaalisessa muodossa.

2.1.2 Syyt tiedon pitkäaikaissäilytykseen organisaatioissa

Pitkäaikaissäilytyksen näkökulmasta organisaatiot voidaan jakaa kahteen ryhmään. Osa organisaatioista on keskittynyt pelkästään merkittävän tiedon ja kulttuuriomaisuuden säilyttämiseen sekä sen saatavuuden varmistamiseen. Tähän ryhmään kuuluvat esimerkiksi kirjasto- ja arkistolaitosten alaiset julkishallinnolliset organisaatiot. Toinen ryhmä muodostuu niistä organisaatioista, jotka joutuvat säilyttämään tietoa pitkäaikaisesti muun toimintansa tukemiseksi, sen jatkuvuuden varmistamiseksi ja vastataksaan lain asettamiin vaatimuksiin. Jälkimmäiseen ryhmään kuuluvat liikeyritykset ja valtionhallinnon organisaatiot. Isoissa organisaatioissa tiedon pitkäaikaissäilytyksestä saattaa vastata myös itsenäisesti toimiva organisaation osa, joka voi olla myös ulkoistettu. Molemmat organisaatiotyypit voivat järjestää pitkäaikaissäilytyksensä haluamallaan tavalla, kunhan lain asettamat velvoitteet täyttyvät.

Arkistolaki (831/94) koskee valtiota, kuntia, itsenäisiä julkisoikeudellisia laitoksia sekä ”muita yhteisöjä, toimielimiä ja henkilöitä niiden suorittaessa julkista tehtävää lain tai asetuksen taikka lain tai asetuksen nojalla annetun määräyksen perusteella siltä osin kuin niille tämän tehtävän johdosta kertyy yleisten asiakirjain julkisuudesta annetussa laissa (83/51) tarkoitettuja asiakirjoja.” Liikeyritykset ja muut organisaatiot eivät siis varsinaisesti kuulu arkistolain piiriin, vaan niiden pitkäaikaissäilytystä koskevat velvoitteet ovat peräisin muusta lainsäädännöstä. Esimerkiksi liikeyrityksen dokumenttien säilytykseen vaikuttavaa lainsäädäntöä ovat kirjanpitolaki (1336/97), verolainsäädäntö, laki viranomaisten toiminnan julkisuudesta (621/99), osakeyhtiölaki (734/78), työsopimuslaki, henkilötietolaki (523/99) sekä kilpailu- ja tuotevastuulainsäädäntö (Roos, 2001). Lainsäädännöllinen kokonaisuus on siis yllättävän monimutkainen.

Kuinka pitkiä aikoja asiakirjoja sitten säilytetään organisaatioissa? Esimerkiksi yrityksen kirjanpitoon liittyviä asiakirjoja tulee säilyttää asiakirjasta riippuen 6 - 10 vuotta ja jotkut asiakirjat säilytetään pysyvästi. Liiketapahtumia koskevan mahdollisesti laajankin kirjeenvaihdon tulee lain mukaan säilyä vähintään 6 vuotta. Laki ei määrää

tiedon säilytysmuotoa, joka voi molempien kohdalla olla digitaalinen tai ei-digitaalinen. Lisäksi on paljon pitkäaikaissäilytettävää materiaalia, jonka säilyttämiseen lainsäädäntö ei ota kantaa. Silti ne voivat olla organisaation toiminnan kannalta merkittäviä. Tällaisia asiakirjoja ovat esimerkiksi laatujärjestelmää koskevat dokumentit, joiden säilytysaikasuositus on 10 vuotta. Lisätietoa kaupallisen organisaation pitkäaikaissäilytettävästä tiedosta ja niiden säilytysaikasuosituksesta löytyy liikearkistoyhdistyksen julkaisusta, josta myös nämä esimerkit on otettu. (Liikearkistoyhdistys, 2001)

Perusvaatimukset tiedon säilyttämiselle organisaatioissa tulevat siis lainsäädännöstä. Muut syyt ovat organisaation toiminnan jatkuvuuden varmistaminen sekä tarve tukea organisaation operatiivista toimintaa ja päätöksentekoa. Monissa organisaatioissa kuten liikeyrityksissä tieto on tärkeintä omaisuutta ja sen on oltava saatavilla pitkänkin ajan kuluttua sen luomishetkestä. Seuraukset voivat olla vakavat, mikäli tietoa menetetään tai sitä ei löydetä riittävän tehokkaasti. Liikeyrityksen kohdalla tuotekehitys ja tuotanto viivästyvät, tilauksia, sopimuksia ja asiakkaita menetetään. Myös lailliset seuraamukset voivat olla vakavia. Kokonaisuutena tarpeellisen tiedon menettämisestä seuraa ylimääräisiä kuluja, jotka johtavat kilpailukyvyyn laskuun. (Roos, 2001)

2.1.3 Digitaalisen tiedon merkitys organisaatioissa

Ennen digitaalisen tiedon säilymiseen liittyvien ongelmien käsittelyä on syytä myös tarkastella digitaalisen tiedon merkitystä organisaatioissa. Digitaalisen tiedon merkitystä voi lähestyä parhaiten pohtimalla digitaalisten tietomuotojen käytön etuja. Rothenberg (1999a, 3) on kerännyt yhteen näistä keskeisempiä. Digitaalinen tieto on koneluettavaa ja siten tietoa voidaan prosessoida täysin automaattisesti kaikissa sen elinkaaren vaiheissa. Tämä mahdollistaa tiedon täydellisen kopioinnin ja monipuolisten hakujärjestelmien rakentamisen. Nykyisellä infrastruktuurilla tietoon päästään myös käsiksi hyvin pienin lisäkustannuksin maantieteellisistä rajoista riippumatta. Lisäksi digitaalisuus on tehnyt mahdolliseksi täysin uudenlaiset, dynaamiset tietomuodot kuten hypertekstin ja hypermedian, joissa on uudenlaista toiminnallisuutta. Tällainen on esimerkiksi mahdollisuus yhdistää eri muodoissa ja eri paikoissa olevaa tietoa yhdeksi

kokonaisuudeksi. Kaikkien digitaalisesta tallennusmuodoista saatavien etujen kartoittaminen on kuitenkin vaikeaa, ellei mahdotonta. Digitaalisten tallennusmuotojen käyttömahdollisuudet ovatkin osoittautuneet lähes rajattomiksi.

Organisaatioiden digitaalisen tiedon määrää koskevia tieteellisiä arvioita on yllättävän vaikea löytää, johtuen ilmeisesti arvioinnin vaikeudesta. Koulopoulos ja Frappaolo (1995, 14) viittaavat kahteen, nyt jo ikääntyneeseen konsulttitutkimukseen, joissa kartoitettiin tiedon säilytysmuotoja Yhdysvaltalaisissa organisaatioissa. Selvityksen perusteella vuonna 1987 95 % organisaatioiden tiedosta oli paperimuodossa, 4 % mikrofilmimuodossa ja 1 % digitaalisessa muodossa. Viisi vuotta myöhemmin tehdyssä vastaavassa tutkimuksessa paperin osuus oli 90 %, mikrofilmin 3 % ja digitaalisten tallennusmuotojen osuus 7 %. Kehitystä on siis tapahtunut, mutta suurin osa tiedosta on näiden lukujen valossa edelleenkin paperilla.

McNurlin ja Sprague (2001) viittaavat puolestaan toiseen konsulttitutkimukseen, jossa arvioitiin, että 90 % organisaatioiden digitaalisista dokumenteista tulostettiin vuonna 1998 paperille käsittelyä varten. Samassa tutkimuksessa osuuden kuitenkin ennustettiin laskevan 40 prosenttiin vuoteen 2005 mennessä. Kaikki edellä esitellyt luvut ovat väistämättä karkeita arvioita, eikä niiden laskentaperusteita kuvattu lähteessä kovinkaan tarkasti. Ne kuitenkin antavat viitteitä digitaalisen teknologian merkittävästä asemassa organisaatioissa. Paperille tulostettavan tiedon suuri määrä taas osoittaa, että digitaalisiin tallennusmuotoihin liittyy vielä merkittäviä käyttämättömiä mahdollisuuksia.

2.2 Digitaalisen tiedon säilymisen ongelmat

Digitaalisen median läpimurto on tuonut mukanaan paljon myönteistä kehitystä. Samanaikaisesti monet digitaalisen tiedon ominaisuudet asettavat organisaatioiden vanhat toimintatavat kyseenalaiseksi. Tiedon pitkäaikaissäilytys ei tee tässä suhteessa poikkeusta. Seuraavassa osuudessa tarkastellaan, mitä haasteita digitaalinen tallennusmuoto asettaa tiedon pitkäaikaissäilytykselle organisaatiossa, ja mistä ne ovat peräisin.

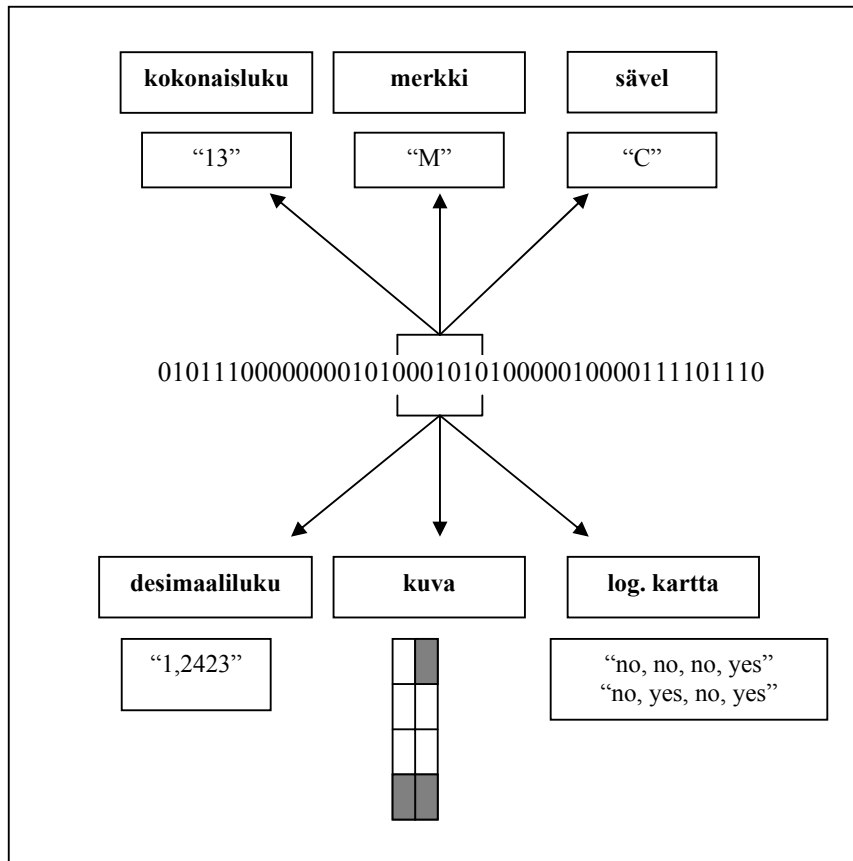
2.2.1 Eheys, tulkinta ja digitaalinen objekti

Pohjimmiltaan digitaalisen pitkäaikaissäilytyksen haasteellisuuden taustalla ovat digitaalisen objektin eheyteen liittyvät ongelmat. Tässä yhteydessä *eheys* määritellään sellaisiksi digitaalisen objektin ominaisuuksiksi, jotka tekevät siitä kokonaisen (Garret & Waters, 1996, 12). Digitaalisen objektin eheyteen liittyvät erityisongelmat selviävät parhaiten verrattaessa perinteisestä paperidokumenttia digitaaliseen dokumenttiin. Tavallinen nykyaikainen paperidokumentti sisältää samassa yksiselitteisessä kokonaisuudessa sisällön, rakenteen ja esitysmuodon. Se on myös käytettävissä ilman teknisiä apuvälineitä ainakin missä tahansa valaistussa ympäristössä (Dollar, 1992, 36). Paperidokumentin kaikki aspektit myös säilyvät, jos dokumentti ja sille tehdyt merkinnät säilyvät ehjänä (Bullock, 1999).

Digitaalisen dokumentin eri aspektit voidaan erottaa toisistaan huomattavasti helpommin, eikä dokumentin lukeminen onnistu ilman tietokonelaitteistoa ja siinä toimivaa ohjelmistoa, joka ymmärtää dokumentin rakenteen. Digitaalinen dokumentin sisältö on myös helposti erotettavissa sitä kuljettavasta artefaktista tai mediasta (Mackenzie ja Walle, 1996, 47). Se ei siis ole samalla tavalla yhtenäinen kokonaisuus kuten paperidokumentti. Esimerkiksi digitaalisen dokumentin ulkoasuun liittyviä määrityksiä voidaan muokata sisällöstä riippumatta. Mahdollisuus aspektien erottamiseen toisistaan on eräs digitaalisen tallennusmuodon tarjoamista eduista, mutta tiedon säilyvyyden kannalta se voi olla hyvin ongelmallista.

Digitaalisen objektin pitkäaikaissäilytystä vaikeuttavat myös sen tulkintaan liittyvät ongelmat. Analogisenkin tiedon tulkintaan liittyy inhimillisestä näkökulmasta katsottuna omat ongelmansa. Eri ihmiset ymmärtävät saman tekstin eri tavoin ja esimerkiksi käytetty kieli asettaa omat rajoitteensa. Näiden ongelmien lisäksi digitaalisen tiedon tulkintaan liittyy teknisiä ongelmia, jotka tekevät asiasta huomattavasti monimutkaisemman. Alhaisimmalla abstraktiotasolla digitaalinen tieto muodostuu pelkästä bittivirrasta, joka voi kuvata lähes mitä tahansa. Tämän seurauksena bittijono voidaan tulkita monin eri tavoin, mikäli sen sisältöä ja kontekstia ei tunneta. Rothenberg (1999b) havainnollistaa digitaalisen informaation

tulkitsemisongelmaa oheisella kuvalla (KUVIO 2), jossa kuvataan tietyn 8 bitin muodostaman jonon tulkitsemismahdollisuuksia. Kuten kuvasta käy ilmi, sovituisista tulkinnoista ja asiansyhteydestä riippuen jono ”00010101” voidaan ymmärtää esimerkiksi kokonaisluvuksi, kirjainmerkiksi, desimaaliluvuksi tai bittikartaksi.

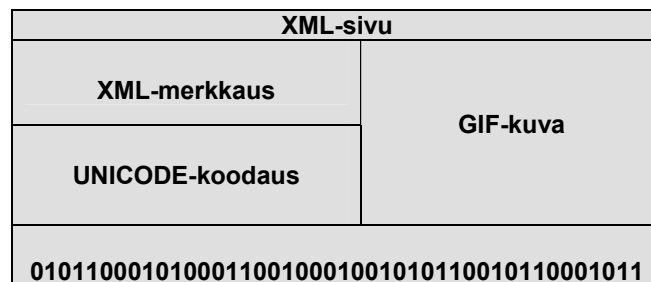


KUVIO 2. Yhden tavun mittaisen bittijonon tulkitsemismahdollisuuksia. (Rothenberg, 1999b).

Digitaalisten dokumenttien eheyteen ja tulkinnanvaraisuuteen liittyvien ongelmien vuoksi on erittäin tärkeää pohtia, millainen on ehyt dokumentti sen pitkäaikaissäilytyksen näkökulmasta. Tätä tarkoitusta varten joudutaan pohtimaan mistä osatekijöistä dokumentin sisältö muodostuu ja millaisia ominaisuuksia siihen pitkäaikaissäilytyksen kannalta liittyy sekä miten digitaalisuus vaikuttaa näihin ominaisuuksiin. Samalla haetaan vastausta siihen, mitä piirteitä dokumentista tulisi

säilyä, jotta se säilyisi täysin ehyenä. Digitaalisen dokumentin sisältö voidaan jakaa ainakin kolmeen osaan, jotka esitellään seuraavassa:

Tietosisältö: Tällä viitataan digitaalisen dokumentin varsinaiseen sisältöön. Sisältö voidaan jakaa eri abstraktiotasolle. Alimmalla abstraktiotasolla kysymyksessä on vain bittijono, eikä se perusteella voida sanoa mitään dokumentin tietosisällöstä ilman kontekstuaalista tietoa dokumentin rakenteesta. Ylimmällä tasolla on kyse dokumentin varsinaisesta informaatioisällöstä. Näiden välissä voi olla useita muita abstraktiotasoja, joita tarvitaan varsinaisen bittivirran esittämiseksi ylemmillä tasoilla. (Garrett & Waters, 1996, 12) Metsäranta (1999) havainnollistaa sisällön käsitettä kuvaamalla (KUVIO 3) esimerkinomaisesti yhden GIF-kuvan sisältävän XML-dokumentin abstraktiotasot. Tietokoneen ymmärtämä bittimuoto tehdään ihmiselle ymmärrettäväksi erilaisten koodausten ja merkkeysten avulla. Varsinaisen kääntämisen suorittaa tietokonelaitteisto siinä toimivien ohjelmistojen ohjaamana, joka esimerkiksi kuvan (KUVIO 3) tapauksessa muodostaa biteistä ihmisen ymmärtämiä kirjaimia ja merkkejä UNICODE-standardin mukaisen koodauksen avulla. XML-merkkkaus koostuu puolestaan näistä merkeistä. Dokumentin sisältö osoittautuu yllättävän laajaksi käsitteeksi ja siihen kuuluu myös tieto siitä, miten bittivirta saadaan käännettyä ihmisen ymmärtämään muotoon (Metsäranta, 1999). Ilman tällaista laitteisiin ja ohjelmistoihin piilotettua teknistä tietoa bittivirran todellista informaatioisältöä on mahdoton saada selville (Beagrie & al, 2001). Tämä tekee digitaalisen tiedon sisällöllisen tulkinnan täysin ohjelmistoriippuvaiseksi ja siten helposti haavoittuvaksi.



KUVIO 3. XML-sivun abstraktiotasot (Metsäranta, 1999).

Ulkoasu: Ulkoasuun kuuluvat tiedot esimerkiksi kirjasinlajeista, väreistä, taitosta, marginaaleista ja alaviitteistä. Monissa dokumenttityypeissä ulkoasu määritellään toisessa dokumentissa, erillään varsinaisesta sisällöstä. Ulkoasu saattaa vaikuttaa merkittävästi dokumentin ymmärrettävyyteen ja siksi sen säilyminen on tärkeää. (Bullock, 1999)

Toiminnallisuus: Toisin kuin paperidokumentti, digitaalinen dokumentti voi sisältää myös toiminnallisuutta. Digitaalinen dokumentti saattaa sisältää esimerkiksi tietokannasta haettua dynaamista sisältöä, dokumenttiin upotettuja hakujärjestelmiä, navigointipainikkeita, linkkejä ja lomakkeita. Dokumentin tarkoituksenmukainen toiminta saattaa kokonaan perustua näihin toimintoihin. (Bullock, 1999) Digitaalisen dokumenttien tai muiden objektien kohdalla toiminnallisuuden merkitys on kasvanut niin paljon, että se on johtanut jopa uuden käsitteen syntymiseen. Shepherd ja Watters (1998) esittelevätkin käsitteen *kybergenre*. Analoginen genre, kuten sanomalehti, voidaan johdannossa esitellyn määritelmän mukaisesti tunnistaa sisällön ja muodon perusteella. Kybergenren kuten hypertekstiympäristössä toimivan verkkolehden kuvaavat tekijät ovat puolestaan sisältö, muoto ja toiminnallisuus.

Vaikka digitaalisen dokumentin sisältö, ulkoasu/muoto ja toiminnallisuus säilyisivätkin, on dokumentti pitkäaikaissäilytyksen kannalta hyödytön, ja vailla todistusvoimaa ilman oheisessa luettelussa esiteltävien ominaisuuksien säilymistä:

Autenttisuus: Käyttäjän on voitava varmistua dokumentin aitoudesta ja muuttumattomuudesta, jotta dokumentti olisi todistusvoimainen. Samanaikaisesti digitaalisen dokumentin muokkaaminen, täydellisten kopioiden ja toisistaan poikkeavien versioiden luominen on erittäin helppoa. Verkkoympäristössä olevan hypertekstidokumentin ongelmaksi voi muodostua myös dokumentin rajojen selvittäminen. Arkistotieteessä onkin herännyt kysymys onko digitaalisen materiaalin autenttisuuden varmistaminen edes mahdollista nykyisiä arkistointimenetelmiä käyttäen, vai joudutaanko tarkoitusta varten luomaan aivan uudenlaisia menetelmiä ja toimintatapoja (Hirtle, 2000).

Viitattavuus: Tietty digitaalinen dokumentti on voitava erottaa selkeästi muista dokumenteista, jotta sen yhtenäisyydestä ja eheydestä voidaan varmistua. Dokumentilla on tällöin oltava yksiselitteinen nimi ja sen eri versiot on pystyttävä erottamaan toisistaan. (Garrett & Waters, 1996, 15 - 16) Linkkien käyttöön perustuvassa hypertekstiympäristössä ongelmaksi voi muodostua myös tietyn dokumentin rajojen selvittäminen.

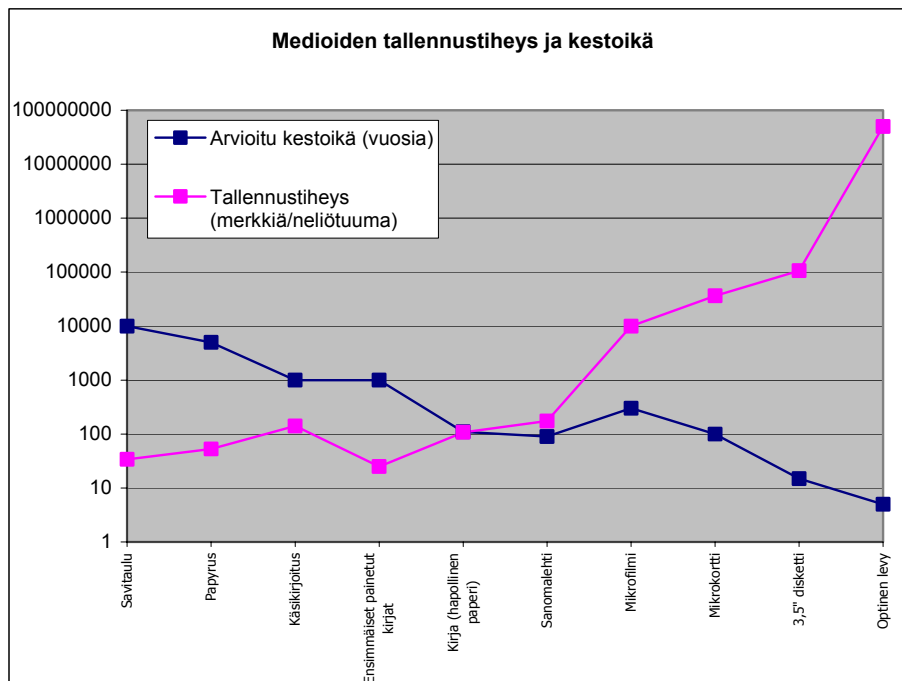
Proveniensi: Alkuperä eli *proveniensi* on yksi arkistotieteen keskeisimpiä periaatteita ja se liittyy läheisesti myös autenttisuuteen. Sillä viitataan dokumentin taustaan, omistusketjuun ja prosessiin, jonka tuloksena tietty dokumentti on syntynyt (Dollar, 1992, 48). Käsitteen taustalla on ajatus siitä, että tiedot dokumentin taustasta auttavat sen eheyden ja autenttisuuden varmistamisessa. Proveniensiin liittyy läheisesti myös kontekstin käsite. *Kontekstilla* viitataan tässä yhteydessä laajempaan tekniseen ja sosiaaliseen ympäristöön, jonka kanssa dokumentti on vuorovaikutuksessa. (Garrett & Waters, 1996) Digitaalisen dokumentin kohdalla kontekstia hämärtää se, ettei dokumenttia ole olemassa fyysisenä kokonaisuutena. Beagrie & al. (2001) käyttävät tästä esimerkkinä perinteistä paperidokumenttien arkistointijärjestelmää. Siinä tiedot dokumentin kontekstista ja provenienssista sisältyvät osaltaan dokumenttien fyysiseen ryhmittelyyn arkistointijärjestelmässä. Tiedot dokumentit on sijoitettu tiettyyn hyllyyn ja niihin läheisesti liittyvät dokumentit viereiseen hyllyyn. Digitaalisten dokumenttien maailmassa tämä ei enää toimi, koska esimerkiksi verkkoympäristössä tieto on vapaata paikan asettamista rajoitteista. Tällöin proveniensiin ja kontekstiin liittyvät tiedot on pitkäaikaissäilytystä varten tietoisesti tallennettava ja dokumentoitava. Tämä on ongelmallista, koska useimmat tietojärjestelmät eivät perinteisesti ole tukeneet tällaisen tiedon tallentamista (Dollar, 1992).

Pitkäaikaissäilytyksen näkökulmasta digitaalinen dokumentti poikkeaa ratkaisevasti perinteisestä dokumentista. Sen sisältö, ulkoasu ja rakenne ovat erotettavissa toisistaan sekä niitä kuljettavasta tallennusmediasta. Lisäksi monet ei-digitaalisen tiedon pitkäaikaissäilytykseen liittyvät konseptit, kuten proveniensi ja konteksti eivät sellaisenaan ole toimivia säilytyksen kohteen ollessa digitaalisessa muodossa. Niitä ei myöskään ole otettu riittävästi huomioon tietojärjestelmiä rakennettaessa. Onkin

epätodennäköistä, että digitaalisia tietokokonaisuuksia voidaan perinteisen arkistoitavan materiaalin tapaan säilyttää alkuperäisenä alkuperäisellä materiaalilla. Mackenzie ja Walle (1996, 46) arvelevatkin säilytyksen painopisteen siirtyvän yhä enemmän varsinaiseen informaatioisisältöön. Syynä tähän ovat pitkäaikaissäilytykseen käytettävien resurssien rajallisuus sekä digitaalisen tiedon riippuvuus nopeasti vanhentuvasta ohjelmisto- ja laitteistoteknologiasta.

2.2.2 Epävarmat tallennusmediat

Nykyaikaisille tallennusmedioille² on tyypillistä teknisen kehityksen suuri nopeus ja epätasapainoisuus. Tallennustiheys on kasvanut räjähdysmäisesti viimeisen kahden vuosikymmenen aikana. Samalla tallennusmedioiden käyttöikä ja luotettavuus ovat laskeneet jyrkästi. Conway (1996) on tiivistänyt tämän nykyaikaisten medioiden ongelman oheiseen kuvaan (KUVIO 4).



KUVIO 4. Tallennusmedioiden tallennustiheys verrattuna kestoikään (Conway, 1996).

² Tallennusmedia tai väline viittaa tässä työssä sekä tiedon sisältävään välineeseen että sitä käsittelevään laitteeseen.

Kuvassa verrataan eri-ikäisten tallennusmedioiden arvioitua kestoikää ja tallennustiheyttä. Kuvaajista voi nähdä selkeästi, miten kestoikä on tasaisesti laskenut tallennustiheyden kasvaessa. Erityisesti digitaaliseen teknologiaan pohjautuvien medioiden kohdalla tallennustiheyden ja kestoiän välinen ero on suuri. Digitaalisten tallennusmedioiden todellisesta kestoiästä onkin esitetty hyvin vaihtelevia arvioita sekä tutkijoiden että laitevalmistajien taholta. Selvää kuitenkin on, että tämän hetkisten tallennusmedioiden fyysinen kestoikä on hyvin rajallinen verrattuna ei-digitaalisiin medioihin (Rothenberg, 1999a). Magneettisten ja optisten tallennusvälineiden kestoikä on välineestä riippuen 5 - 20 vuotta, arkistointilaatuisen paperin ja mikrofilmin säilyessä oikeissa olosuhteissa jopa satoja vuosia (Beagrie & al., 2001; Hedstrom, 1995).

Arkistointitieteessä huomio on perinteisesti kiinnittynyt tietoa kuljettavan materiaalin kuten paperin säilyttämiseen, jolloin siinä oleva tietokin säilyy alkuperäisessä muodossa. Digitaalisen tiedon kohdalla tämä ei kuitenkaan toimi ja kirjallisuudessa vallitseekin laaja yksimielisyys siitä, ettei digitaalinen tieto säily alkuperäisellä materiaalilla, vaan se on ajoittain siirrettävä uusille tallennusmedioille (Hedstrom, 1995; Garrett & Waters, 1996). Pikaisesti ajateltuna tämä tuntuukin toimivalta vaihtoehdolta, sillä digitaalinen tietoa voidaan käsitellä täysin koneellisesti ja sen täydellinen kopiointi on helppoa.

Todellinen ongelma muodostuu teknologisen kehityksen nopeudesta ja ohjelmistojen vanhentumisesta. Garrett ja Waters (1996, 5) arvioivat raportissaan digitaalisen tiedon käsittelyyn tarvittavien laitteistojen, ohjelmistojen ja prosessien korvautuvan uusilla 2 - 5 vuoden jaksoissa. Digitaalisten julkaisujen pitkäaikaissäilytykseen erikoistunut raportti suhtautuu asiaan hieman valoisammin ja arvioi digitaalisen julkaisun vanhenevan teknisesti 10 - 25 vuodessa. (Mackenzie & Walle, 1995, 7). Uudemmat artikkelit puhuvat vain 18 kuukaudesta (Kuny, 1997) tai muutamasta vuodesta (Rothenberg, 1999a). Joka tapauksessa on selvää, että tällainen kehitysvauhti tekee toimintaympäristön epävakaaksi ja muodostaa uhan ohjelmisto- ja laitteistoteknologian jatkuvuudelle pitkällä tähtäimellä. Digitaalisen materiaalin käyttökelpoisena pitäminen eri teknologiasukupolvien yli vaatii perinteistä pitkäaikaissäilytystä enemmän aktiivista

toimintaa tiedon elinkaaren alkuvaiheista lähtien (Beargie & al., 2001). Tämä muodostaa tiedon säilyttämisestä vastaavalle taholle median rappeutumista suuremman haasteen.

2.3 Organisatoriset ongelmat

Digitaalisen tiedon pitkäaikaissäilytys on tekninen ongelma, mutta suurelta osin kyse on myös organisatorisesta ongelmasta. Organisatorisen ongelman taustalla olevia tekijöitä ovat teknistaloudellisen kehityksen nopeus ja sen aikaansaama tietomäärän kasvu. Samanaikaisesti nykyaikainen informaatioteknologia on pakottanut organisaatiot muuttamaan toimintatapojaan.

Dollar (1992, 40 - 42) pohtii informaatioteknologian vaikutusta organisaatioihin ja niiden toimintaan. Hänen mukaansa informaatioteknologia on vähentänyt riippuvuutta ajasta ja paikasta sekä lisännyt työntekijöiden keskinäistä vuorovaikutusta. Yhdessä nämä ovat johtamassa organisaatioiden hierarkkisen rakenteen hajoamiseen. Organisaation hierarkiat on alun perin tarkoitettu tehostamaan työskentelyä rajoittamalla viestintää organisaation osien välillä. Paperilla tapahtuvan viestinnän aikakaudella tämä onnistuikin, mutta informaatioteknologia on rappeuttanut koko ajatusmallia. Viestintä organisaation rakenteiden yli on uusien välineiden myötä helpottunut ja samanaikaisesti tiedon byrokraattinen kontrollointi on vaikeutunut.

Tehostunut viestintä ei kuitenkaan ole ainoa syy organisaation rakenteen hajautumiseen. Hierarkkiset organisaatiot ovat yksinkertaisesti liian hitaita muuttumaan toimiakseen jatkuvasti kehittyvässä ympäristössä. Viimeisen vuosikymmenen aikana yritykset ovatkin nostaneet mukautumiskyvyn kriittiseksi selviytymistekijäksi epävakaa toimintaympäristössä. Hierarkkiset organisaatiot ovat korvautuneet rakenteeltaan hajautetuilla verkosto-organisaatioilla tai hybridiorganisaatioilla, joissa on sekä hierarkkisen että verkottuneen organisaation piirteitä. (Simpson, 1998, 89) Tiedon pitkäaikaissäilytyksen kannalta tämä on ongelmallista, koska vallitsevat arkistointimenetelmät soveltuvat huonosti muihin kuin hierarkkisiin organisaatioihin (Dollar, 1992, 42). Suurissa organisaatioissa arkistointi on perinteisesti ollut

keskittynyttä toimintaa, tietoresurssit siirretään keskitetyn arkiston vastuulle niiden elinkaaren aktiivisen vaiheen päätyttyä. Dollar (1992, 53 - 55) ennustikin tämän olevan muuttumassa ja pitkäaikaissäilytyksen siirtyvän enemmän osaksi organisaation operatiivista toimintaa. Dollarin mukaan tämä vaatii myös monien perinteisten toimintatapojen ja siten myös organisaation toimintaa tukevien tietojärjestelmien muuttumista. Dollarin ennusteiden toteutumista tarkastellaan lähemmin luvussa 3.

Organisaatioiden muuttaessa rakenteitaan tekninen kehitys kulkee eteenpäin ja hallittavan tiedon määrä kasvaa. Eräs merkittävä syy tietomäärän kasvuun ja tiedon hallinnan vaikeuteen organisaatioissa on henkilökohtaisten tietokoneiden 1980-luvulla alkanut yleistyminen, kehittyminen ja verkottuminen (Morris, 1998). Yksittäiset ihmiset ja työntekijät voivat helposti tuottaa suuria määriä helposti jaettavaa tietoa henkilökohtaisten tietokoneiden avulla. Toinen tiedon pitkäaikaissäilytystä vaikeuttava tekijä on organisaatioiden ohjelmisto- ja laitteistoympäristöjen heterogeenisyys. Nykyisten organisaatioiden tiedonhallinnasta puhuttaessa käytetäänkin informaatioosaariston käsitettä. Organisaatioihin on nopean kehityksen myötä syntynyt paljon eri-ikäisiä ja keskenään yhteensopimattomia tietojärjestelmiä (Koulopoulos & Frappaolo, 1995, 16). Tieto on hajautunut pieniin järjestelmiin ja siirtynyt lähemmäksi käyttäjää. Samanaikaisesti tietoon liittyvät vastuukysymykset ovat tulleet monimutkaisemmiksi ja tiedonhallinnan tasossa on suuria eroja eri organisaatioiden välillä (Simpson, 1998). Lisäksi Simpson (1998, 92) toteaa, että organisaatioiden informaatioteknologiaan liittyvässä toiminnassa painotetaan usein teknologian osuutta varsinaisen informaation jäädessä heikompaan asemaan. Hänen mukaansa organisaatiot jättävät merkittäviä mahdollisuuksia käyttämättä unohtaessaan kiinnittää riittävästi huomiota tiedon strategiseen hallintaan.

2.4 Yhteenveto

Tiedon pitkäaikainen säilyminen on yhteiskunnan kehittymisen perusedellytys ja siksi sen toteutuminen on haluttu turvata lainsäädännön avulla. Ilman sitä organisaatiot eivät myöskään pysty täyttämään juridisia velvoitteitaan. Tutkimuksessa käytettiin esimerkkinä liikeyrityksen pitkäaikaissäilytykseen vaikuttavaa lainsäädäntöä ja tällaisen

organisaation pitkäaikaissäilytystä koskeva lainsäädäntö osoittautui yllättävän laajaksi ja monimutkaiseksi kokonaisuudeksi. Organisaatiot eivät säilytä tietoa pelkästään vastatakseen lainsäädännön vaatimuksiin. Nykyaikaiset organisaatiot ovat hyvin tietointensiivisiä ja elinkaaren eri vaiheissa olevaa tietoa tarvitaan tukemaan päätöksentekoa ja operatiivista toimintaa.

Digitaalisen tiedon pitkäaikaissäilytyksen ongelmallisuuden taustalla olevat päätekijät ovat nopea teknologinen kehitys ja digitaalisen tiedon hajanainen luonne verrattuna analogiseen tietoon. Digitaalisen tiedon hajanainen luonne asettaa monet arkistoinnin käsitteet ja toimintatavat kyseenalaisiksi. Samanaikaisesti teknistaloudellinen kehitys on johtanut tietomäärän nopeaan kasvuun ja pakottanut organisaatiot jatkuvaan rakennemuutokseen. Tieto on siirtynyt keskitetyistä järjestelmistä huomattavasti pienempiin kokonaisuuksiin kuten henkilökohtaisiin tietokoneisiin. Tämä on tehnyt tiedon hallinnan vaikeaksi ja aiheuttaa ongelmia sen pitkäaikaissäilytykselle. Kaikki nämä tekijät kietoutuvat yhteen muodostaen monitahoisen ja vielä huonosti ymmärretyn ongelmakentän. Käynnissä onkin eräänlainen siirtymävaihe analogisen ja digitaalisen maailman välillä. Tulevaisuus näyttää miten tiedon pitkäaikaissäilytyksessä onnistutaan. Tulevilla historian kirjoittajilla saattaa olla kaksi vaihtoehtoista ongelmaa heidän tutkiessaan 2000-luvun alun tietoyhteiskuntia; joko hallitsemattoman suuri ajanjaksoa koskeva tietomassa on säilynyt tai suuri määrä tarpeellista tietoa on menetetty pysyvästi (Morris, 1998).

3. LÄHESTYMISTAPOJA DIGITAALISEN TIEDON PITKÄAIKAISSÄILYTYKSEEN

3.1 Tekniset lähestymistavat

Pitkäaikaissäilytyksen ongelmien ratkaisemiseksi on esitetty useita teknisiä ja organisatorisia lähestymistapoja, joilla tiedon säilyminen voitaisiin varmistaa. Seuraavassa käydään läpi näistä keskeisempiä ja pohditaan niiden toimivuutta.

3.1.1 Esitetyt teknisiä lähestymistapoja

Keskeiset kirjallisuudessa esitetyt lähestymistavat voidaan jaotella ryhmiin, joita ovat tulostaminen erilaisille materiaaleille, tietokoneuseot, siirtomenetelmä eli migraatio sekä emulointi (Bullock, 1999; Rothenberg 1999a, 9 - 15). Tulostamisella tarkoitetaan digitaalisessa muodossa olevan tiedon tulostamista perinteiselle arkistokelpoiselle materiaalille kuten paperille tai mikrofilmille. Tätä ratkaisua on käytetty paljon ja se onkin toimiva ratkaisu organisaatioissa, joissa arkistoitavaa materiaalia on vähän ja se on pääosin tekstimuotoista. Kokonaisuutena menetelmä ei kuitenkaan ole varsinainen ratkaisu. Digitaalisen materiaalin käytöstä saatavat edut ja lisätoiminnallisuus menetetään koneluettavuuden katoamisen seurauksena (Rothenberg, 1999a, 10). Ratkaisu ei myöskään ota kantaa multimediatiedon, kuten kuva- ja äänitiedostojen säilyttämiseen (Bullock, 1999).

Toinen yleisesti tyrmätty ratkaisu on luottaminen teknologian säilymiseen. Tässä lähestymistavassa digitaalisen tiedon lukemiseen tarvittava teknologia säilytettäisiin erityisissä tietokoneuseoissa. Lähestymistavan ongelmia ovat sen kalleus ja epäkäytännöllisyys. Kaikkien mahdollisten laiteyhdistelmien säilyttäminen yhdessä paikassa ei vaikuta realistiselta ja tällöinkin ollaan riippuvaisia museoiden maantieteellisestä sijainnista. Tietokoneuseot eivät myöskään tarjoa minkäänlaista ratkaisua pitkäaikaissäilytyksen organisatorisiin ongelmiin. (Rothenberg, 1999a, 12)

Vakavammin otettavina ratkaisuinä on pidetty *siirtomenetelmää* eli *migraatiota* sekä *emulointia*. Siirtomenetelmä pitää sisällään useita eri menetelmiä ja niiden yhdistelmiä. Perusideana kuitenkin on digitaalisen aineiston siirtäminen vanhentuvasta ohjelmisto- ja laitteistoympäristöistä uusiin. Migraatiota pidetään yleisesti hyväksyttynä ratkaisuna ja digitaalisen tiedon pitkäaikaissäilytystä käsittelevät keskeiset selvitykset suosittelevat sen käyttöä (Garrett & Waters, 1995; Mackenzie ja Walle, 1996). Tässä yhteydessä on huomattava, että eri lähteet ymmärtävät migraation eri tavoin, esimerkiksi Garrett & Waters (1995, 27) katsovat myös tulostamisen olevan eräs migraatiomenetelmä.

Siirtomenetelmää kohtaan on esitetty myös kritiikkiä. Migraation on sanottu vaativan runsaasti henkilö-, aika- ja taloudellisia resursseja johtuen toimenpiteen monimutkaisuudesta. Aiheellisesti voidaan kuitenkin todeta saman ongelman koskevan myös muita lähestymistapoja (Granger, 2000). Lisäksi virheiden todennäköisyys on suuri ja ne kertautuvat helposti prosessia toistettaessa (Rothenberg, 1999a). Tästä huolimatta siirtomenetelmää pidetään ainoana kattavana ratkaisuna digitaalisen tiedon pitkäaikaissäilytyksen ongelmiin ja se on myös käytössä monissa organisaatioissa (Bearman, 1999; Bullock, 1999; Hedstrom & Montgomery, 1998; Hodge, 2000). Tämän vuoksi siirtomenetelmä on myös tässä työssä valittu pääasialliseksi tekniseksi lähestymistavaksi ja sitä tarkastellaan yksityiskohtaisemmin hiukan myöhemmin.

Ehkä eniten keskustelua herättänyt lähestymistapa lienee emulointi. Siinä ideana on ”matkia” vanhentuneiden laitteistojen ja ohjelmistojen toimintaa uudemmassa ympäristössä erityisen emulaattoriohjelmiston avulla. Tällöin digitaalisen materiaalin pitkäaikaissäilyvyys voitaisiin varmistaa ja digitaalista dokumenttia tarkastella lähes alkuperäisen kaltaisena alkuperäisessä ohjelmisto- ja laitteistoympäristössä. (Rothenberg, 1999a, 17; Bullock 1999) Teknisesti emulointi onkin havaittu toimivaksi ratkaisuksi (Hedstorm & Lampe, 2001; Rothenberg, 2000). Lähestymistapaan sisältyy kuitenkin paljon heikkouksia. Emulointiohjelmistot ovat teknisesti monimutkaisia. Niiden rakentaminen vaatii runsaasti resursseja, eikä emulointiohjelmistojen laatiminen kaikista yleisesti käytössä olevista laitteisto- ja ohjelmistoympäristöistä sekä näiden yhdistelmistä ole mahdollista (Granger, 2000). Lisäksi emulointi on tarpeettoman laaja

ja huonosti mukautuva ratkaisu, sillä dokumentin säilyminen täysin muuttumattomana on harvoin ehdoton vaatimus. Myös tekijänoikeuskysymykset muodostavat merkittävän esteen emuloinnin laajamittaiselle käytölle (Bearman, 1999). Tästä huolimatta emuloinnilla on varmasti merkitystä tilanteissa, joissa säilytettävän digitaalisen materiaalin tulkitseminen on poikkeuksellisen riippuvaista tietyn ohjelmisto- tai laitteistoympäristön säilymisestä (Bullock, 1999). Pääasiallisesti lähestymistavaksi suurten tietomassojen säilymisen varmistamiseen se ei kuitenkaan sovellu.

3.1.2 Migraatio lähestymistapana

Garrett & Waters (1996) määrittelevät migraation tarkoittavan digitaalisen aineiston jaksottaista siirtämistä tietystä ohjelmisto- ja laitteistoympäristöstä tai sukupolvesta toiseen. Migraation tarkoituksena on digitaalisen objektin eheyden ja käyttökelpoisuuden säilyttäminen kehittyvässä teknologisessä ympäristössä.

Yksinkertaisimmillaan migraatiossa on kyse säilytettävän materiaalin siirtämisestä (refresh) muuttumattomana vanhalta tallennusmedialta uudelle. Toimenpide on tarpeellinen, jos nykyinen tallennusmedia uhkaa tulla teknisesti vanhanaikaiseksi tai muuttuu ikääntymisen seurauksena epäluotettavaksi (Bullock, 1999). Tällainen siirtäminen soveltuukin parhaiten tilanteisiin, joissa tieto on jo valmiiksi mahdollisimman ohjelmistoriippumattomassa muodossa kuten ASCII-tekstinä, eikä konvertoinnista ole merkittävää lisähyötyä (Garrett & Waters, 1996, 27). Menetelmän etuja ovat sen yksinkertainen toteutus, kustannustehokkuus ja hyvä saatavuus. Näin yksinkertainen migraatio vastaa kuitenkin ainoastaan fyysisen median vanhenemisen aiheuttamiin ongelmiin, mutta ei tarjoa ratkaisua ongelmiin, jotka aiheutuvat ohjelmistojen- ja tietokonelaitteistojen nopeasta vanhenemisestä.

Toisessa migraatiotyypissä vaihdetaan median lisäksi/sijasta tiedon formaattia eli tallennusmuotoa. Tämä tulee kysymykseen tietyn ohjelmiston tai sen laitteistoympäristön tullessa vanhanaikaiseksi tai harvinaiseksi. Organisaatiossa saattaa esimerkiksi olla digitaalisia dokumentteja, joita pystytään käsittelemään ainoastaan tietyllä sovelluksella, joka on katoamassa tai kadonnut markkinoilta. Tällöin kyseiset

dokumentit pyritään mahdollisuuksien mukaan muuttamaan toiseen, yleisempään tallennusmuotoon. Kirjallisuudessa suositellaan mahdollisimman pysyvien, standardoitujen ja laitteistoriippumattomien tallennusmuotojen käyttöä (Garrett & Waters, 1996, 28 - 29). Tällaisessa migraatiossa tiedon eri aspektien säilyminen riippuu käytetyistä formaateista ja migraation toteutustavasta.

Kaikkein monimutkaisimmassa migraatiossa digitaalinen objekti siirretään teknologiaympäristöstä toiseen, esimerkiksi VAX-ympäristöstä UNIX-ympäristöön (Bullock, 1999). Tämän tapainen migraatio on tarpeellista, mikäli digitaalisen objektin toiminnallisuuden säilyminen on sen sisällön säilymisen edellytys. Yleisesti on huomattava, että erilaisia migraatiostrategioita sovelletaan usein samanaikaisesti.

Resurssivaatimusten lisäksi migraatioon liittyy monia ongelmia. Migraation ajoitus ja käytettävien tallennusmuotojen valinta ovat vaikeita ratkaisuja teknologisen kehityksen ja rajallisten resurssien hallitsemassa toimintaympäristössä. Ohjelmistot, laitteistot, standardit ja toimintatavat muuttuvat jatkuvasti ja usein myös yllättävästi (Rothenberg, 1999a, 14). Migraatio joudutaan myös toistamaan säännöllisin väliajoin, jotta säilytettävä tieto säilyisi käyttökelpoisena. Joissakin hyvin suurien tietomäärien säilyttämisestä vastaavissa organisaatioissa migraatio on lähes jatkuvaa toimintaa, uusi migraatiokierros joudutaan aloittamaan edellisen ollessa vielä kesken (Carrol & Hodge, 1999). Tämä luonnollisesti vaatii organisaatiolta runsaasti resursseja ja kasvattaa virheiden mahdollisuutta. Mahdolliset virheet myös kertautuvat helposti jokaisella migraatiokerralla. Migraation ajoittamista vaikeuttavat myös informaatioteknologialle tyypilliset äkillisesti ilmaantuvat uudet ohjelmistoparadigmat (Rothenberg 1999a, 14). Tällaisia paradigmoja ovat esimerkiksi oliotietokannat, www-pohjaiset hypermediadokumentit sekä sähköpostiliitteet. Kaikki nämä paradigmat ovat yleistyneet suhteellisen lyhyessä ajassa, aiheuttaneet merkittäviä muutoksia toimintatavoissa, eikä niiden yleistymistä ole osattu ennakoida. On myös todennäköistä, että vastaavia muutoksia tapahtuu runsaasti tulevaisuudessa. Tuleviin migraatiokierroksiin liittyviä teknisiä ongelmia on siis hyvin vaikea ennakoida ja tiettyihin tallennusmuotoihin migraatiota ei voida soveltaa lainkaan. (Hakala, 2001).

Myös MacKenzie ja Walle (1996, 51 – 52) pohtivat migraation suunnitteluun liittyviä ongelmia. He suosittelevatkin migraation jakamista suunnitteluvaiheessa lyhyt- ja pitkäaikaiseksi migraatioksi. Lyhytaikainen tarkoittaa tässä yhteydessä noin 5 vuoden pituista ajanjaksoa, jolloin huomio kiinnittyy nykyisten tallennusmedioiden ja -muotojen pysyvyyttä parantaviin migraatiotoimenpiteisiin. Pitkäaikaisen 10 – 25 aikajänteellä tapahtuvan migraation suunnittelun myös MacKenzie ja Walle toteavat miltei mahdottomaksi tehtäväksi. Näin pitkällä aikavälillä migraatio saattaa korvautua jollain toisella lähestymistavalla ja teknologisen kehityksen nopeudessa voi tapahtua vaihtelua. Myös MacKenzie ja Walle viittaavat paradigmanmuutoksiin puhuen ”tappajateknologioista” (killer technologies). Tällä he tarkoittavat teknologioita jotka aiheuttavat merkittäviä muutoksia hyvin lyhyellä aikavälillä.

Kokonaisuutena migraatio on monimutkainen ja monin tavoin myös puutteellinen lähestymistapa tiedon pitkäaikaissäilytykseen. Kirjallisuudessa se on kuitenkin kattavasti todettu ainoaksi lähestymistavaksi, joka toimii myös laajemmassa mittakaavassa. Migraatio ei oikeastaan ole yksittäinen lähestymistapa, vaan enemmänkin kokoelma tiedon siirtämiseen perustuvia lähestymistapoja. Tämä tekee migraatiosta lähestymistavan, joka voidaan muokata sopivaksi erilaisiin tarpeisiin. Hakala (2001) kiteyttää migraation toteamalla sen olevan tilanteesta riippuen välillä yksinkertaista ja välillä äärimmäisen haastavaa. Migraation soveltaminen vaatii kuitenkin runsaasti asiantuntemusta ja resursseja, sillä virheet saattavat olla peruuttamattomia ja ne havaitaan vasta tulevaisuudessa.

3.1.3 Standardit ja tallennusmuodot

Migraatiossa, emuloinnissa ja muussa pitkäaikaissäilytyksessä standardit ovat tärkeässä asemassa. Standardit helpottavat digitaalisen tiedon hallintaa ja pitkäaikaissäilytystä vähentämällä tiedon tallennusmuotoja ja vaadittavien sovellusten määrää. Hyvin laaditut avoimet standardit parantavat myös ohjelmistojen ja tietojärjestelmien keskinäistä yhteensopivuutta. On myös todennäköistä, että standardimuotoon tallennettu tieto säilyy pidempään käyttökelpoisena. Standardit ovat pitkäaikaissäilytyksen kannalta

merkityksellisiä ja ”pysyvien” standardien luomista on esitetty myös ratkaisuksi tiedon pitkäaikaissäilytykseen (Bearman, 1999).

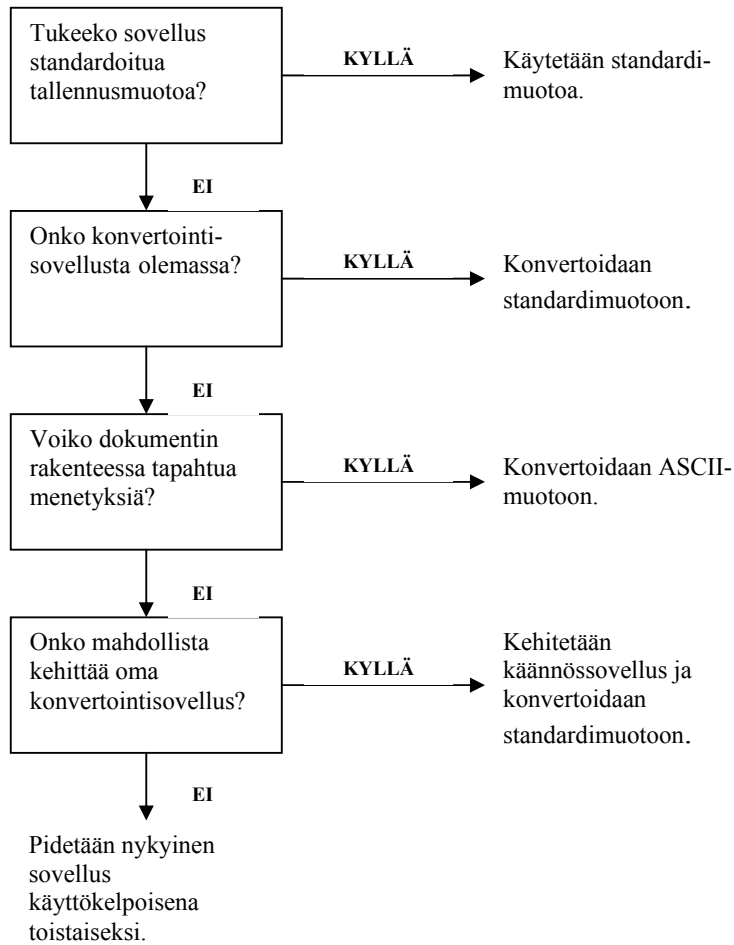
Itsenäiseksi lähestymistavaksi standardit eivät kuitenkaan sovellu johtuen niiden lyhytikäisyydestä ja epävakaudesta. Nopea tekninen kehitys tekee hyvätkin standardit nopeasti vanhanaikaiseksi ja pitkällä tähtäimellä on vaikea tietää, mikä lukuisista kilpailevista standardeista on lopulta pitkäikäisin. Kokonaisuutena standardit kuitenkin tukevat merkittävästi tiedon pitkäaikaissäilytystä ja jotkut standardit ovat myös vakiinnuttaneet asemansa. (Rothenberg, 1999a) Standardien käyttömahdollisuudet pitkäaikaissäilytyksessä ovatkin monipuoliset, tallennusmuotojen valinnan lisäksi standardeja sovelletaan myös arkistoitavaan tietoon liittyvän metadatan määrittelyssä ja tallentamisessa. Lisäksi tiedon pitkäaikaissäilytykseen liittyvää organisatorista prosessia on standardoitu ja uusia malleja kehitetään (Bullock, 1999). Organisaation tasolla tapahtuvasta kattavasta standardoinnista ja käytettävien tallennusmuotojen keskitetystä valinnasta on myös saatu hyviä kokemuksia pitkäaikaissäilyvyyden varmistamisessa (Salminen & al., 2001).

Yleisesti standardit voidaan jaotella kolmeen kategoriaan. *De facto* -standardi tarkoittaa epävirallista standardia, joka on päässyt niin hallitsevaan asemaan, että sen määrittelijän kanssa kilpailevat valmistajat tai muut tahot käyttävät sitä vertailussa referenssinä ja yhteensopivuusmuutokset ovat standardin tuottajan hallittavissa. Esimerkiksi tätä kirjoittaessa Microsoft Word on *de facto* -standardi tekstinkäsittelyssä. Usein myös tietyn alan johtavat valmistajat muodostavat konsortiumin sopiakseen tuotteiden yhteensopivuuden varmistavasta standardista. Tällaisesta standardista käytetään nimitystä *PAS (Publicly Available Specification)* eli standardi, joka pohjautuu julkisesti saatavilla olevaan spesifikaatioon. Kolmas ryhmä muodostuu *de jure* -standardeista, jotka ovat virallisten standardointiorganisaatioiden määrittelemiä. Tällainen on esimerkiksi rakenteisten tekstien kuvaamisessa käytettävä SGML. Yleensä *de facto* -standardit muuttuvat vakiintuessaan lopulta *de jure*-standardeiksi. (DLM-forum, 1997) Pitkäaikaissäilytyksen näkökulmasta standardit voidaan jakaa myös avoimiin ja suljettuihin standardeihin. Avoimet standardit ovat yleisesti saatavilla ja käytettävissä,

suljetut standardit ja niiden spesifikaatiot ovat puolestaan tietyn valmistajan omaisuutta, eikä niiden käyttäminen ilman lupaa ole mahdollista.

Pitkäaikaissäilytyksessä on luonnollisesti kannattavinta käyttää avoimiin ja virallisiin standardeihin pohjautuvia tallennusmuotoja. Käytännössä ratkaisevana on kuitenkin pidetty tietyn tallennusmuodon käytön laajuutta sekä sen taustalla olevan valmistajan pysyvyyttä. Digitaalisten dokumenttien pitkäaikaissäilytyksessä käytettävät tallennusmuodot ovat jossain määrin vakiintuneet ja niiden määrä on viime vuosina jopa vähentynyt pienempien valmistajien hävitessä markkinoilta. Tätä kirjoittaessa tekstimuotoinen tieto tallennetaan esimerkiksi ASCII-muotoon, rakenteisen tekstin kohdalla käytetään virallisesti standardoitua SGML-merkkäusta tai siihen pohjautuvia HTML- ja XML-kieliä. Kuvissa vakiintuneita tallennusmuotoja ovat TIFF, JPEG ja PDF. Myös sähköisten asiakirjojen de facto –standardiksi on vakiintunut PDF. (Hedstorm & Montgomery 1998; Carrol & Hodge, 1999) Jotkut standardit kuten SGML ovat myös osoittautuneet suhteellisen pysyviksi.

Bullock (1999) suosittelee standardien soveltamiseksi kolmevaiheista lähestymistapaa. Uusien digitaalisten objektien luonnissa käytetään tämän hetkisiä standardeja, toisessa vaiheessa niiden muuttumista seurataan ja kolmannessa vaiheessa objektit muutetaan uusien standardien mukaisiksi migraatiota hyödyntäen, jolloin kierros alkaa periaatteessa alusta. Seuraava esimerkinomainen tallennusmuotoihin liittyvää päätöksentekoa kuvaava malli (KUVIO 5) on luotu käyttäen pohjana DLM-forumin (1997) esittämää mallia. Mallissa tarkastellaan standardien soveltamista ennen kaikkea dokumenttiedon näkökulmasta, mutta vastaavat vaiheet löytyvät käsiteltäessä muunkin tyyppistä tietoa.



KUVIO 5. Esimerkki tallennusmuotojen valintaan liittyvästä päätöksentekoprosessista (DLM-forum, 1997).

Mallissa tarkastellaan aluksi tukeeko sovellus tai tietojärjestelmä standardinmukaista tallennusmuotoa. Mikäli sovellus ei tue tällaista tallennusmuotoa tarkistetaan konvertointisovelluksen olemassaolo, jolla sovelluksen tuottama tieto voidaan tarvittaessa muuttaa standardimuotoon. Ellei tähän tarkoitukseen sopivaa konvertointisovellusta ole saatavilla, vaihtoehdoksi jää dokumentin muuttaminen ohjelmistoriippumattomaan muotoon tai uuden konvertointisovelluksen kehittäminen. Valintaan vaikuttavat dokumentin rakenteessa tapahtuvien menetysten hyväksyttävyyss sekä käytettävissä olevat resurssit. Esimerkiksi matemaattisia kaavoja sisältävää dokumenttia ei voida muuttaa ASCII-muotoon ilman kaavojen menetystä. Toisaalta oman käännössovelluksen rakentaminen voi olla liian kallista. Tällaisessa tilanteessa

järkevin vaihtoehto onkin jatkaa nykyisen sovelluksen ylläpitoa, kunnes parempi ratkaisu löytyy.

3.1.4 Metatieto

Digitaalisen tiedon pitkäaikaissäilytyksen tueksi tarvitaan myös metatietoa eli tietoa tiedosta. Käsitteellä metadata viitataan puolestaan koneluettavaan eli formaaliin metatietoon. On myös huomattava, ettei metatietoa ole mahdollista käsitellä tämän työn puitteissa kovinkaan laajasti ja siksi esimerkiksi metatietostandardeja ei voida esitellä tarkemmin. Kokonaan aihepiiriä ei kuitenkaan voida ohittaa tiedon pitkäaikaissäilytystä käsiteltäessä.

Dempsey ja Heery (1998) määrittelevät *metatiedon* tarkoittavan informaatio-objekteihin liittyvää tietoa, joka auttaa potentiaalista käyttäjää saamaan tietoa varsinaisen objektin olemassaolosta ja ominaisuuksista. Käyttäjä voi tässä yhteydessä olla tietokoneohjelma tai henkilö. Konkreettinen esimerkki metatiedosta voisi olla esimerkiksi kirjastokortisto, joka sisältää dataa objektien eli tässä tapauksessa kirjojen ominaisuuksista. Digitaalisen tiedon yhteydessä metatietoa käytetään monenlaisiin tarkoituksiin kuten tiedonhakuun ja oikeuksien hallintaan, mutta sillä on myös erittäin merkittävä rooli sekä emuloinnissa että migraatiossa. Metatietoa tarvitaan digitaaliseen objektin tekniseen ja organisatoriseen kontekstiin liittyvän tiedon sekä migraatiohistorian tallentamiseen. (Day, 1999). Toisin sanoen metatieto tukee informaatio-objektin eri aspektien kuten autenttisuuden ja provenienssin säilymistä. Tämä on yksi metatiedon keskeisimmistä tehtävistä myös analogisen tiedon pitkäaikaissäilytyksessä, mutta digitaalisen tiedon yhteydessä sen merkitys korostuu tiedon itsetulkitsevuuden puuttuessa.

Millaista metatietoa sitten käytännössä tarvitaan pitkäaikaissäilytyksen tueksi? Bullockin (1999) mukaan tällaista metatietoa saattaisivat olla objektin tunnistamisessa käytettävä yksiselitteinen tunnus (ISBN, URL), tiedot tallennusmedioista ja –muodoista, objektin migraatiohistoria, autenttisuuden selvittämistä tukevat tiedot, oikeuksien hallintaan liittyvät tiedot sekä versiot ja päivämäärät. Lista on kuitenkin ainoastaan suuntaa-antava ja tarkkojen metatietokenttien sekä niihin liittyvien

standardien määrittelemiseksi tehdään runsaasti tutkimusta (Day, 1999; Hakala, 2001; Stenvall, 2001). Kovin vakiintuneita standardeja ei kuitenkaan ole vielä olemassa ja digitaalisiin dokumentteihin liittyvää metadataa on tutkittu tähän mennessä niukasti kirjasto- ja arkisto-organisaatioiden ulkopuolella (Päivärinta, Tyrväinen & Ylimäki, 2002). Edellä mainituista lähteistä löytyy lisätietoa myös metatietostandardeista, joita ei tämän gradun puitteissa ollut mahdollista käsitellä.

3.2 Organisatoriset lähestymistavat

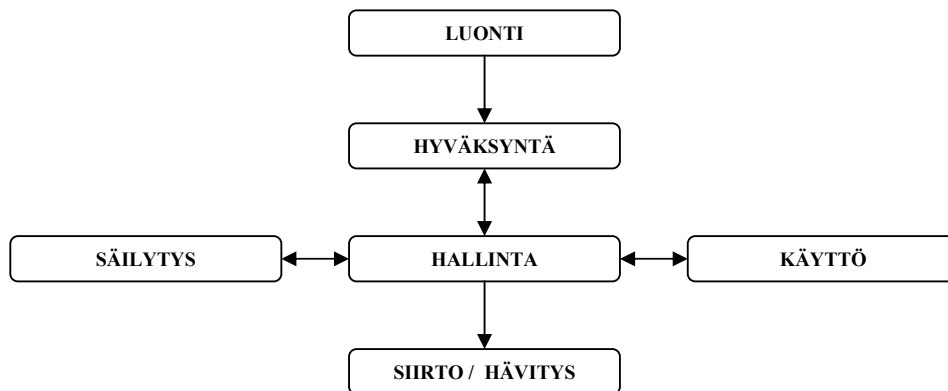
Kuten aiemmin todettiin, on tiedon pitkäaikaissäilytys suurelta osin myös organisatorinen ongelma. Teknisiä ratkaisuja on olemassa, mutta niiden käyttö ja käytön suunnittelu on hankalaa vanhojen toimintatapojen hallitsemisessa organisaatioissa. Seuraavassa luvussa esitellään lähestymistapoja organisatoristen ongelmien hallitsemiseksi. Yksi keskeisimmistä lähestymistavoista on soveltaa digitaaliseen tietoon elinkaariajattelua.

3.2.1 Informaation elinkaarimalli ja jatkumomalli

Perinteisesti informaation elinkaari on ymmärretty kaksiosaiseksi kokonaisuudeksi. Dokumentti on aktiivisessa vaiheessa tukiessaan organisaation operatiivista toimintaa. Tämän jälkeen dokumentti hävitetään tai siirretään arkistoon pitkäaikaissäilytystä varten, jolloin dokumentti siirtyy passiiviseen vaiheeseen. Tämä lähestymistapa on kuitenkin riittämätön, mikäli kyseinen dokumentti on digitaalisessa muodossa. Säilymisen varmistaminen digitaalisessa toimintaympäristössä voi nimittäin olla myöhäistä, mikäli siihen johtava päätös tehdään vasta dokumentin aktiivisen elinkaaren loppuvaiheessa. (Day, 1999) Digitaalisen objektin eri aspekteihin ja pitkäaikaissäilytykseen on kiinnitettävä huomiota tiedon elinkaaren alkuvaiheista lähtien. Tämän toteavat myös Carrol & Hodge (1999) pitkäaikaissäilytykseen liittyviä projekteja kartoittavassa tutkimuksessaan.

DLM-forum (1997) kuvaa informaation elinkaarta kuvassa (KUVIO 6) esitetyllä mallilla. Lähde on kuitenkin sikäli puutteellinen, ettei kuvassa esiteltyjä vaiheita

käsitellä työssä kovinkaan yksityiskohtaisesti. Lähes vastaava, mutta paremmin dokumentoitu viitekehys löytyy kuitenkin lähteestä Beagrie & al. (2001). Käyttäen molempia edellä mainittuja lähteitä toisiaan täydentävästi esitellään seuraavassa luetelmassa informaation elinkaaren vaiheet. Vaiheet on kuvattu melko yleisluontoisesti, koska niitä sovelletaan eri organisaatioissa hyvin vaihtelevasti. Tarkoituksena onkin enemmän pitkäaikaissäilytyksen tarkasteluun soveltuvan viitekehysten esittely, kuin elinkaaren tarkka analysointi.



KUVIO 6. Informaation elinkaari pitkäaikaissäilytyksen näkökulmasta (DLM-forum, 1997).

Luonti / hankinta: Tässä vaiheessa digitaalinen objekti luodaan, joko ”tyhjästä” tai analogisessa muodossa oleva dokumentti muutetaan digitaaliseksi. Dokumentti voidaan myös hankkia, jolloin olemassa oleva objekti tai siihen liittyvä vastuu siirtyy organisaatiolle esimerkiksi toisesta organisaatiosta. Vaiheessa päätetään useista seikoista kuten dokumentin sisällöstä, rakenteesta, tallennusmuodoista ja sen kuvaamiseen käytettävästä metadatatista (Beagrie & al., 2001). Vaikutusmahdollisuudet ovat luonnollisesti vähäisemmät, mikäli dokumentti on saatu organisaation ulkopuolelta. Kuten aiemmin todettiin, vaiheessa tehtävät päätökset vaikuttavat ratkaisevasti tiedon hallitsemiseen ja säilyttämisen onnistumiseen tulevaisuudessa. Hodge (2000) suositteleeikin selvityksessään käytäntöä, jossa digitaalisen dokumentin

pitkäaikaissäilytyksen tarve arvioidaan heti dokumentin luonnin yhteydessä. Tämän on havaittu myös käytännössä antavan hyvän kuvan dokumentin arvosta.

Hyväksyntä (appraisal): Hyväksymisen yhteydessä päätetään dokumentin hyväksymisestä organisaation kokoelmaan tai tietojärjestelmään. Dokumentin hyväksyttävyyden arvioinnissa käytetään monenlaisia kriteerejä ja niiden yhdistelmiä riippuen tilanteesta ja organisaatiosta. Pitkäaikaissäilytyksen näkökulmasta arvioitavia seikkoja ovat dokumentin arvo, eheys, autenttisuus, tallennusmuoto ja kuinka hyvin se noudattaa organisaation käyttämiä standardeja (DLM-forum, 1997, 27 - 28). Hyväksymisvaiheessa tulee huolehtia myös kaikkien tarpeettomien tai käyttökelvottomien dokumenttien hävittämisestä. Dokumentti on syytä myös konvertoida mahdollisuuksien mukaan organisaation käyttämään standardimuotoon.

Hallinta: Digitaalisen tiedon hallinnan tavoitteena on säilyttää digitaalinen objekti ja siihen liittyvät aspektit käyttökelpoisena ajan kuluessa (DLM-forum, 1997, 30). Beagrie & al (2001) mukaan tähän vaiheeseen kuuluvat päätökset digitaalisen objektin käsittelystä ja sen kuvaamiseen tarvittavasta metatiedosta. Kuten todettiin, tällaista metatietoa ovat tiedot objektin rakenteesta, sisällöstä, kontekstista, provenienssista ja historiasta. DLM-forum (1997) toteaa tämän olevan yksi vaativimmista ja tärkeimmistä tehtävistä, koska vain objektin ja sen ympäristön tarkka kuvaaminen tekee digitaalisesta dokumentista todistusvoimaisen. Luonnollisesti myös itse digitaalisen objektin on säilyttävä ja seuraavaksi tarkastellaankin digitaalisen tiedon säilytystä.

Säilytys: Säilytys kattaa sekä tiedon operatiiviseen varastointiin että pitkäaikaissäilytykseen liittyvät päätökset ja toimenpiteet. Tiedon varastointiin kuuluvat päätökset käytettävistä tallennusmuodoista, ohjelmisto- ja laitteistoalustoista, tietojen mahdollisesta hajauttamisesta, tietoturvasta sekä tallennusmedioiden kunnosta huolehtiminen. Päämääränä on digitaalisten objektien eheyden ja käyttökelpoisuuden varmistaminen lyhyellä tähtäimellä. Pitkällä aikavälillä vaiheeseen kuuluvat myös objektin ja sen aspektien säilyttäminen mahdollisimman muuttumattomana teknologisesti muutoksesta huolimatta. Tässä vaiheessa sovelletaan tarvittaessa erilaisia teknisiä lähestymistapoja kuten migraatiota ja emulointia. (Beagrie & al, 2001)

Käyttö: Digitaalinen objekti on hyödytön, mikäli siihen ei päästä käsiksi käyttäjien haluamalla tavalla haluttuna ajankohtana. Vaiheen toteutusta suunniteltaessa on tarkasteltava laajasti digitaalisen objektin käyttäjien ja mahdollisten käyttäjien tarpeita sekä mahdollisuuksia vastata näihin vaatimuksiin käytettävissä olevien resurssien puitteissa. Pohdittavia seikkoja ovat esimerkiksi tiedon jakaminen muiden organisaatioiden kanssa, lailliset ja taloudelliset vaatimukset, tietoverkkojen tarjoamat mahdollisuudet sekä käyttäjien laitteisto- ja ohjelmistoympäristöihin kohdistuvat vaatimukset. (Beagrie & al, 2001)

Siirto / hävittäminen: Tiedon aktiivisen elinkaaren päättyessä päätetään digitaalisen objektin hävittämisestä tai siirtämisestä arkistoon passiivista säilytystä varten. Tosin säilytys on digitaalisten objektien kohdalla erittäin harvoin täysin passiivista. Vaihe muistuttaa monessa suhteessa hyväksymisvaihetta ja kyseessä voi olla joko informaatio-objektin tai siihen liittyvän vastuun siirtäminen toiseen organisaatioon tai saman organisaation toiselle osalle, joka vastaa tiedon pitkäaikaissäilytyksestä. (DLM-forum, 1997) Objektin arkistointikelpoisuus tarkistetaan ja tarvittaessa se muunnetaan paremmin arkistoitavaan tallennusmuotoon.

Varsinaisen elinkaarimallin lisäksi arkistotieteessä on jo vuosikymmeniä keskusteltu myös toisenlaisista pitkäaikaissäilytyksen hallintaa käsittelevistä malleista. Malleista käytävä tieteellinen keskustelu onkin voimistunut digitaalisten tallennusmuotojen arkistoinnille tuomien erityisongelmien myötä. Näiden mallien joukossa eräs suosituimmista on ollut 90-luvun alkupuolella syntynyt jatkumomalli (records continuum model), joka on saanut kannatusta joidenkin arkistotieteilijöiden keskuudessa. Jatkumomalli poikkeaa elinkaarimalleista siten, ettei siinä jaotella informaatio-objektin elinkaarta erillisiin vaiheisiin, vaan sen hallinnan katsotaan olevan jatkuva prosessi. Se suhtautuu elinkaarimallista poikkeavalla tavalla myös vastuukysymyksiin. Jatkumomallin mukaan aineistoon liittyviä vastuita ei pidä jakaa eri tahojen kesken, vaan niitä on pyrittävä enemmänkin yhdistämään. Jatkumomalli onkin herättänyt kiivasta tieteellistä keskustelua, eikä tämä ole ihme, sillä käytännössä mallin noudattaminen merkitsisi perinteisen arkistotoimen asteittaista katoamista. (Bantin, 2000)

3.2.2 Vastuu ja pitkäaikaissäilytyksen organisointi

Karkeasti sanottuna tiedon pitkäaikaissäilytys voidaan organisoida kahdella eri tavalla (Bantin, 2000). Ensinnäkin se voi muodostaa organisaatiossa keskitetyn arkistointifunktion, joka on ”irrallaan” organisaation jokapäiväisestä toiminnasta. Tällöin ideana on, että vastuu tiedosta siirtyy tälle keskitetylle erillisfunktiolle organisaation muilta osilta tiedon siirtyessä elinkaarensa passiiviseen vaiheeseen. Juuri tämä onkin se tapa, jolla tiedon pitkäaikaissäilytys on perinteisesti järjestetty organisaatiossa (Dollar, 1992, 52). Toinen tapa tiedon pitkäaikaissäilytyksen organisoimiseksi on hajauttaa arkistointi osaksi muita organisaation jokapäiväisiä prosesseja ja funktioita, jotka vastaavat yleisestä informaationhallinnasta ja siinä tarvittavasta informaatioteknologiasta.

Edellä esiteltyjen lähestymistapojen eduista ja haitoista käytävä keskustelu on voimistunut digitaalisten tallennusmuotojen yleistymisen myötä. Bantin (2000) käykin artikkelissaan läpi keskitettyä mallia puolustavien ja vastustavien tutkijoiden esittämiä perusteluja. Hänen mukaansa keskitetyn arkistointimallin kannattajat pitävät sen keskeisenä etuna arkistoinnin vaatiman erityisosaamisen keskittymistä yhteen paikkaan. He myös epäilevät muista tehtävistä vastaavien työntekijöiden halua ja kykyä huolehtia arkistomateriaalin säilymisestä. Lisäksi on pelätty eturistiriitojen syntymistä, mikäli samat tahot vastaavat sekä tiedon tuottamisesta että sen säilyttämisestä. Hajautettua organisaatiomallia on puolestaan puolustettu kustannustekijöille ja uusien digitaalisten tallennusmuotojen vaatimalla erityisosaamisella. Nykyisissä tietointensiivisissä organisaatioissa tiedon säilyttäminen erillisessä organisaation osassa on katsottu vaativan liikaa resursseja. Toisaalta on myös varsin kyseenalaista, pystyvätkö arkistoalan asiantuntijat hankkimaan ja ylläpitämään riittävää informaatio- ja viestintäteknologian osaamista vastatakseen näitä teknologioita hyödyntävän tiedon säilymisestä.

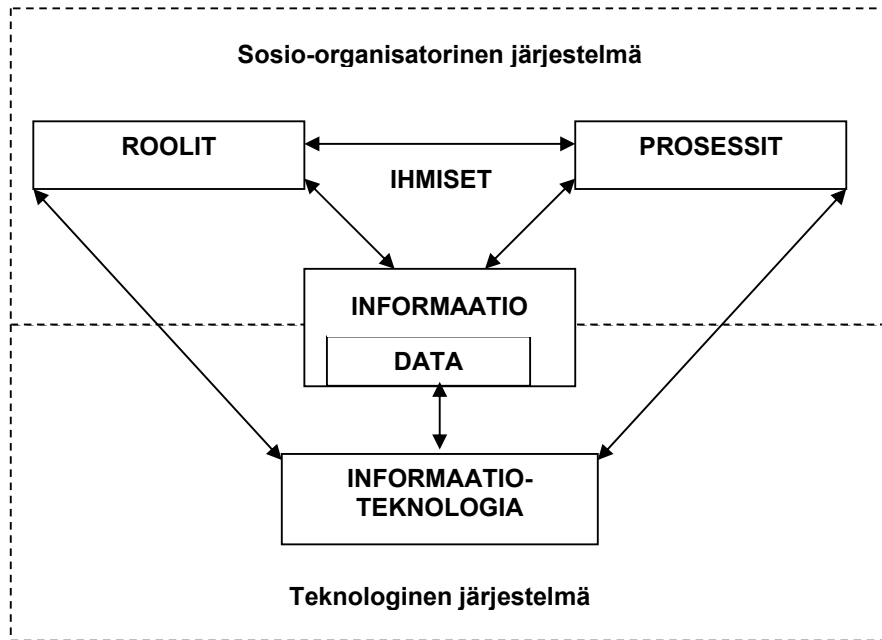
Tässä kirjallisuuskatsauksessa aiemmin tehtyjen havaintojen perusteella on syytä uskoa hajautetun mallin toimivuuteen tulevaisuudessa. Digitaalinen tieto on muutenkin monimutkaistanut pitkäaikaissäilytyksen organisointia ja siihen liittyviä

vastuukysymyksiä. Kuten kohdassa 3.2 todettiin, henkilökohtaisten tietokoneiden ja tehokkaiden tietojärjestelmien yleistymisen myötä vastuu pitkäaikaissäilytyksestä on siirtymässä yhä pienemmille yksiköille. Hodge (2000) toteaaakin, että digitaalisen tiedon aikakaudella ryhmät ja yksilöt ovat lähes huomaamattaan jo nyt joutuneet arkistojän rooliin organisaation rakenteessa tapahtuneiden muutosten ja muiden digitaalisen teknologiaan liittyvien tekijöiden seurauksena. Tämän perusteella vaikuttaisi siltä, että teknologisen muutoksen hallitsemassa ympäristössä tiedon pitkäaikaista säilytystä on mahdoton erottaa sen lyhytaikaisesta hallinnasta. Myös Dollar (1992, 53 - 55) arvioi työssään erillisen pitkäaikaissäilytysfunktion katoavan tiedon digitalisoitumisen seurauksena. Mielenkiintoisempi kysymys sitten onkin, onko tämä 11 vuotta sitten lausuttu ennustus vielä lainkaan toteutunut.

Ottaen huomioon digitaalisten tallennusmuotojen vaikutuksen pitkäaikaissäilytyksen vastuukysymyksiin ja organisointiin on yllättävää, että tiedon pitkäaikaissäilytystä osana yleisempää organisaation informaationhallintaa käsittelevää kirjallisuutta on hankala löytää. Tämän vuoksi tässäkin työssä tarkoitukseen soveltuvaa mallia jouduttiin hakemaan organisaation informaatioresurssien hallinnan yleisempään tarkasteluun tarkoitettujen viitekehysten joukosta. Tietojärjestelmätieteessä organisaation informaatioresurssien hallinnan tarkasteluun soveltuvia viitekehyksiä ja malleja onkin viimeisten vuosikymmenien aikana esitetty lukuisia, eikä niiden pintapuolinenkaan esittely ole tässä yhteydessä mahdollista. Seuraavassa kuitenkin esitellään esimerkinomaisesti yksi viitekehys, jonka avulla voidaan tarkastella organisaation informaationhallintaa. Tämän jälkeen pohditaan pitkäaikaissäilytyksen sijoittumista tähän viitekehykseen.

Päivärinta ja Tyrväinen (2001) käyttävät esittämänsä viitekehysten (KUVIO 7) pohjana huomattavasti varhaisempaa, jo vuonna 1965 julkaistua organisatorisen muutoksen kuvaamiseen tarkoitettua timanttimallia, jota on sovellettu runsaasti myös tietojärjestelmätieteessä. Alkuperäisen malli on kuitenkin sikäli puutteellinen, ettei siinä pidetä informaatiota itsessään osana organisaatiota. Päivärinta ja Tyrväinen (2001) mukauttavatkin mallia informaatioresurssien tarkasteluun sopivammaksi lisäämällä

malliin uudeksi käsitteeksi informaation. Perusidealtaan Leavittin malli säilyy kuitenkin lähes alkuperäisen kaltaisena.



KUVIO 7. Organisaation informaatioresurssien hallinnan tarkasteluun tarkoitettu timanttimalli (Päivärinta & Tyrväinen, 2001).

Viitekehyksessä organisaation informaationhallinta jaetaan teknologiseen ja sosio-organisatoriseen osaan, joiden välisenä rajapintana toimii informaatio. Ihmiset työskentelevät organisaatiossa enemmän tai vähemmän tunnistettujen roolien ja prosessien välityksellä. *Rooli* ymmärretään tiettyyn ihmiseen tai organisaation osaan liitettäväksi oikeuksien ja velvollisuuksien kokoelmaksi. Roolilla siis viitataan organisaation hierarkkiseen rakenteeseen. *Prosessi* puolestaan viittaa organisaation varsinaisiin toimintoihin, joiden kautta organisaatio pääsee asettamiinsa tai sille asetettuihin tavoitteisiin. (Päivärinta & Tyrväinen, 2001)

Malli tarjoaa hyvän välineen myös organisaation pitkäaikaissäilytyksen tarkasteluun ja kaikki mallissa esitellyt osa-alueet tulisi huomioida sitä organisoitaessa. Kuten aiemmin todettiin, tiedon pitkäaikaissäilytystä käsittelevässä kirjallisuudessa huomio on kohdistunut lähinnä teknologiseen järjestelmään ja mallin sosio-organisatorinen osa on

jäänyt vähemmälle huomiolle. Myös seuraava pitkäaikaissäilytyksen sijoittumista malliin käsittelevä osa jää hyvin pinnalliseksi johtuen lähteiden puutteesta.

Perinteisesti tiedon pitkäaikaissäilytystä on organisatorisesta järjestelmästä katsottuna ajateltu lähinnä organisatoristen roolien kuten arkiston ja arkistonhoitajan kautta. Nyt painotus on kuitenkin kirjallisuuden valossa siirtymässä enemmän prosesseille, koska nykyaikaisissa organisaatioissa informaation määrä ja merkitys ovat kasvaneet sekä informaation elinkaaren alkupäässä tehdyt päätökset vaikuttavat sen säilyvyyteen. Käytännössä tämä tarkoittaa sitä, että tiedon pitkäaikaissäilytys organisaatiossa on muuttumassa keskitetystä vertikaalisesta funktiosta kohti hajautettua horisontaalista prosessia. Samalla tiedon pitkäaikaissäilytys liittyy osaksi organisaation yleistä informaatioresurssien hallintaa. Tällöin myös kaikki timanttimallin osatekijät tulisi huomioida sitä organisoitaessa. Esimerkiksi informaatioteknologian soveltamisessa tehdyt päätökset vaikuttavat suoraan organisaation tietoresurssien käyttökelpoisuuteen tulevaisuudessa ja siten myös niiden säilyvyyteen.

3.3 Yhteenveto

Digitaalisen tiedon säilyvyyden varmistamiseen voidaan käyttää useita erilaisia teknisiä lähestymistapoja, joista käytännössä sovelletuin paperille tulostamisen lisäksi on migraatio. Emulointi soveltuu puolestaan niihin suhteellisen harvoihin tilanteisiin, joissa halutaan säilyttää digitaalisen objektin alkuperäisen toiminnallisuus. Minkään menetelmän soveltaminen ei kuitenkaan ole yksinkertaista, saati ongelmatonta. Niiden soveltaminen vaatii runsaasti asiantuntemusta ja monimutkaista päätöksentekoa. Kaikkien menetelmien toimivuutta voidaan tehostaa soveltamalla mahdollisimman vakiintuneita standardeja sekä metatietoa.

Tehokkaammin tiedon säilyvyyteen liittyviä ongelmia voidaan kuitenkin välttää kiinnittämällä asiaan huomiota organisaation eri tasoilla mahdollisimman varhaisessa vaiheessa. Tämä pätee sekä yksittäisten työntekijän tasolla kuten myös koko organisaation informaationhallintaa suunniteltaessa. Esimerkiksi informaation elinkaaren pohjautuvat viitekehykset tarjoavat tähän tarkoitukseen sopivan työkalun.

Kirjallisuudessa on myös selvästi havaittavissa pitkäaikaissäilytyksen integroitumista korostava kehityssuunta. Teknologinen kehitys on ajamassa tiedon pitkäaikaissäilytystä ulos passiivisesta arkistosta osaksi organisaatioiden informaatioresurssien hallintaa. Yhteenvedona voidaan siis sanoa, että nykyisessä uuden teknologian hallitsemassa tilanteessa toimivan informaatioresurssien hallinta vähentää myös pitkäaikaissäilytykseen liittyviä ongelmia. Nykyisissä jatkuvasti muuttuvissa ja informaatioosaaristosta kärsivissä organisaatioissa hyvin toimiva informaationresurssien hallinta on kuitenkin usein vallitsevan asiantilan sijasta enemmänkin haastava tavoite.

4. TAPAUSTUTKIMUS – KOHDEORGANISAATIO JA TUTKIMUSPROSESSI

4.1 Kohdeorganisaation valinta ja tutkimusongelma

Tapausosuuden tarkoituksena on tarkastella tiedon pitkäaikaissäilytystä ja siihen liittyviä kysymyksiä yhdessä organisaatiossa, joka tämän tutkimuksen kohdalla on Jyväskylän yliopiston informaatioteknologian tiedekunta. Kohdeorganisaation valintaan vaikuttivat sen saatavuuden lisäksi muutkin tekijät. Tiedekunnassa on monia nykyaikaiselle organisaatiolle tyypillisiä piirteitä. Se on syntynyt kahden erillisen organisaation yhdistyessä, ollut jatkuvan muutoksen kohteena koko olemassaolonsa ajan ja sen päätuote on informaatio. Lisäksi organisaatio on kooltaan pienehkö ja sen tietovarastoja on kartoitettu aiemmissa tutkimuksissa, joiden tulokset ovat nyt käytettävissä. Tämä tekee mahdolliseksi organisaation tiedon pitkäaikaissäilytyksen kattavan selvittämisen näin pienimuotoisen tutkimuksen puitteissa. Organisaatioissa säilytetään myös monentyyppistä tietoa hallintoon tarvittavasta operatiivisesta tiedosta tutkimustuloksiin. Tämä yhdessä aiemmin esiteltyjen ominaisuuksien kanssa parantaa tutkimuksen yleistettävyyden mahdollisuuksia. Organisaation tekee aihepiirin näkökulmasta mielenkiintoiseksi myös sen digitaalisten opetusmateriaalien käyttöön tähtäävät tulevaisuudensuunnitelmat³.

Tapaus tutkimuksen keskeinen tutkimusongelma on selvittää kohdeorganisaation informaatioresurssien hallinnan tila digitaalisen tiedon pitkäaikaissäilytyksen näkökulmasta. Tähän keskeisesti liittyviä alakysymyksiä ovat, miten organisaation tietoa voidaan numeerisesti luokitella pitkäaikaissäilytyksen näkökulmasta sekä kuinka suurella osalla organisaation tiedosta on pitkäaikaista merkitystä. Näihin tavoitteisiin pääseminen edellyttää myös tiedon pitkäaikaissäilytyksen analysointiin soveltuvien mittareiden kehittämistä ja testaamista. Lisäksi osatavoitteena on tutkia, löytyykö kohdeorganisaation tietomassasta ominaisuuksia, jotka saattaisivat olla yleistettävissä

³ esim. Jyväskylän yliopiston virtuaaliyliopistostrategia <<http://virtuaaliyliopisto.jyu.fi/strategia>>

muihin organisaatioihin. Tietoon liittyvät vastuukysymykset rajautuvat pääosin tapaustutkimuksen ulkopuolelle.

4.2 Tapaustutkimuksen tausta

Kappaleessa esitellään ensimmäiseksi kohdeorganisaatio. Tämän jälkeen käydään läpi tapaustutkimuksen tausta-aineisto ja kerrotaan, miten kyseistä aineistoa täydennettiin tutkimuksen kuluessa.

4.2.1 Kohdeorganisaatio

Jyväskylän yliopiston informaatioteknologian tiedekunta syntyi, kun kaksi eri tiedekunnissa sijainnutta kyseisen yliopiston osaa yhdistyi samaksi hallinnolliseksi kokonaisuudeksi vuonna 1999. Näitä liitoksen tuloksena muodostuneen tiedekunnan laitoksia ovat tietotekniikan laitos (TTL) ja tietojenkäsittelytieteiden laitos (TKTL), johon kuuluu erillisenä osana myös tietotekniikan tutkimusinstituutti (TITU). Tässä tutkimuksessa keskitytään kahteen ensimmäiseksi mainittuun laitokseen ja TITU rajataan tutkimuksen ulkopuolelle. Kahdessa tarkasteltavassa laitoksessa on yhteensä noin 1700 pääaineopiskelijaa ja hieman alle 200 muuta työntekijää. Organisatorisesti tiedekunta on melko yksinkertainen, molemmilla laitoksella on oma pienehkö hallintohenkilökuntansa tiedekunnan yhteisen hallintohenkilöstön lisäksi.

Informaatioteknologiaa tiedekunnassa sovelletaan luonnollisesti runsaasti niin hallinnossa kuin opetustoiminnassakin. Kovin yhtenäisiä järjestelmiä ei kuitenkaan ole käytössä ja tiedekunnassa on useita erilaisten järjestelmien muodostamia saarekkeita, jotka ovat osittain myös päällekkäisiä. Hallinnossa ja tiedon tuottamisessa käytetään pääosin toimisto-ohjelmistojen tekstinkäsittely- ja taulukkolaskentasovelluksia. Lisäksi käytössä on kurssitietojen hallintaan, tilavarauksiin, hankkeisiin, opintosuorituksiin sekä taloushallintoon liittyviä itsenäisiä järjestelmiä ja tietokantoja.

Tiedon pitkäaikaissäilytys tiedekunnassa tapahtuu hyvin perinteisesti. Sen pohjana toimii Jyväskylän yliopiston arkistosääntö sekä siihen liittyvä arkistointisuunnitelma, jossa kuvataan säilytettävät asiakirjat ja niiden säilytysajat. Arkistosäännön mukaan erillisyksiköt ovat vastuussa tiedon pitkäaikaissäilytyksestä yksikön esimiehen nimeämän vastuuhenkilön välityksellä. Käytännössä tämä toimii siten, että erillislaitokset toimittavat arkistointisuunnitelmassa pitkäaikaissäilytettäväksi mainitun materiaalin yliopiston arkistolle, jolloin myös vastuu tiedosta siirtyy laitoksilta yliopiston hallintovirastolle. Virallisesti pitkäaikaissäilytettävät asiakirjat säilytetään tiedekunnassa lähes kokonaan paperimuodossa. Ainoan poikkeuksen muodostaa opiskelija- ja opintosuoritusrekisteri, joka on keskitetty kaikkien yliopiston laitosten osalta opiskelijapalvelun ylläpitämään tietokantaan.

4.2.2 Tutkimusmenetelmä

Tapaustutkimuksen lähtökohtana toimii kohdeorganisaatiossa syksyllä 2001 toteutettu tutkimusprojekti, PROMI, jossa kartoitettiin yksikössä kulkevia tietovirtoja genrepohjaisesti. Kuten johdannossa mainittiin, genrellä tarkoitetaan tässä työssä organisaatiossa yhteisesti ymmärrettyä toistuvaa kommunikatiivista ilmaisua tai tilannetta, joka voidaan tunnistaa samanlaisen sisällön ja muodon perusteella (Yates & Orlikowski, 1992). Käytännön genrejä kohdeorganisaatiossa ovat esimerkiksi matkalasku, tutkimussuunnitelma, luento ja tenttivastaus. Genren tunnettavuudessa ja standardoinnin asteessa voi olla eroja, eikä yksittäinen organisaation jäsen useinkaan tunnista kaikkia organisaation genrejä. Organisaation toisiinsa liittyvät genret muodostavat *genrejärjestelmän*. Tyypillisesti tiettyyn organisatoriseen prosessiin liittyy oma genrejärjestelmänsä. (Yates, Orlikowski ja Rennecker 1997, 51) Esimerkiksi tiedekunnassa järjestettäviin tentteihin voi liittyä genrejärjestelmä, johon kuuluvia genrejä ovat tenttitilavaraus, tenttikysymys, tenttivastaus ja tenttitulos. Tietojärjestelmätieteessä genrejä on sovellettu tietojärjestelmien suunnittelussa (Päivärinta, Halttunen ja Tyrväinen, 2001) sekä mitattaessa organisaatioiden informaatioresurssien määrää, laatua ja digitalisoitumisen astetta (Tyrväinen & Päivärinta 2003). Ensimmäiseksi mainittua menetelmää sovellettiin PROMI-projektissa

kohdeorganisaation genrejä tunnistettaessa. Tässä työssä näitä tuloksia puolestaan täydennetään viimeksi mainittua mittausmenetelmää soveltaen.

Tapaustutkimuksen aineistoa ei siis kerätty kohdeorganisaation arkistointia koskevan ohjeistuksen pohjalta, vaan lähteessä Tyrväinen & Päivärinta (2003) kuvattua organisaation tietoresurssien kartoitukseen tarkoitettua genre-pohjaista analyysimenetelmää soveltaen. Syitä tähän oli useita. Ensinnäkin näin voidaan arvioida kyseisen menetelmän soveltuvuutta tiedon pitkäaikaissäilytyksen kartoittamiseen organisaatiossa ja hyödyntää kohdeorganisaatioon aiemmin kohdistuneen PROMI-projektin tuloksia. Toisaalta mainittua menetelmää käyttäen tutkimus ei rajoitu ainoastaan tällä hetkellä arkistosuunnitelmassa kuvattuun pitkäaikaissäilytettävään aineistoon. Tällöin kyseisen aineiston määrää voidaan verrata organisaatiossa tapahtuvan viestinnän kokonaismäärään ja tarkastelun kohteeksi saadaan myös sellainen materiaali, jotta ei tällä hetkellä säilytetä pitkäaikaisesti, mutta joka voi esimerkiksi organisaatiossa käytettävän teknologian muuttuessa tulla säilytysvelvollisuuden alaiseksi.

Kohdassa 4.2.3 kuvatussa tiedonkeräysprosessissa käytettävät muuttujat ovat pääosin samoja kuin Tyrväisen ja Päivärinnan (2003) kuvaamassa analysointimenetelmässä. Myös muuttujien laskentaperiaatteet ovat lähes vastaavia. Joiltakin osin tässä tutkimuksessa käytettävä menetelmä kuitenkin eroaa alkuperäisestä. Merkittävimmät poikkeukset ovat genrejen säilytysajan ottaminen mukaan tiedonkeräykseen ja alkuperäistä menetelmää yksinkertaisempi tapa luokitella organisaatiossa käytettäviä medioita. Nämä ja muut poikkeamat mainitaan erikseen tiedonkeräysprosessin kuvauksessa. Muilta osin tapaustutkimuksessa käytettävät muuttujat ovat samoja kuin Tyrväisen ja Päivärinnan (2003) menetelmässä. Arkistointisuunnitelmaa hyödynnettiin lisätietolähteenä tutkimuksen kuluessa.

Aiemmin mainitussa PROMI-projektissa tunnistettiin ja nimettiin kohdeorganisaation genret sekä niitä hyödyntävät entiteetit eli informaation tuottajat ja käyttäjät soveltaen lähteessä Päivärinta, Halttunen ja Tyrväinen (2001) kuvattua suunnittelumenetelmää. Nämä tiedot kerättiin kyseisessä projektissa tietolistoihin, joista selviää genre, sen

tuottaja/lähde sekä tiedon käyttäjä/kohde. Myös jokaisen genren luontiin ja käyttöön liittyvät prosessit kartoitettiin mainitussa selvityksessä. Tietolistat on jaoteltu karkeasti organisaation toimintojen mukaisesti, joita tutkimuksessa olivat opintohallinto, opetus ja opiskelu, projektit, jatkotutkimus sekä talous- ja henkilöstöhallinto. Oheisessa taulukossa (TAULUKKO 1) on näyte eräästä projektin tuloksia sisältävästä tietolistasta. Kohdeorganisaatiosta löydettiin kartoituksessa yhteensä 310 erillistä genreä.

TAULUKKO 1. Näyte eräästä PROMI-projektin tietolistasta, joka sisältää genren nimen sekä sen tuottajan ja käyttäjän.

Lähde/tuottaja	Genre / Dokumentti	Kohde/käyttäjä
HALLINTO-PÄÄLLIKKÖ	Tutkintoasetus	AMANUENSSI
HALLINTO-PÄÄLLIKKÖ	Opiskelijavalinnan perusteet	AMANUENSSI
HALLINTO- PÄÄLLIKKÖ	Muuntopyrkijät-lista+alustava esitys	AMANUENSSI
HALLINTO- PÄÄLLIKKÖ	Pääsykoe-opiskelijavalinta	OPETTAJA
HALLINTO- PÄÄLLIKKÖ	Oikaisuvaatimus	OPETTAJA
HALLINTO- PÄÄLLIKKÖ	Opiskelijahaku-ilmoitukset	OPISKELIJA
HALLINTO- PÄÄLLIKKÖ	Hylkäyskirje (sis. Oikaisuos.)	OPISKELIJA

4.2.3 Aineiston keräys

Tämän tutkimuksen ensimmäisessä vaiheessa edellä esitellystä genrelistasta (TAULUKKO 1) poistettiin organisatoristen prosessien välivaiheiden tukena toimineet genret. Esimerkiksi opiskelijavalinnassa alustavia, päätöksenteon tukena toimineita genrejä ei analysoitu, vaan ainoastaan lopullinen valintapäätös otettiin huomioon. Näin huomio saatiin kiinnitettyä organisatoristen prosessien sijasta niiden lopputuloksiin, eli varsinaiseen säilytettävään materiaaliin. Samalla listasta poistettiin tunnistamattomat ja päällekkäiset genret. Toisessa vaiheessa suodatuksen tuloksena syntyneitä 143 genreä käsittävää listaa täydennettiin keräämällä tiedot Tyrväisen ja Päivärinnan (2003) menetelmässään kuvaamista genrekohtaisista muuttujista. Samalla kerättiin myös muuta tiedon pitkäaikaissäilytyksen kannalta merkityksellistä lisätietoa. Tämä tapahtui keväällä 2002 haastattelemalla tiedekunnan hallintopäällikköä, tietojenkäsittelytieteen laitoksen varajohtajaa ja amanuenssia sekä tiedekunnan talousasioista vastaavaa

osastosihteeriä. Haastateltavien kanssa käytiin läpi edellä kuvatun suodatuksen tuloksena syntynyt genrelista ja kerättiin mahdollisimman monesta listan genrestä ohessa esitellyt tiedot. Seuraavassa luettelussa esiteltävät muuttujat ovat Tyrväisen ja Päivärinnan (2003) menetelmäkuvauksen mukaisia, ellei toisin mainita:

Uniikki ilmentymä (UI): Ainutkertaisella eli uniikilla ilmentymällä viitataan tietyn genren erilliseen ilmentymään. Esimerkiksi jokainen erillinen matkalasku on matkalasku-genren uniikki ilmentymä. Kenttä kertoo tietyn genren ilmentymien määrän kappaleina organisaatiossa vuositasona. Tätä kysyttäessä haastateltavia pyydettiin esimerkiksi arvioimaan, montako matkalaskua organisaatiossa syntyy vuositasona.

Kopiot/Ilmentymä (K / UI): Kyseinen luku kertoo kuinka monta kopiota yhdestä tietyn genren uniikista ilmentymästä otetaan keskimäärin. Kaikista ilmentymistä ei välttämättä oteta kopiota. Luvun avulla voidaan laskea yhdestä ainutkertaisesta ilmentymästä luotavien kopioiden määrä. Luku lasketaan kappaleina vuositasona, esimerkiksi montako kopiota yhdestä matkalaskusta otetaan keskimäärin sen käsittelyn aikana.

Sivut (S): Tämä kertoo kuinka paljon tietyn genren uniikki ilmentymä sisältää keskimäärin informaatiota. Tämä on siis eri asia, kuin informaation vaatima tallennustila, joka riippuu käytetystä tallennusmediasta. Yksi puhtaasti tekstiä sisältävä sivu vie ASCII-muodossa tilaa arviolta 3 – 5 kilotavua tai bittikarttakuvana noin yhden megatavun. Sivun on ymmärrettävä tässä ihmisen havaitsemaksi visuaaliseksi sivuksi, ei niinkään tarkaksi A4-sivuksi, vaikka se pääosin tarkoittaakin juuri sitä. Luku voi olla myös pienempi kuin yksi, mikäli kyseessä on genre, jonka yksittäinen ilmentymä on hyvin pieni. Tältä osin tässä käytettävä sivu-käsite poikkeaa myös Tyrväisen ja Päivärinnan (2003) menetelmästä. On myös huomattava, että tuloksissa esiteltävien suullisten genrejen kokoyksikkönä käytetään tapaustutkimuksessa myös sivua. Tällöin viitataan yksittäisen suullisen ilmentymän keskimääräiseen kokoon, jos se kirjoitettaisiin paperille tekstimuotoiseksi ”käsikirjoitukseksi”.

Säilytysaika: Säilytysajalla viitataan tietyn genren ilmentymien säilytysaikaan organisaatiossa. Tämä genrekohtainen muuttuja otettiin käyttöön tätä tutkimusta varten, eikä sitä mainita Tyrväisen ja Päivärinnan (2003) menetelmässä. Säilytysaika selvitettiin tarkistamalla se ensisijaisesti organisaatiossa käytössä olevasta arkistointisäännöstä. Mikäli tietoa ei löytynyt arkistointisäännöstä, selvitettiin organisaatiossa vallitseva käytäntö tarkasteltavan genren osalta. Mikäli säilytysaika vaihteli saman genren eri ilmentymien välillä, valittiin pisin käytetty säilytysaika. Genret jaettiin säilytysajan perusteella kolmeen luokkaan. Luokituksella haluttiin erottaa toisistaan operatiivinen ja säilytettävä aineisto. Säilytettävä aineisto jaettiin puolestaan kahteen luokkaan keskipitkästi säilytettävän aineiston ja pitkäaikaisesti säilytettävän aineiston erottamiseksi toisistaan. Seuraavassa esiteltävät aikarajat muodostuivatkin melko luontevasti kohdeorganisaation omien säilytysaikasuosituksen perusteella. Genre kuuluu operatiiviseen luokkaan, mikäli sen säilytysaika on vuosi tai vähemmän. Keskipitkän säilytysajan luokkaan kuuluvia genrejä säilytetään 1 – 5 vuotta. Pitkän säilytysajan luokkaan genre kuuluu, mikäli sitä säilytetään enemmän kuin 5 vuotta.

Medialuokka: Haastatteluissa kartoitettiin erikseen tietyn genren ilmentymien luonti-, käsittely- ja säilytysmedia kohdassa 3.2.1 esiteltyjen vaiheiden mukaisesti. Näiden tietojen perusteella kohdeorganisaation genret voitiin luontevasti jaotella neljään, tiedon elinkaaren eri vaiheet kattavaan medialuokkaan, joita ovat paperimuotoiset genret, toimistosovelluksilla luodut genret, tallentamattomat genret sekä tietojärjestelmätuetut genre, jotka on tallennettu organisaation tietojärjestelmiin. Medioita ei luokiteltu Tyrväisen ja Päivärinnan (2003) menetelmäkuvauksessa esitetyn mediajaottelun mukaisesti, koska se olisi ollut käyttötarkoituksen huomioiden liian yksityiskohtainen. Seuraavassa tutkimuksessa käytettyjen medialuokkien kuvaukset sisältävässä luetelmassa kuitenkin mainitaan, mitä edellä mainitussa menetelmäkuvauksessa esitetyn mediajaottelun luokkia ne vastaavat. Tässä tutkimuksessa käytettiin seuraavia genrekohtaisia medialuokkia:

Paperimuotoiset genret: Nämä ovat nimensä mukaisesti genrejä, joiden käsittely kohdeorganisaation sisällä tapahtuu kokonaan paperimuodossa. Näistäkin genreistä

suurin osa luodaan tietokonetta käyttäen, mutta tulostetaan heti paperille käsittelyä varten. Jotkut luokan genreistä on puolestaan luotu kohdeorganisaation ulkopuolella tietokonetta käyttäen, mutta kohdeorganisaatioon ne tulevat paperimuodossa. Ryhmän genrejä käsitellään paperimuodossa joko arvokkuuden vuoksi tai jostain käytännön syystä. Ensimmäistä ryhmää edustava genre on esimerkiksi tutkintotodistus. Jälkimmäiseen ryhmään kuuluu esimerkiksi opiskelijan tenttitilanteessa käsin kirjoittama tenttivastaus, joita ei tämän hetkisen käytännön mukaisesti digitalisoida lainkaan. Luokka vastaa alkuperäisen medialuokituksen ”analoginen media” (analogue medium) –kategoriaa.

Toimistosovelluksilla tuotetut genret: Toimistosovelluksilla tuotetut genret luodaan jollakin toimistosovelluksella, jotka kohdeorganisaatioissa ovat pääosin tekstinkäsittely- tai taulukkolaskentasovelluksia. Säilytystä ja osittain myös käyttöä varten kyseisten genrejen ilmentymät tulostetaan lähes poikkeuksetta paperille. Sisällöllisesti luokka on melko kirjava, mukana on raportteja, päätöksiä ja tiedotteita. Tyypillisiä luokan genrejä ovat esimerkiksi opiskelijoiden valintapäätös, palkkaraportti ja tutkintovaatimukset. Luokka vastaa alkuperäisen medialuokituksen ”koodattu muoto” (encoded format) -kategoriaa sekä ”digitaalinen kuva” (digital image) –kategoriaa.

Tallentamattomat genret: Tallentamattomiin genreihin kuuluvat genret ovat operatiivisia ja pääosin suullisia, eikä niitä säilytetä virallisesti. Ryhmään kuuluvia genrejä ovat keskustelut, ryhmätapaamiset ja puhelut. Myös henkilökohtaiset sähköpostit lasketaan mukaan tähän luokkaan, koska niiden käyttötapa organisaatiossa on pääosin operatiivinen, eikä niitä säilytetä virallisesti. Tyypillisiä luokan genrejä ovat esimerkiksi tutkimuksen tietomassaltaan suurin genre eli luento-opetus, opintoneuvonta ja atk-ylläpitopyyntö. Selvää myös on, että luokassa on suuria ”hiljaisen” tiedon tietovarantoja, jotka eivät tule esille tässä selvityksessä. Tässä yhteydessä on myös syytä mainita, että esimerkiksi suullinen luento-opetus laskettiin PROMI-projektissa sovelletun käytännön mukaisesti yhdeksi genreksi ja kirjallinen luentomateriaali toiseksi. Näin toimittiin myös useiden muiden tallentamattomien genrejen kohdalla. Tallentamattomat genret laskettiin mukaan tuloksiin, koska osa niistä saattaa muuttua virallisesti säilytettäväksi tulevaisuudessa. Alkuperäisessä medialuokituksessa ryhmä

sisältää ihmisten väliseen suoraan viestintään liittyvät kategoriat (same time, same place) sekä kategorian ”välitettävä ja osittain säilyvä viestintä” (mediated and semi-transient communication).

Tietojärjestelmätuetut genret: Tähän kuuluvat genret, jotka säilytetään kokonaan tai osittain tiedekunnan erilaisissa tietojärjestelmissä. Tiedekunnan tietojärjestelmät ovat melko pieniä ja hajanaisia, näihin kuuluvat kurssi-ilmoittautumisjärjestelmä KORPPI⁴, tilavarausjärjestelmä sekä pieniä taloushallinnon tukena toimivia järjestelmiä. Tähän luokkaan on laskettu kuuluvaksi myös opintosuorituksiin liittyvät genret, vaikka niitä säilytetäänkin kohdeorganisaation ulkopuolisessa opintorekisteritietokannassa, jossa säilytetään myös opiskelijoiden henkilötietoja. Mielenkiintoisesti opintosuoritus onkin organisaation ainoa genre, jonka pitkäaikaissäilytys tapahtuu kokonaan sähköisessä muodossa. Luokka vastaa alkuperäisen medialuokituksen kategorioita ”rakenteinen muoto” (structured format), ”puoli-rakenteinen muoto” (semi-structured format) ja ”formaali data” (formal data).

Lisäksi haastatteluissa kerättyjen tietojen pohjalta laskettiin lisäksi kaksi muuttujaa, joita käytetään seuraavassa luvussa esiteltävien tulosten yhteydessä kuvaamaan eri säilytysaika- ja medialuokissa tapahtuvan kommunikaation määrää kohdeorganisaatiossa. Nämä kaksi muuttujaa esitellään seuraavassa luettelussa:

Uniikkisivut (US): Luku kertoo tietyn genren ainutkertaisten ilmentymien laajuuden sivuina organisaatiossa vuositasona. Luku lasketaan kertomalla keskenään tietyn genren uniikkien ilmentymien (UI) määrä sen yksittäisen ilmentymän keskimääräisellä koolla eli sivumäärällä (S). Uniikkisivujen yksikkönä on sivujen kappalemäärä, esimerkiksi montako sivua organisaatiossa muodostuu matkalaskuja vuoden aikana.

Kokonaissivut (KS): Luku kertoo genren uniikkisivujen ja niistä otettavien kopioiden yhteislaajuuden. Arvo saadaan kertomalla uniikkien ilmentymien (UI) määrä kopioiden määrällä (K / UI) ja kertomalla tulos kyseisen genren sivumäärällä (S). Yksinkertaistetusti tunnusluku kertoo yksittäisen genren ilmentymien

⁴ <http://korppi.it.jyu.fi>

kokonaislaajuuden sivuina organisaatiossa vuositasolla. Laskemalla yhteen organisaation kaikkien genrejen kokonaissivut saadaan selville organisaatiossa tapahtuvan kommunikaation kokonaislaajuus.

Seuraavassa taulukossa (TAULUKKO 2) esitellään neljä genreä kattava näyte tapaustutkimuksen tulostaulukosta kaikkien edellä esiteltyjen muuttujien keräämisen ja prosessoinnin jälkeen. Näitä tuloksia ja niiden pohjalta tehtyjä johtopäätöksiä esitellään tarkemmin luvuissa 5 ja 6. Ennen näiden esittelyä arvioidaan kuitenkin tapaustutkimuksessa kerätyn aineiston kattavuutta kahdella eri tavalla.

TAULUKKO 2. Näyte tapaustutkimuksen aineiston sisältävästä taulukosta edellä esitellyn prosessoinnin jälkeen.

Genre	UI	K / UI	S	US	KS	Säilytysaika	Medialuokka
Haku jatko-opiskelijaksi	40	3	4	160	480	5 vuotta	toimistosovellus
ABI-päivät (suullinen)	1	150	20	150	3000	0	tallentamaton
Tutkintotodistus	100	2	2	200	400	pysyvästi	paperimuotoinen
Tenttitulokset	200	50	1	400	10000	5 vuotta	toimistosovellus

4.2.4 Tulosten kattavuus

Täydelliset tiedot eli kaikki edellisessä kappaleessa mainitut tiedot kerättiin yhteensä 143 kohdeorganisaation genreistä, joita alkuvaiheessa oli yhteensä 310. Suurin osa tästä erotuksesta selittyy mainitulla prosessien välivaiheina toimineiden genrejen karsimisella pois aineistosta. Joitain genrejä ei pystytty tunnistamaan, osasta ei löytynyt kaikkia tietoja käytössä olleiden resurssien puitteissa, ja jotkut genret olivat päällekkäisiä. Tällaisten genrejen määrä oli kuitenkin hyvin vähäinen.

Tapaustutkimuksen kattavuuden arvioimiseksi verrattiin organisaation kopiopaperinkulutuksesta tapaustutkimuksen pohjalta tehtyä arvioita (luku A) haastattelujen perusteella kerättyihin, organisaation todellista paperinkulutusta kuvaaviin tunnuslukuihin (luvut B ja C). Haastattelun kohteena tässä yhteydessä olivat

kohdeorganisaation tulostuspaperin hankinnasta vastaava osastosihteeri sekä tulostuspalvelimien ylläpidosta vastaava insinööri kohdeorganisaation ATK-ylläpidosta.

Kohdeorganisaation paperille tulostettava kommunikaatio genreanalyysin perusteella (A): Kohdeorganisaatiossa vuositasolla tapahtuvan kommunikaation yhteismääräksi arvioitiin genreanalyysin perusteella 5,9 miljoonaa (5'877'044) sivua. Tästä vähennettiin kohdassa 4.2.3 mainittu tallentamaton, tietojärjestelmätuettu ja paperimuotoinen kommunikaatio, jota oli yhteensä 1,6 miljoonaa (1'565'554) sivua sekä sellainen kommunikaatio, jotka tuotetaan paperille kohdeorganisaation ulkopuolella, esimerkiksi yliopistopainossa. Kohdeorganisaation ulkopuolella paperille tuotettavaa kommunikaatiota arvioitiin olevan yhteensä noin 900'000 (879'000) sivua. Vähennyslaskun tarkoituksena on saada paperille tapahtuvan tulostuksen määrää kuvaava luku, joka on vertailukelpoinen kohdeorganisaation todellista tulostuspaperinkulutusta kuvaavien tunnuslukujen kanssa. Mainittujen vähennysten jälkeen tulokseksi saatiin 3,4 miljoonaa (3'432'490) sivua. Tämä siis kertoo organisaatiossa mahdollisesti paperille tulostettavan kommunikaation kokonaismäärän. Myöhemmin haastatteluissa arvioitiin, että noin 60 % tästä materiaalista tulostetaan todellisuudessaakin paperille, joten genreanalyysin perusteella kohdeorganisaation paperille tulostettava kommunikaatio on vuositasolla noin 2,1 miljoonaa sivua.

Opiskelijoiden toimesta tapahtuva tulostus (B): Todellinen tulostuspaperinkulutus on peräisin kohdeorganisaation tulostuspalvelimen tilastoista ja siihen kuuluu kaikki kohdeorganisaation tiloissa tapahtunut tulostaminen vuoden ajalta. Kokonaisuudessaan kohdeorganisaatiossa tulostettiin 1,3 miljoonaa (1'262'044) sivua ja näistä opiskelijoiden tulostamia oli yhteensä 700'000 (710'439) sivua.

Kohdeorganisaation paperille tulostettava kommunikaatio arvioituna todellisen paperinkulutuksen perusteella (C): Kohdeorganisaation vuoden 2002 todellinen paperinkulutus on puolen vuoden käytön perusteella arvioituna 1,7 miljoonaa (1'668'000) arkkia. Kaksipuolisten tulosteiden määräksi arvioitiin haastatteluissa 60 %, joten todellinen tulostusmäärä sivuina on 2,7 miljoonaa (2'668'800) sivua. Vähentämällä tästä luvusta opiskelijoiden tulostus (luku B) lopputulokseksi saadaan noin 2 miljoonaa sivua.

Sekä luku A että luku C olivat hyvin lähellä kahta miljoonaa sivua. Tämä viittaisi siihen, että genreilmentymien määrä on tapaustutkimuksessa jossain määrin yliarvioitu, koska luvun A pitäisi olla lukua C pienempi johtuen väliprosesseihin kuuluvien genrejen puuttumisesta. Syynä voi tosin myös olla todellisen tulostuksen tai kaksipuolisten tulosteiden määrän virhearviointi. Toisaalta erotus on niin vähäinen, ettei genreanalyysissä todennäköisesti ole suuria systemaattisia virheitä. Luultavammin erotus selittyy ongelmilla, jotka johtuvat siitä, ettei alkuperäistä genrejaottelua oltu laadittu tiedon pitkäaikaissäilytyksen analysoimista varten. Asian tarkempi selvittäminen ei kuitenkaan liene tarkoituksenmukaista tämän työn puitteissa. Kattavuusarvion perusteella voidaan silti sanoa, että organisaation mediatyyppien määrän arviointi on yllättävän monimutkaista ja jotkut muuttujista on jouduttu arvioimaan suhteellisen karkeasti. Samoin voidaan todeta, että alkuperäistä genrejaottelua rakennettaessa on otettava huomioon analyysin käyttötarkoitus, jotta menetelmästä saataisiin paras mahdollinen hyöty.

Tapaustutkimuksen tulosten realistisuutta voidaan myös arvioida jakamalla organisaation vuosittainen kokonaiskommunikaatio (5,8 miljoonaa sivua) organisaation työntekijöiden (200) ja käytettävissä olevien vuosittaisten työpäivien (220) määrällä. Näin laskettuna tulokseksi saadaan noin 130 sivua henkilöä kohden yhden työpäivän aikana. On kuitenkin huomattava, että arvioista puuttuvat välivaiheissa tarvittavat genret, joten tapaustutkimuksen perusteella todellinen määrä on korkeampi. Landauerin (1986, 480) arvion perusteella ihmisen luetun tekstin omaksumiskyky on korkeintaan 300 sivua työpäivässä. Toisessa genrekartoituksessa kartoituksessa vastaavaksi luvuksi saatiin puolestaan 90 sivua (Tyrväinen, 2003). Näiden vertailuarvojen perusteella tämän tutkimuksen arviota 130 sivusta voidaan pitää uskottavana.

5. TAPAUSTUTKIMUKSEN TULOKSET

5.1 Tapaustutkimuksen tulokset

Luvussa esitellään edellisessä luvussa kuvatun aineiston pohjalta laskettuja kohdeorganisaatiossa käsiteltävän aineiston määrää kuvaavia tuloksia. Aluksi luodaan yleiskuva kohdeorganisaation tietovarantoihin. Tämän jälkeen esitellään uniikkisivujen määrällistä jakautumista kohdassa 4.2.3 esiteltyihin media- ja säilytysaikalukkiin. Lopuksi tarkastellaan tiedon rakeisuutta kyseisissä luokissa ja esitellään yksityiskohtaisemmin aineistoa, jota säilytetään kohdeorganisaatiossa pitkäaikaisesti. Tässä yhteydessä on syytä todeta etteivät tässä esitetyt luvut ole pyöristämättömydestään huolimatta tarkkoja arvoja, vaan laskennallisia arvioita.

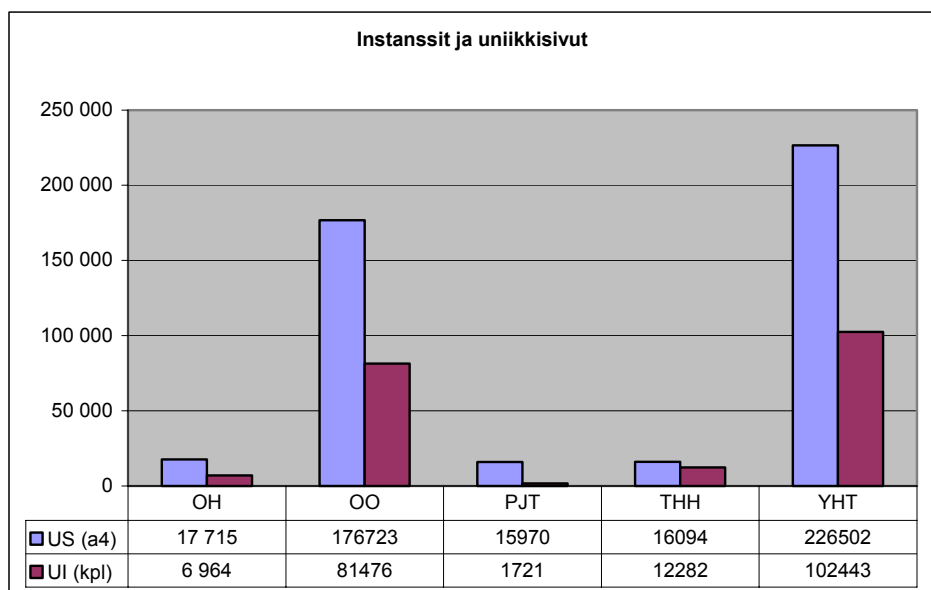
5.1.1 Yleiskuva kohdeorganisaation tietovarannosta

Kohdeorganisaatio jaettiin PROMI-projektissa viiteen eri osa-alueeseen, joita olivat opintohallinto, opetus ja opiskelu, projektit, jatkotutkimus sekä talous ja henkilöstöhallinto. Osa-alueet ovat samat myös tässä tutkimuksessa sillä erolla, että pienehköt projektit -alue ja jatkotutkimus -alue yhdistettiin yhdeksi osa-alueeksi tulosten käsittelyn yksinkertaistamiseksi. Oheisessa taulukossa (TAULUKKO 3) esitellään kohdassa 4.2.3 mainittuja tunnuslukuja jaoteltuna organisaation osa-alueiden mukaisesti.

TAULUKKO 3. Uniikkisivut (US), kokonaissivut (KS), uniikit instanssit (UI) ja genret (G) jaoteltuna organisaation toimintojen mukaisesti. Toimintoja ovat opintohallinto (OH), opetus ja opiskelu (OO), projektit ja jatkotutkimus (PJT) sekä talous- ja henkilöstöhallinto (THH).

	OH	OO	PJT	THH	YHT
G (kpl)	45	41	26	31	143
UI (kpl)	6 964	81 476	1 721	12 282	102 443
UI (%)	6,9 %	79,5 %	1,7 %	12,0 %	100 %
US (kpl)	17 715	176 723	15 979	16 094	226 502
US (%)	7,8 %	78,0 %	7,1 %	7,1 %	100,0 %
KS (a4)	152 115	5 683 300	14 848	26 741	5 877 004
KS (%)	2,5 %	95,1 %	1,9 %	0,4 %	100,0 %

Taulukossa silmiinpistävä on opetus ja opiskelu -luokan hallitseva asema organisaatiossa. Luokkaan kuuluu 78 % uniikkisivuista ja 95 % kokonaissivuista. Tämä näkyy myös oheisessa kuvassa (KUVIO 8), jossa esitetään uniikkisivujen ja instanssien jakautuminen toiminnoittain.



KUVIO 8. Uniikkisivujen ja instanssien jakautuminen toiminnoittain kohdeorganisaatiossa.

Tämä opetus ja opiskelu –luokan hallitseva asema selittyy lähes kokonaan kohdeorganisaation kolmen suurimman genren sijoittumisella kyseiseen. Nämä kolme opetukseen liittyvää genreä esitellään tarkemmin seuraavassa taulukossa (TAULUKKO 4).

TAULUKKO 4. Kohdeorganisaation kolme suurinta genreä.

Genre	UI	K / UI	S	US	KS	Säilytysaika	Medialuokka
Luentomateriaali	200	100	200	40'000	4'000'000	3 vuotta	Toimistosovellus
Luento-opetus	1000	50	30	30'000	1'500'000	0	Tallentamaton
Opiskelijoiden Tenttivastaukset	16'000	1	4	64'000	64'000	0,5 vuotta	Paperimuotoinen

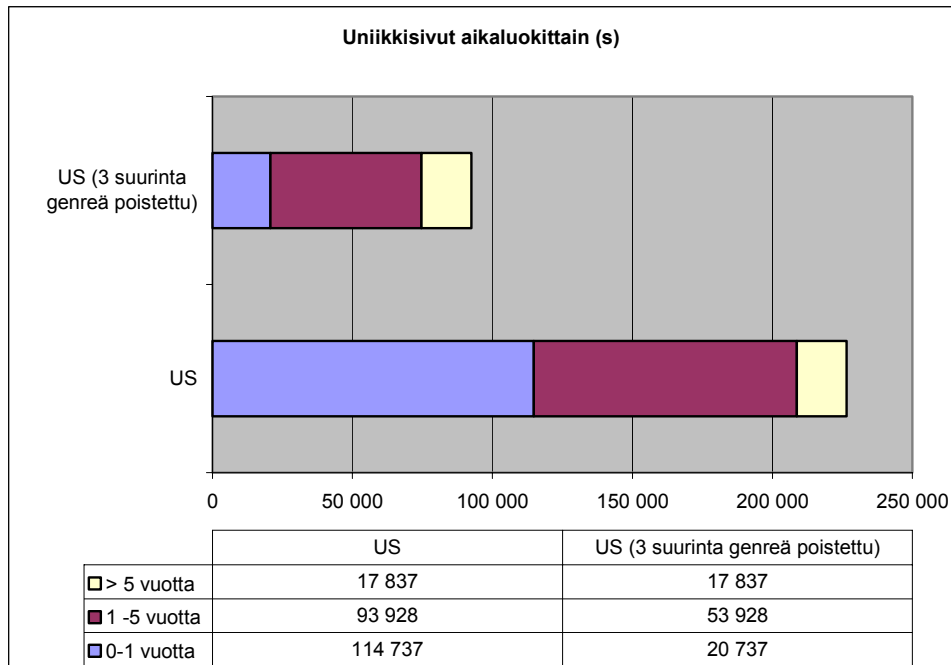
Taulukon perusteella kohdeorganisaation suurimmat genret ovat suullinen luento-opetus (30'000 uniikkisivua ja 1'500'000 kokonaissivua), opiskelijoiden tenttivastaukset (64'000 uniikkisivua) sekä opetusmateriaali (40'000 uniikkisivua ja 4'000'000 kokonaissivua). Näiden genrejen asema on ratkaiseva tarkasteltaessa myös medioita ja säilytysaikoja. Tästä voidaan päätellä, että hyvin suuri osa kohdeorganisaatiossa tapahtuvasta kommunikaatiosta liittyy opetukseen.

5.1.2 Tiedon tallennusmediat ja säilytysajat

Säilytysaikojen perusteella genret jaoteltiin kolmeen eri kategoriaan. Genre kuuluu operatiiviseen luokkaan, mikäli sen säilytysaika on vuosi tai vähemmän. Keskipitkällä ajalla viitataan 1 - 5 vuoden säilytysaikaan. Pitkäaikaisella tarkoitetaan tässä yhteydessä yli 5 vuoden pituista ajanjaksoa.

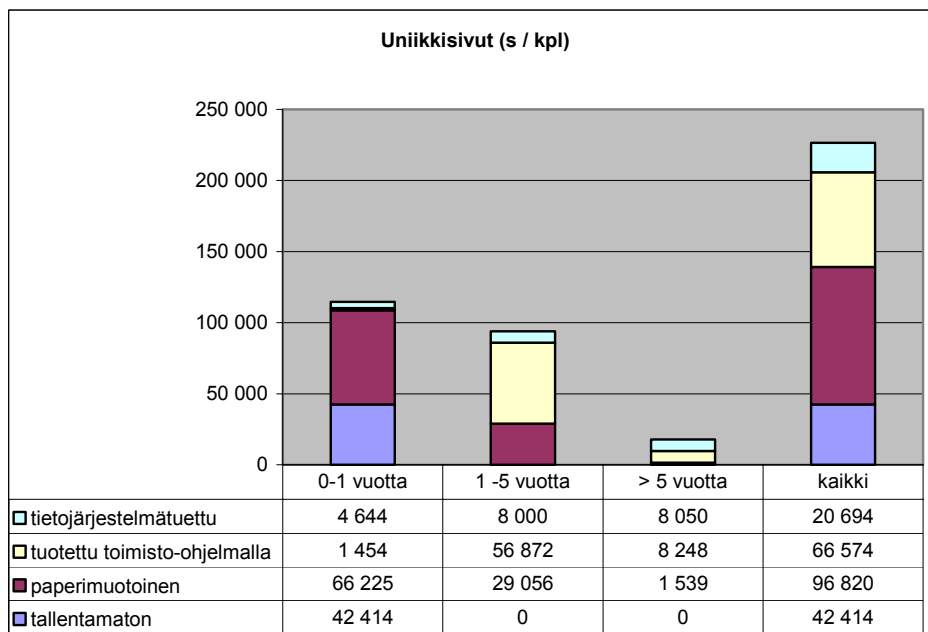
Seuraavassa kuvassa (KUVIO 9) nähdään selvityksessä mukana olleiden genrejen jakautuminen eri säilytysaikaluokkiin. Alempi kuvaaja kuvaa tilannetta silloin kun mukana ovat kaikki genret. Tällöin noin 50 % tutkimuksessa mukana olevasta organisaation tiedosta säilytetään lyhytaikaisesti alle vuoden, 41 % 1 - 5 vuotta ja 8 %

yli 5 vuotta. Ylemmässä kuvaajassa nähdään puolestaan tilanne, mikäli organisaation 3 suurinta genreä lasketaan pois kokonaissivumäärästä. Uniikkisivujen kokonaismäärä putoaa 226'502 sivusta 92'502 sivuun. Näistä sivuista 22 % säilytetään lyhytaikaisesti, 58 % 1-5 vuotta ja 19 % enemmän kuin 5 vuotta.



KUVIO 9. Uniikkisivujen määrä organisaatiossa aikaluokkiin jaettuna.

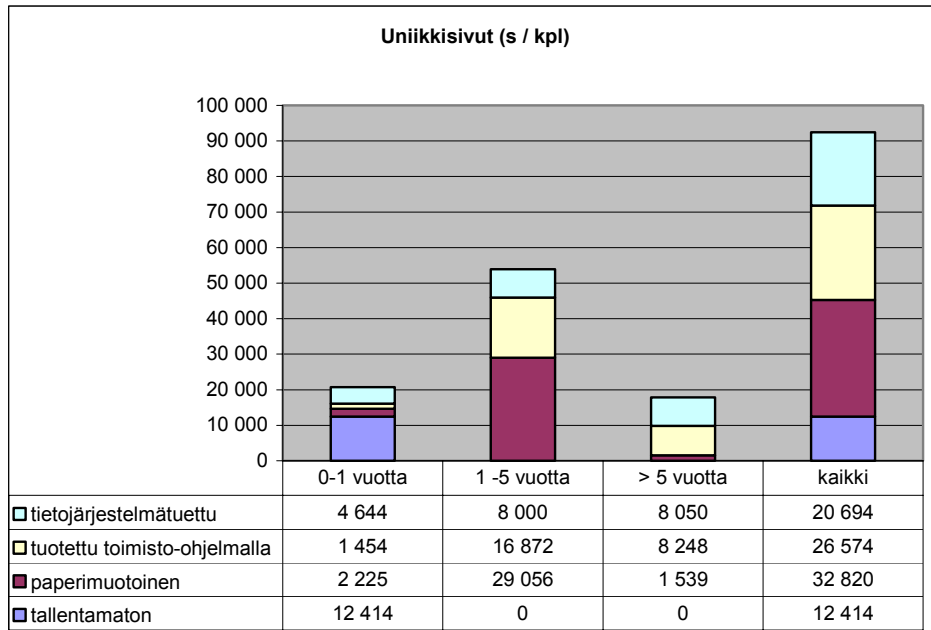
Kartoitettujen genrejen luonti-, käyttö ja tallennusmedian pohjalta kohdeorganisaation genret voitiin jakaa neljään medialuokkaan, jotka esiteltiin kohdassa 4.2.3 Seuraavassa kuvassa (KUVIO 10) nähdään uniikkisivujen määrä näissä media- ja säilytysaikalukissa.



KUVIO 10. Uniikkisivujen määrä jaoteltuna kategorioihin mediatyyppien ja säilytysaikojen perusteella.

Tallentamattomia genrejä löytyi luonnollisesti ainoastaan operatiivisesta luokasta. Mielenkiintoista tuloksissa on paperin osuuden väheneminen ja digitaalisen tiedon suhteellisen osuuden kasvu säilytysajan pidentyessä. Tämä selittyy osittain kolmen suurimman organisaatiokohtaisen genren vaikutuksella ja osittain opintosuoritus - genren tietokantapohjaisella pitkäaikaissäilytyksellä.

Kolmen suurimman genren merkittävän vaikutuksen vuoksi tuloksia on syytä tarkastella myös ilman näiden genren vaikutusta (KUVIO 11). Tällöin suurimmat muutokset tapahtuvat operatiivisessa luokassa, jossa paperin osuus tenttivastausten poislaskemisen seurauksena vähenee huomattavasti. Toisaalta myös keskipitkän ajan säilytettävässä tiedossa toimistosovelluksilla tuotetun tiedon osuus laskee merkittävästi johtuen luentomateriaali-genren poistumisella. Kokonaisuutena kolmen suurimman genren poistaminen pudottaa kokonaissivumäärä yli puolella.



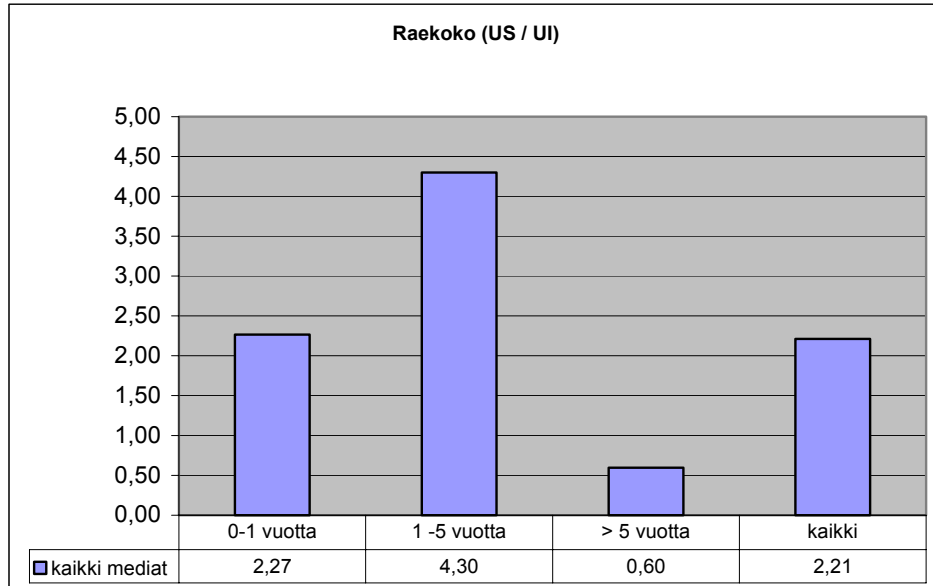
KUVIO 11. Uniikkisivut jaoteltuna kategorioihin medioiden ja säilytysaikojen perusteella. Kuvasta on poistettu organisaation kolme suurinta genreä.

5.1.3 Tiedon raekoko

Raekoolla tarkoitetaan digitaalisen tiedon yhteydessä hallittavien tietoyksikköjen kokoa. Tässä tutkimuksessa sillä puolestaan viitataan tiettyyn luokkaan kuuluvien genreilmentymien keskimääräiseen kokoon sivuina ja se saadaan jakamalla kaikkien kiinnostuksen kohteena olevaan luokkaan kuuluvien uniikkisivujen (US) summa uniikkien instanssien (UI) summalla. Esimerkiksi yli 5 vuotta säilytettävän tiedon keskimääräinen raekoko saadaan jakamalla yli 5 vuotta säilytettävien genrejen uniikkisivujen kokonaislaajuus kyseisten genrejen uniikkien instanssien kokonaisuudella. Raekoko siis kertoo tietyn genren ilmentymän keskimääräisen koon sivuina (S) tietyssä luokassa.

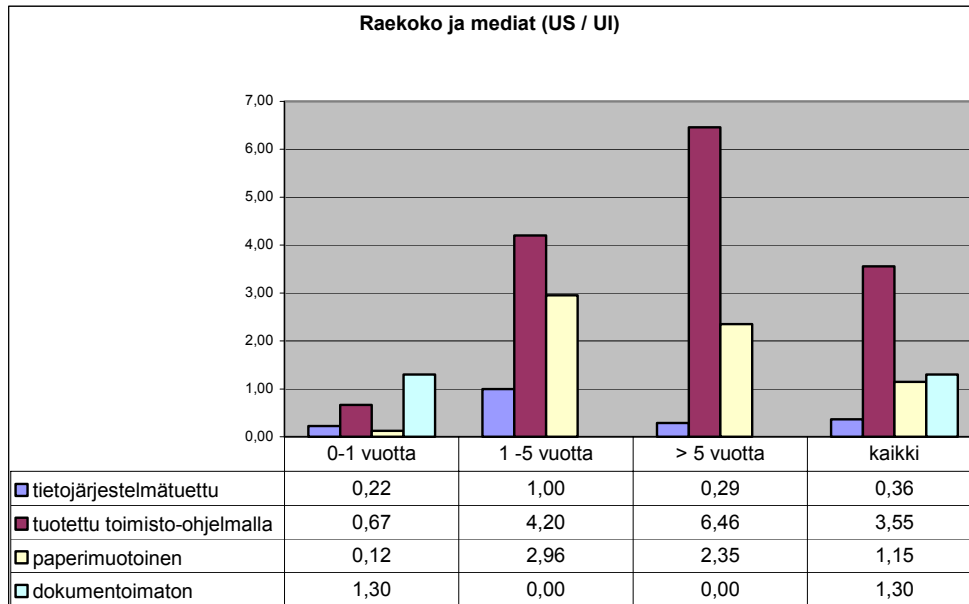
Kuvassa (KUVIO 12) esitellään uniikkisivujen ja uniikkien ilmentymien välinen suhde eri säilytysaikaluokissa. Mukana on myös kaikkien mukana olleiden genrejen raekoko, joka on 2,2 sivua. Operatiivisessa säilytysaikaluokassa raekoko oli puolestaan 2,3 sivua, keskipitkällä ajalla 4,3 sivua ja yli 5 vuotta säilytettävissä genreissä 0,6 sivua.

Viimeisen luokan alhainen arvo johtuu mukana olevista tietokantamuotoisista genreistä, joiden ilmentymiä mahtuu yhdelle sivulle useita.



KUVIO 12. Tiedon raekoko eri säilytysaikaluokissa.

Raekokoa on syytä tarkastella myös hieman tarkemmalla tasolla ja kuvassa (KUVIO 13) arvioidaan raekokoa eri media- ja säilytysaikaluokissa. Kuvasta on havaittavissa tietojärjestelmätuettujen genrejen melko alhainen (0,36) raekoko, joka kertoo organisaation tietojärjestelmissä säilytettävän tiedon olevan pitkälti tietokantamuotoista. Toimistosovelluksessa tuotettujen genrejen raekoko on puolestaan melko suuri, keskimäärin 3,55. Tallentamattomien ja paperimuotoisten genrejen raekoko pysyy melko lähellä yhtä sivua.



KUVIO 13. Tiedon raekoko eri media- ja säilytysaikalukissa.

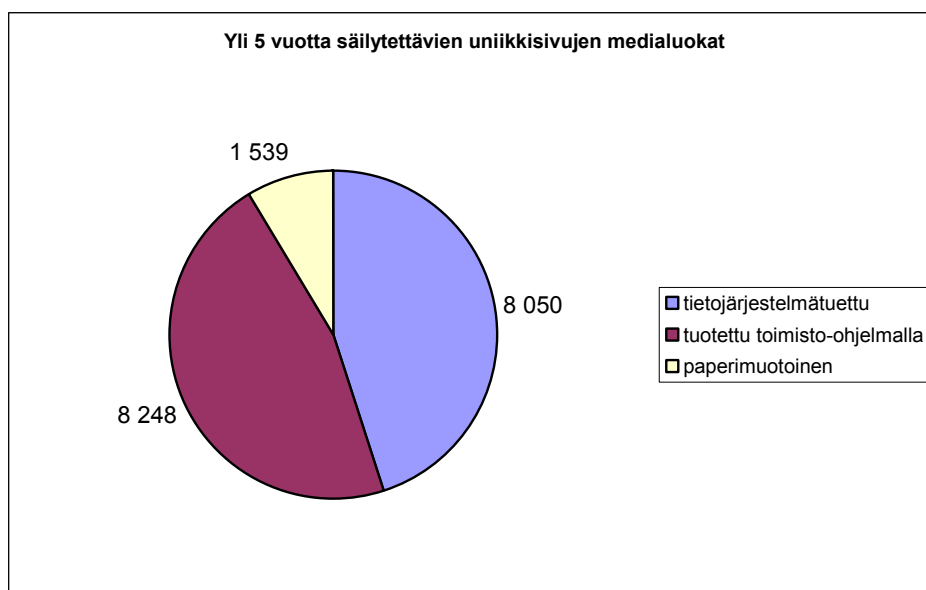
Etenkin viimeksi esitellyn kuvion (KUVIO 13) kohdalla on huomattava, että tällä tarkkuusasteella raekokoja kannattaa pitää aineiston keräämisessä käytettävien menetelmien epätarkkuudesta johtuen ainoastaan suuntaa-antavina. Niiden tarkkuus riippuu myös pitkälti PROMI-projektissa luodun genrejaottelun tarkkuudesta, johon ei ole voitu vaikuttaa tätä selvitystä tehdessä.

5.1.4 Organisaation pitkäaikaissäilytettävä tieto

Tapaustutkimuksessa analysoidusta organisaation uniikkisivuista 8 % (17837 sivua) säilytetään yli 5 vuotta. Seuraavassa tarkastellaan tätä pitkäaikaissäilytettävää tietomassaa lähemmin.

Pitkäaikaissäilytettävän tiedon jakautuminen eri medialuokkien kesken esitetään oheisessa kuvassa (KUVIO 14). Kuvaa tarkasteltaessa selviää, että 9 % (1539 sivua) tästä tiedosta on koko elinkaarensa paperimuodossa, 50 % (8248 sivua) tuotetaan toimitusovelluksilla ja tulostetaan paperille pitkäaikaissäilytystä varten ja 45 % (8050 sivua) säilytetään tietojärjestelmissä tai -kannoissa. Yllättävästi viimeisen ryhmän suuri osuus selittyy lähes kokonaan yhden tietokannan eli opintosuoritus- ja

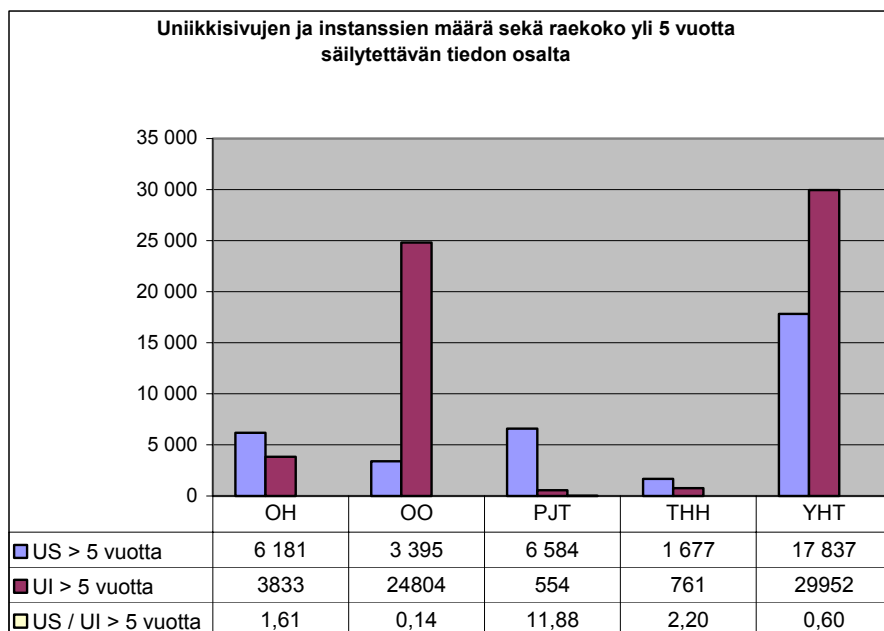
henkilötietorekisterin olemassaololla. Opintosuoritus on tietomäärältään suurin yksittäinen pitkäaikaissäilytettävä genre. Lisäksi tietojärjestelmätuetujen genrejen joukossa on useita taloushallintoon ja tutkimukseen liittyviä genrejä, joista on olemassa useita rinnakkaisia tallennusmuotoja, mutta ainakin yksi arkistointikelpoinen paperiversio. Opintosuorituksia lukuun ottamatta yli 5 vuotta säilytettävä tieto tulostetaan kohdeorganisaatiossa tapaustutkimuksen perusteella paperille säilytystä varten.



KUVIO 14. Pitkäaikaissäilytettävien uniikkisivujen jakaantuminen eri medialuokkiin

Seuraavassa kuvassa (KUVIO 15) esitetään yli 5 vuotta säilytettävän tiedon määrä organisaation eri osa-alueilla uniikkisivuina ja instansseina sekä näiden arvojen välinen suhde eli raekoko. Suurin osa tähän ryhmään kuuluvasta tiedosta liittyy projekteihin ja jatkotutkimukseen sekä opintohallintoon. Näihin ryhmiin sijoittuvatkin suurimmat pitkäaikaissäilytettävät genret kuten opintosuoritus ja ensimmäiseksi mainittuun ryhmään kuuluva tieteellinen julkaisu. Opetus ja opiskelu -luokka sisälsi puolestaan toiseksi vähiten (3395 uniikkisivua) pitkäaikaissäilytettävää materiaalia, vaikka kyseinen luokka olikin hyvin merkittävä kokonaisuutta tarkasteltaessa. Vähiten yli 5 vuotta säilytettävää aineistoa oli talous- ja henkilöstöhallinta -luokassa.

Opetus ja opiskelu –toimintojen alhainen raekoko (0,14) selittyy opintosuoritusgenrellä, jonka ilmentymiä säilytetään pieninä kokonaisuuksina tietokannassa. Projektit ja jatkotutkimus –ryhmän korkea raekoko (11,88) johtuu puolestaan tieteellinen julkaisugenren ilmentymien suuresta koosta. Kahden muun ryhmän raekoot pysyvät melko lähellä yhtä sivua.



KUVIO 15. Uniikkisivujen ja instanssien määrä sekä raekoko yli 5 vuotta säilytettävän aineiston osalta. Tiedot on jaoteltu organisaation toimintojen mukaan.

6. TAPAUSTUTKIMUKSEN JOHTOPÄÄTÖKSET

Luvussa esitellään tapaustutkimusosuuden keskeiset johtopäätökset. Aluksi tarkastellaan tutkimuksessa käytettyä menetelmää ja tämän jälkeen varsinaisia tuloksia.

6.1 Keskustelua tapaustutkimuksessa käytetystä menetelmästä

Genre-pohjainen tietovarantojen analysointimenetelmä osoittautui toimivaksi tavaksi arvioida organisaation tietovarantoja pitkäaikaissäilytyksen näkökulmasta, kunhan käytettävän tutkimusaineiston luotettavuudesta voidaan varmistua. Sen avulla päästiin hyvin yksityiskohtaiselle tarkkuustasolle ja raekoon käsite tarjosi toimivan mittarin tallennusratkaisujen teknisten yksityiskohtien arvioimiseen, koska se esimerkiksi kertoi, milloin säilytettävä tieto on tietokantamuotoista ja milloin dokumenttimuotoista. Myös säilytysajan liittäminen mukaan menetelmään kävi ongelmitta.

Menetelmässä ja sen soveltamisessa oli kuitenkin puutteita, joista merkittävimmät liittyivät juuri tiedonkeräysprosessiin ja siinä käytettyyn tausta-aineistoon. Genreperhjaista menetelmää käytettäessä tulosten tarkkuus ja luotettavuus ovat luonnollisesti suoraan riippuvaisia tavasta, jolla genret on tunnistettu, nimetty ja jaoteltu. Tämän tutkimuksen kohdalla käytössä oleva genrejaottelu oli peräisin kohdeorganisaation prosessien kartoittamiseen liittyvästä selvityksestä. Materiaali oli siis kerätty toista käyttötarkoitusta varten ja sen muuttaminen pitkäaikaissäilytyksen tutkimukseen soveltuvaksi oli tietyissä vaiheissa hankalaa ja vaikutti heikentävästi tulosten tarkkuuteen. Esimerkiksi organisatoristen prosessien ja niiden lopputuloksena muodostuneen aineiston erottaminen toisistaan oli genrejen nimeämiskäytännön vuoksi ongelmallista. Osittain tämän johdosta myös tulosten kattavuuden arviointi oli hankalaa.

Toinen, pienempi ongelma liittyi tapaustutkimuksessa käytetyn mediajaotteluun. Monet kohdeorganisaation genreistä tallennetaan elinkaarensa aikana rinnakkain usealle eri tallennusmedialle, eivätkä medialuokkien rajat ole aina kovinkaan yksiselitteisiä. Tämä kuitenkin voitiin huomioida tutkimuksessa käytetystä melko karkeasta mediajaottelusta

johtuen ainoastaan rajallisesti. Toisaalta tarkemman mediajaottelun soveltaminen olisi tutkimuksen luonteen huomioiden ollut tarpeettoman työlästä.

Edellä todetut ongelmat antoivat myös viitteitä kahdesta genrepohjaiseen tietoresurssien analyysimenetelmään liittyvästä heikkoudesta. Ensimmäkin menetelmän käyttötarkoitus on syytä olla tiedossa ennen genrejen tunnistamista ja nimeämistä, jotta tuloksista saataisiin mahdollisimman luotettavia. Toiseksi lähes kokonaista organisaatiota koskevan genrepohjaisen aineiston kattavuuden ja luotettavuuden arviointi voi olla ongelmallista, koska saaduille tuloksille on hankala löytää vertailukohtaa kohdeorganisaatiosta. Tässä tutkimuksessa kattavuutta arvioitiin vertaamalla genrekartoituksen pohjalta arvioitua organisaation paperinkulutusta sen todelliseen paperinkulutukseen. Kovin luotettava mittari tämäkään ei tosin ollut, koska organisaation paperinkulutuksen laskeminen tarkasti ostetun paperimäärän ja tulostustilastojen perusteella on yllättävän monimutkainen tehtävä. Tämän vuoksi tulosten realistisuutta arvioitiin laskemalla myös henkilöä kohti päivässä syntyvän kommunikaation määrä.

Näiden vertailulukujen laskemisen jälkeen tutkimusaineiston luotettavuudesta voitiin kuitenkin varmistua siinä määrin, jotta sen analysointi voitiin aloittaa. Tulosten analysoinnissa genrepohjainen menetelmä osoittautuikin varsin toimivaksi. Sen avulla organisaation tietoresurssit voitiin jakaa vaivattomasti erilaisiin luokkiin ja tunnuslukujen laskeminen oli varsin suoraviivaista. Genret siis tarjoavat tämän tutkimuksen valossa toimivan välineen organisaation informaatioresurssien ja niiden pitkäaikaissäilytyksen analysointiin, kunhan tutkimusaineiston keräämiseen kiinnitetään riittävästi huomiota. Genreihin pohjautuvan menetelmän käyttötarkoitus on myös syytä olla tiedossa jo genrejen nimeämisvaiheessa, jotta tulosten luotettavuudesta voitaisiin varmistua paremmin.

Kokonaisuutena on huomattava, ettei edellä esitetyn perusteella voi tehdä kovin pitkälle meneviä päätelmiä itse menetelmästä. Osa ongelmista selittyy varmasti myös tutkijan kokemattomuudella ja toisaalta genrepohjaista organisaation tietoresurssien analysointimenetelmää ei ole käytetty laajassa mittakaavassa kovinkaan paljon. Lisäksi

on huomioitava, ettei kyseistä menetelmää voitu verrata muihin mahdollisiin menetelmiin tämän tutkimuksen puitteissa. Menetelmän suhteen tapaustutkimuksen keskeisenä opetuksena voidaankin pitää, että nykyaikaisessa organisaatiossa tapahtuvan viestinnän määrälliseen arviointiin tarkoitettujen menetelmän laatiminen ja testaaminen on vaativa tehtävä.

6.2 Keskustelua tapaustutkimuksen tuloksista

Tapaustutkimuksen varsinaiset tulokset olivat pääosin ennakko-oletusten mukaisia. Organisaation kommunikaatio jakaantui eri säilytysaikaluokkiin melko odotetusti eli materiaalin määrä väheni säilytysajan kasvaessa. Medialuokkia tarkasteltaessa puolestaan selviää, että merkittävä osa organisaation kommunikaatiosta ainakin luodaan digitaalisesti, mutta muutetaan analogiseksi käsittelyä varten. Yllättävää kuitenkin oli, että suurin osa organisaation tiedosta sisältyy hyvin pieneen joukkoon genrejä. Sama asia havaittiin myös toiseen organisaatioon kohdistuneessa laajemmassa genrekartoituksessa (Tyrväinen, 2003). Näiden kahden tutkimuksen perusteella siis vaikuttaisi, että suurin osa organisaation tiedosta sijoittuu muutamaankin organisaatiokohtaiseen genreen. Ennakko-odotusten perusteella oli myös yllättävää, ettei digitaalisessa muodossa pitkäaikaissäilytettävää materiaalia ollut opintosuoritusten lisäksi oikeastaan lainkaan.

Lisäksi oli jossain määrin yllättävää havaita, että valtaosa kohdeorganisaation kokonaistietovarannoista (78 % uniikkisivuista ja 95 % kokonaissivuista), kuten myös neljä suurinta genreä liittyivät opetukseen ja opetusmateriaaleihin. Osittain tämä oli odotettavissa, sillä onhan kohdeorganisaatio ennen kaikkea opetusorganisaatio, mutta silti tämän tyyppisen aineiston osuus vaikuttaa yllättävän suurelta. Tämä viittaa siihen, että sähköisten oppimateriaalien käytölle kohdeorganisaatiossa asetetut odotukset ovat perusteltuja. Asia raportoitiinkin tutkimusaineiston pohjalta laaditussa julkaisussa, jossa asiaa tarkasteltiin sähköisiin oppimateriaaleihin liittyvien mahdollisuuksien näkökulmasta (Tyrväinen, Järvenpää ja Sievänen, 2003). Samalla se osoittaa, että digitaalisen aineiston osuus tulee kasvamaan merkittävästi tulevaisuudessa, mikäli kohdeorganisaatio siirtyy toteuttamaan sähköisiin opetusmateriaaleihin liittyviä

suunnitelmia. Tällöin myös digitaalisen tiedon säilyvyyden varmistamiseen on kiinnitettävä huomattavasti nykyistä enemmän huomiota.

Kirjallisuudessa (Dollar, 1992; Morris 1998; Koupoulos & Frappolo, 1995; Simpson, 1998) esitetyt organisaatioiden tiedon pitkäaikaissäilytyksen ja informaatioresurssien hallintaan liittyviä ongelmia koskevat väitteet pitävät paikkaansa kohdeorganisaatiossa. Kohdeorganisaatiossa on nähtävissä merkkejä informaatioaaristosta, käytössä on useita keskenään yhteensopimattomia järjestelmiä, jotka ovat osittain päällekkäisiä. Suurin osa tiedosta tuotetaan verkotettujen henkilökohtaisten tietokoneiden avulla. Tiedekunnassa ei ole myöskään keskitettyä informaatioresurssien hallintaa, vaan tämä vaikuttaa pohjautuvan enemmänkin tiedekunnassa työskentelevien henkilöiden kokemukseen ja hiljaiseen tietoon. Varsinainen tiedon pitkäaikaissäilytys tapahtuu keskitetyssä arkistossa, jossa keskeisin tallennusmedia on paperi.

Informaatioaaristo ei kuitenkaan ole muodostunut itsestään, vaan syyt ovat selvästi nähtävissä kohdeorganisaatiossa ja nekin vastaavat pitkälti kirjallisuudessa esitettyjä. Tiedekunta on muodostunut äskettäin kahdesta erilaisesta yliopiston osasta, joiden tavoitteissa ja toimintatavoissa on luonnollisesti eroja. Kohdeorganisaatio on myös laajentunut ja monipuolistunut erittäin nopeasti, esimerkiksi opiskelijoiden määrä on kasvanut viimeisten vuosien aikana erittäin nopeasti. Pääsyy tähän on informaatioteknologian merkityksen nopea kasvu ja alan jatkuva kehittyminen viimeisen viiden vuoden aikana.

Toistaiseksi vallitseva järjestelmä on kuitenkin pääosin toiminut, koska kohdeorganisaatio on kooltaan rajallinen ja hallinnollisesti melko yksinkertainen. Pitkäaikaissäilytettävää materiaalia ei myöskään ole paljon ja se on pitkälti tekstimuotoista, joten se voidaan helposti muuntaa analogiseen muotoon arkistointia varten. Ongelmia voi kuitenkin muodostua, mikäli tiedekunta kasvaa vielä kokoa ja/tai digitaalisen tiedon rooli organisaatiossa kasvaa, koska tällöin informaatioresurssit eivät enää ole yksittäisten työntekijöiden hallittavissa. Asialla on merkitystä myös resurssien tehokkaan hyödyntämisen kannalta, sillä nyt tietovarannot ovat osittain päällekkäisiä ja tietoresurssien hallitsemiseksi joudutaan tekemään runsaasti ylimääräistä työtä.

Haastatteluissa myös ilmeni, ettei haluttuihin tietoresursseihin päästy aina käsiksi toivotulla tavalla niiden hallintaan liittyvien ongelmien seurauksena.

Arkistoiva aineisto on kuitenkin pääosin siirtynyt aktiivisen vaiheensa ulkopuolelle, sille ei siis ole enää käyttöä kohdeorganisaation jokapäiväisessä toiminnassa. Tällöin digitaalisista tallennusmuodoista saatavien lisäominaisuuksien menetystä ei myöskään koeta organisaatiossa ongelmallisena. Toisaalta on mahdollista, ettei esimerkiksi hakujärjestelmien puutteen vuoksi vaikeasti saavutettavia tietoresursseja osata edes kaivata. Mielenkiintoinen kysymys myös on, toteutuvatko organisaation omat säilytysaikaosuudet sähköisen aineiston osalta. Esimerkiksi kurseilla käytettävän opetusmateriaalin tulisi nykyisen käytännön mukaisesti säilyä kolme vuotta. Tapahtuuko sitten näin esimerkiksi verkkosivuille tai erittäin nopeasti kehittyviin sähköisiin oppimisympäristöihin laitetun materiaalin osalta onkin moniselitteisempi asia. Tähän ei kuitenkaan ollut mahdollista ottaa kantaa tässä tutkimuksessa, eikä opetusmateriaali kuulu viralliseen arkistointisuunnitelmaan.

Varsin aiheellista on myös pohtia, onko analogisia tallennusmuotoa hyödyntävä keskitetty arkisto paras tapa pitkäaikaissäilytyksen toteuttamiseksi tulevaisuudessa. Digitaalisten aineistojen määrän kasvaessa ja niiden hallinnan vaikeutuessa käyttöön joudutaan mahdollisesti ottamaan esimerkiksi dokumenttien hallintajärjestelmä tai soveltamaan jotain metatietostandardia. Myös käytettyjen tallennusmuotojen määrä tulee todennäköisesti kasvamaan digitaalisten aineistojen monipuolistuessa. Näiden asioiden suhteen joudutaan luonnollisesti tekemään valintoja, jotka vaikuttavat merkittävästi aineiston säilyvyyteen ja käyttökelpoisuuteen tulevaisuudessa. Tällöin tiedon pitkäaikaissäilytyksestä vastaavien tahojen tulisi ottaa kantaa näihin kysymyksiin. Arkistointifunktion täydellinen hajauttaminen ei luultavasti ole yliopiston kaltaisessa monitahoisessa organisaatiossa ainakaan lähitulevaisuuden vaihtoehto, mutta tiedon pitkäaikaissäilytyksestä vastaavan tahon tulisi päästä vaikuttamaan nykyistä enemmän kohdeorganisaation jokapäiväiseen toimintaan ja informaatioressurssien hallintaa koskevaan päätöksentekoon.

Yhteenvedon tapaustutkimuksesta voidaan sanoa, että tiedon pitkäaikaissäilytys toimii kohdeorganisaatiossa perinteisesti keskitetyssä arkistossa, jossa käytetään analogisia tallennusmuotoja, vaikka varsinainen organisatorinen kommunikaatio tapahtuu pitkälti digitaalisesti. Tämä ei kuitenkaan uhkaa tiedon säilyvyyttä, sillä arkistointia säädellään organisaatiossa tarkasti. Tällä hetkellä säilytyksen kohteena olevan aineiston pitkäaikaissäilytyksen toteutuminen ei siis ole digitalisoitumisen seurauksena uhattuna kohdeorganisaatiossa, eikä kirjallisuuskatsauksessa esitetyille menetelmille kuten migraatioille ja emuloinnille ole ainakaan vielä tarvetta. Tutkimuksessa analysoidusta aineistosta on silti nähtävissä, että digitaalisiin materiaaleihin liittyy edelleen merkittäviä käyttämättömiä mahdollisuuksia. Tämä pätee erityisesti opetukseen ja opetusmateriaaleihin, koska ne muodostavat huomattavan osan organisaation kokonaiskommunikaatiosta. On siis oletettavissa, että digitaalisten aineistojen osuuden ja organisaation koon kasvaessa myös digitaalisen tiedon säilymiseen liittyviin ongelmiin joudutaan kiinnittämään huomiota. Tapaustutkimuksen perusteella näyttää myös varsin selvältä, että nykytilanteessa tiedon pitkäaikaissäilytystä on lähes mahdoton erottaa informaatioresurssien yleisestä hallinnasta.

7. YHTEENVETO

Luvussa käydään läpi tutkimuksen keskeiset tulokset vastaamalla johdannossa esitettyihin tutkimuskysymyksiin. Lopuksi esitellään tutkimuksen kuluessa muodostuneet johtopäätökset ja esitellään mahdollisia jatkotutkimusaiheita.

7.1 Tulokset

Osakysymys 1 – Mikä digitaalisen tiedon pitkäaikaissäilytyksessä on kirjallisuuden valossa ongelmallista ja miten ongelmaa on yritetty ratkaista?

Digitaalisen tiedon pitkäaikaissäilytykseen liittyy sekä teknisiä että organisatorisia ongelmia. Tekniset ongelmat johtuvat pääosin siitä, etteivät pitkän kokemuksen tuloksena syntyneet, analogisen tiedon pitkäaikaissäilytykseen liittyvät käsitteet, toimintatavat ja periaatteet ole sellaisenaan toimivia käsiteltäessä digitaalista tietoa. Esimerkiksi dokumentin eri aspekteihin kuten sisältöön ja proveniensiin liittyvä informaatio, jotka analogisen tiedon kohdalla syntyy pitkälti dokumentin luovan prosessiin ohessa, joudutaan digitaalisen tiedon kohdalla luomaan ”tyhjästä” ja tallentamaan varsinaiseen digitaaliseen objektiin liitettäväksi metatiedoksi. Joihinkin teknisiin kysymyksiin ei vielä ole edes yksiselitteistä vastausta. Tällaisia ovat esimerkiksi kysymykset nykyaikaisten tallennusmedioiden todellisesta kestoikästä tai metatiedon liittämistä digitaalisiin objekteihin.

Organisatoriset ongelmat kietoutuvat yhteen teknisten ongelmien kanssa ja näiden erottaminen toisistaan onkin hankalaa. Teknisten ongelmien voidaankin sanoa olevan sekä organisatoristen ongelmien syy että seuraus. Teknologinen kehitys on osaltaan muuttanut taloudellista toimintaympäristöä ja pakottanut organisaatiot jatkuvaan rakennemuutokseen. Tämä vuoksi ne joutuvat soveltamaan entistä enemmän myös uutta teknologiaa. Perinteinen pitkäaikaissäilytyskonsepti, jossa tieto siirtyy keskitetyn arkiston vastuulle sen aktiivisen linkaaren päätyttyä, on digitaalisen tiedon aikakaudella osoittautumassa vanhanaikaiseksi. Digitaalisen objektin pitkäaikaissäilytys

on nimittäin sitä yksinkertaisempaa, mitä varhaisemmassa vaiheessa informaation elinkaarta pitkäaikaissäilytykseen liittyvät ongelmat huomioidaan organisaatiossa.

Erityisesti teknisten ongelmien ratkaisemiseksi on esitetty runsaasti erilaisia ideoita ja lähestymistapoja. Jotkut ideoista, kuten tietokonemuseoiden tai pysyvien standardien luominen ovat hyvin utopistisia ja mielenkiintoisia lähinnä ideatasolla. Toimivimpina ratkaisuuina pidetäänkin migraatiota eli digitaalisten aineistojen muuttamista muodosta toiseen sekä emulointia, jossa vanhojen ohjelmisto- ja laitteistoympäristöjen toimintaa matkitaan uudemmissa ympäristöissä. Näiden soveltuvuudesta ja käyttötarkoituksista vallitsee kirjallisuudessa laaja yhteisymmärrys. Migraatio soveltuu pääasialliseksi ratkaisuksi ja emulointia tarvitaan tilanteissa, joissa digitaalisen objektin säilyminen on erityisen riippuvaista tietyn ohjelmistoympäristön säilymisestä.

On kuitenkin huomattava, ettei emuloinnissa tai migraatiossa ole kyse yksinkertaisesta ratkaisusta, vaan kokoelmasta erilaisia ratkaisuja, joita sovelletaan hyvinkin vaihtelevasti tilanteesta riippuen. Näiden lähestymistapojen käyttäminen vaatii myös runsaasti resursseja ja erityisesti emulointi on myös teknisesti hyvin monimutkaista ja tämänkin vuoksi toimiva ratkaisu ainoastaan erityistilanteissa. Emulointi- ja migraatoratkaisujen toimivuuden lisäksi kirjallisuudessa vallitsee yhteisymmärrys metatiedon ja standardien pitkäaikaissäilytystä tukevasta roolista.

Parhaiten digitaalisten tallennusmuotojen tiedon pitkäaikaissäilytykselle aiheuttama murros näkyy tarkasteltaessa organisatorisia ratkaisuja käsittelevää kirjallisuutta. Käytävässä tieteellisessä keskustelussa voidaan kärjistetysti erottaa kaksi toisistaan poikkeavaa suuntausta. Ensimmäisen suuntauksen mukaan tiedon pitkäaikaissäilytyksen tulisi teknologisesti muutoksesta huolimatta tapahtua organisaatiossa kuten tähänkin asti, eli keskittyessä koko organisaation kattavassa arkistossa, jonne tieto siirretään kun sille ei ole enää käyttöä organisaation muissa osissa. Toisen suuntauksen mukaan tiedon pitkäaikaissäilytysfunktio tulisi hajauttaa osaksi organisaation yleistä informaatioresurssien hallintaa. Tällöin tietoresurssin hallinnan erilaisiin vaiheisiin jakavista malleistakin voitaisiin luopua jopa kokonaan. Jälkimmäistä vaihtoehtoa perustellaan erityisesti sillä, että digitaalisen objektin säilyvyyttä on vaikea varmistaa, ellei säilyvyyteen kiinnitetä huomiota objektin elinkaaren alkuvaiheista lähtien.

Toisaalta ensimmäistä vaihtoehtoa on puolustettu kysymällä, että kiinnitetäänkö tiedon pitkäaikaissäilytykseen yleensäkin riittävästi huomiota, jos se siirretään osaksi organisaation informaatioresurssien yleistä hallintaa. Tässä tutkimuksessa läpikäydyn kirjallisuuden perusteella on syytä uskoa hajautettujen ratkaisujen toimivuuteen, vaikka elinkaarimalleja jossain määrin sovellettaisiinkin. Digitaalisen tiedon säilyttäminen vaatii analogista tietoa huomattavasti enemmän aktiivista toimintaa ja tämän seurauksena organisaation pitkäaikaissäilytysfunktio tulee mitä todennäköisimmin muuttumaan passiivisesta erillisfunktioista aktiiviseksi tiedonhallintaprosessiksi. Kirjallisuudessa tätä ei kuitenkaan ole erityisemmin huomioitu. Yleisesti todetaan, että pitkäaikaissäilytys tulee siirtymään enemmän osaksi yleistä informaatioresurssien hallintaa. Tämän tutkimuksen puitteissa ei kuitenkaan löytynyt asian selkeästi huomioivaa mallia tai viitekehystä

Osakysymys 2 - Soveltuuko kommunikatiivisiin genreihin perustuva mittausmenetelmä organisaation tietoresurssien tarkasteluun tiedon pitkäaikaissäilytyksen näkökulmasta?

Genreihin pohjautuva mittausmenetelmä osoittautui toimivaksi ja kattavaksi tavaksi kartoittaa organisaation tietovarantoja pitkäaikaissäilytyksen näkökulmasta, mikäli tausta-aineiston laadusta voidaan varmistua. Se antoi hyvän yleiskuvan kohdeorganisaation tietovarannoista ja toi esiin yllättäviäkin piirteitä kuten tutkimuskysymyksen käsittelyn yhteydessä tulee esille. Tässä tutkimuksessa tulosten luotettavuutta ja mittausmenetelmän soveltuvuuden arviointia vaikeutti kuitenkin se tosiasia, että käytetty genrejaottelu oli peräisin toiseen tarkoitukseen tehdystä tutkimuksesta. Tämä on kuitenkin enemmän menetelmän soveltamistapaan, kuin varsinaiseen menetelmään liittyvä puute. Tutkimuksen perusteella genrepohjaista analyysimenetelmää voidaan hyödyntää luotettavammin, mikäli analyysin lopulliset käyttötarkoitukset otetaan huomioon jo genrejen tunnistamis- ja nimeämisen vaiheessa. Menetelmän jatkokehityksessä tulisi myös pohtia uusia tapoja tulosten kattavuuden ja luotettavuuden arvioimiseksi, sillä tässä tutkimuksessa käytetyt arviointimenetelmät eivät olleet erityisen luotettavia. Kovin pitkälle meneviä päätelmiä genreihin perustuvasta analyysimenetelmästä ei kuitenkaan voida yhden tapaustutkimuksen perusteella tehdä, eikä se kuulunut esitetyn tutkimuksen tavoitteisiinkaan.

Tutkimuskysymys: Saavatko kirjallisuudessa esitetyt digitaalisen tiedon pitkäaikaissäilytystä koskevat näkemykset tukea tarkasteltaessa kohdeorganisaation tilaa aihepiirin näkökulmasta?

Tiedon pitkäaikaissäilytys tapahtuu kohdeorganisaatiossa koko taustaorganisaation eli yliopiston kattavassa keskitetyssä arkistossa analogisia tallennusmuotoja käyttäen, vaikka organisaation sisäinen kommunikaatio on kasvavissa määrin digitaalista. Vallitseva järjestelmä on toistaiseksi toiminut hyvin, koska tällä hetkellä pitkäaikaissäilytettävää materiaalia on melko vähän, pitkäaikaissäilytys on hyvin ohjeistettu ja aineisto on pitkälti tekstimuotoista ja siten helposti muunnettavissa analogiseen muotoon pitkäaikaissäilytystä varten. Aineisto on myös siirtynyt elinkaarensa passiiviseen vaiheeseen ja sitä ei tarvita varsinaisessa organisaatiossa.

Tällä hetkellä tiedon pitkäaikaissäilytys ei siis ole digitaalisten tallennusmuotojen yleistymistä huolimatta vakava ongelma kohdeorganisaatiossa johtuen organisaation pienestä koosta ja keskeisten organisatoristen prosessien kuten opetuksen alhaisesta digitaalisuusasteesta. Samalla tutkimus osoitti, että tietoresurssien kuten opetusmateriaalin digitalisointiin liittyy edelleen merkittäviä käyttämättömiä mahdollisuuksia johtuen jo yksistään aineiston suuresta määrästä. Tämä raportoitiinkin gradun pohjalta laaditussa erillisessä julkaisussa (Tyrväinen, Järvenpää ja Sievänen, 2003). Tiedon pitkäaikaissäilytystä teknisestä näkökulmasta tarkasteleva kirjallisuus ei siis ole kohdeorganisaation kannalta nykytilanteessa kovinkaan ajankohtaista. Opetusmateriaalien suuri määrä, yhdessä niihin liittyvien digitalisoimissuunnitelmien kanssa antavat kuitenkin viitteitä siitä, että digitaalisten aineistojen osuus tulee kasvamaan organisaatiossa merkittävästi lähitulevaisuudessa. Tällöin myös tutkimuksessa esitellyt ongelmat tulevat ajankohtaisemmiksi.

Tiedon pitkäaikaissäilytystä ja informaatioresurssien hallintaa organisatorisesta näkökulmasta tarkastelevat lähteet saavat kuitenkin tukea tarkasteltaessa kohdeorganisaation tilannetta laajemmin. Toimintaympäristössä tapahtuneet nopeat muutokset ovat pakottaneet myös kohdeorganisaation muuttumaan nopeasti. Organisatorista muutosta kuvannee parhaiten se, että kohdeorganisaatio on muodostunut

vasta äskettäin kahden erillisen organisaation osan yhdistyessä. Kasvun nopeudesta saa puolestaan käsityksen tarkastelemalla pääaineopiskelijoiden määrää kohdeorganisaatiossa, joka vuonna 2002 kasvoi lähes kolmella sadalla.

Nopea kasvu ja tekninen kehitys ovat myös vaikuttaneet organisaation tiedonhallintaan, ja siten myös tiedon pitkäaikaissäilytykseen kirjallisuudessa kuvatulla tavalla. Organisaatio muuttuu nopeammin kuin sen informaationhallinta. Kohdeorganisaatiossa on nähtävissä esimerkiksi Koulopouloksen ja Frappaolon (1995) kirjassaan kuvamaan informaatioaariston piirteitä. Morrisin (1998) ja Simpsonin (1998) mainitsemmalla tavalla vastuu tiedosta on myös siirtynyt lähemmäksi sen käyttäjiä ja hajautunut, eli informaatioresurssit ovat siirtyneet työntekijöiden henkilökohtaisiin tietokoneisiin. Käytössä on useita pieniä ja keskenään yhteensopimattomia tietojärjestelmiä, jotka sisältävät osittain päällekkäistä tietoa. Suurin osa tiedosta myös tuotetaan henkilökohtaisten tietokoneiden avulla, mutta tulostetaan paperille käyttöä ja säilytystä varten. Informaatioresurssien hallinta organisaatiossa vaikuttaakin olevan enemmän yksittäisten työntekijöiden kokemuksen, kuin keskitetyn tietoresurssien hallintajärjestelmän varassa.

Myös Dollarin (1992) ja Bantinin (2000) kuvaama, keskitetyn ja hajautetun arkistointimallin välisistä eroista käytävä keskustelu on tapaustutkimuksen valossa hyvin aiheellista. Keskitetty arkistointimalli on todennäköisesti toimiva yliopiston kaltaisessa monitahoisessa organisaatiossa. Tässä kuitenkin piilee vaara, että arkistointifunktio irtaantuu teknisen kehityksen myötä organisaation jokapäiväisestä toiminnasta. Tällöin voi syntyä tilanne, jossa arkisto saa arkistointisuunnitelman mukaisen analogisen aineiston kohdeorganisaatiossa. Samalla kohdeorganisaation on kuitenkin syntynyt uusia arvokkaita digitaalisia informaatioresursseja, joiden olemassaolosta ei olla sen ulkopuolella edes tietoisia. Tämäntapaisen tilanteen välttämiseksi huomio onkin kiinnitettävä informaation koko elinkaareen, eikä vain sen passiiviseen vaiheeseen. Lisäksi tarvitaan entistäkin enemmän organisaation eri osien välistä yhteistyötä.

7.2 Johtopäätökset

Tutkimus lähti liikkeelle esittelemällä tiedon pitkäaikaissäilytyksen merkitystä ja tavoitteita. Vaiheen johtopäätöksenä todettiin, että kyse on yllättävän monimutkaisesta asiasta, jota digitaalisen tiedon aikakaudella on vaikea irrottaa yleisesti organisaation informaationhallintaan liittyvistä ongelmista. Tarkasteltaessa seuraavaksi asiaan liittyviä teknisiä ja organisatorisia ongelmia vaikutelma vahvistui. Samalla selvisi esiteltyjen teknisten ratkaisujen monipuolisuus ja ongelman teknisten aspektien kirjallisuudessa saaman huomion suuri määrä.

Yksi kirjallisuuskatsauksen keskeisistä johtopäätöksistä oli, että digitaalisen tiedon säilymisessä kyse on ennen kaikkea organisatorisesta ongelmasta. Teknisiä ratkaisuja on saatavilla runsaasti, mutta niiden soveltaminen on vaikeaa heterogeenisten tietojärjestelmien, rajallisten resurssien sekä nopeiden teknistaloudellisten muutosten hallitsemisessa organisaatioissa. Perinteinen keskitetty pitkäaikaissäilytyksen organisointimalli, jossa tiedon elinkaari jaetaan aktiiviseen ja passiiviseen vaiheeseen on osoittautumassa vanhanaikaiseksi. Huonosti dokumentoidun digitaalisen aineiston säilyvyyden ja todistusvoimaisuuden varmistaminen voi olla mahdotonta, mikäli asiaan ei ole kiinnitetty huomiota jo aineiston luonnin yhteydessä. Tämän vuoksi kirjallisuudessa korostetaan uusia organisointimalleja, jossa tiedon pitkäaikaissäilytys nähdään organisaation yleiseen informaatioresurssien hallintaan liittyvänä jatkuvana prosessina. Tällaisesta näkökulmasta asiaa on kuitenkin tähän mennessä tarkasteltu kirjallisuudessa melko vähän. Toisaalta, vaikka teknologisia ratkaisuja on käsitelty tieteellisessä kirjallisuudessa huomattavasti enemmän, oli niistäkin käytävä keskustelu pitkälti teoreettista. Menetelmistä ja käytännöistä ei ole ehtinyt syntyä riittävästi kokemusta, jotta ”hyllyltä” saatavia ratkaisuja olisi voinut muodostua.

Tapaustutkimuksesta kokonaisvaikutelmaksi jäi, että kohdeorganisaatio on keskellä analogisen ja digitaalisen maailman välistä siirtymävaihetta. Tämä näkyi konkreettisesti siten, että merkittävä osa organisaation tiedosta tuotetaan digitaalisesti, mutta muutetaan käsittelyä ja säilytystä varten analogiseksi. Tämän seurauksena digitaalisen tiedon pitkäaikaissäilytykseen ei kohdeorganisaatioissa ainakaan vielä liity merkittäviä

ongelmia. Siirtymävaiheen edetessä ja digitaalisten tallennusmuotojen käytön tullessa monipuolisemmaksi tilanne voi kuitenkin muuttua nopeasti. Akuuttien ongelmien puuttuminen ei siis tarkoita sitä, ettei asiaan tarvitse kiinnittää huomiota. Digitaalisen tiedon pitkäaikaissäilytykseen liittyvien ongelmien ratkaiseminen on kuitenkin helpompaa, jos ne huomioidaan mahdollisimman varhaisessa vaiheessa.

7.3 Jatkotutkimusaiheita

Gradussa käsitelty aihepiiri on vasta muotoutumassa omaksi tutkimusalueekseen, ja siten myös mahdollisia jatkotutkimusaiheita löytyi tutkimuksen kuluessa lukuisia. Seuraavassa esityksessä tyydytäänkin esittelemään yksityiskohtaisten aiheiden sijasta mahdollisia tutkimussuuntia.

Ensinnäkin digitaalisen tiedon säilytysongelman laajuuden ja ulottuvuuksien kartoittamiseen tähtäävälle empiiriselle tutkimukselle olisi selvästi tarvetta. Tämän hetkissä kirjallisuudessa ongelman merkitystä kyllä korostetaan, mutta sen todellisesta laajuudesta löytyy niukasti tutkittua tietoa. Tällaisen tutkimuksen voisi hyvin kytkeä myös osaksi organisaation tietoresursseihin kohdistuvaa yleistä tutkimusta käyttäen hyväksi esimerkiksi genreihin perustuvia menetelmiä. Digitaalisen tiedon pitkäaikaissäilytystä koskevaa tutkimusta tulisi tehdä nykyistä enemmän myös muiden kuin kirjasto- ja arkisto-organisaatioiden näkökulmasta. Asian tieteellinen tarkastelu esimerkiksi liikeyrityksen näkökulmasta toisi varmasti esiin uusia ja mielenkiintoisia näkökulmia.

Lähestymistapojen tutkimuksessa tulisi siirtyä enemmän käytännölliseen suuntaan. Teoreettisesti mielenkiintoisten, mutta hieman utopististen kaikki ongelmat kerralla ratkaisevien ”hopealuotien” sijasta tarvittaisiin enemmän ohjeita, standardeja, suuntaviivoja ja hyväksi havaittuja käytäntöjä, joista olisi apua digitaalisen tiedon säilyvyysongelman jokapäiväisissä toiminnoissaan kohtaaville organisaatioille. Tähän liittyy olennaisesti myös metatiedon luomista organisaatioissa ja sen roolia pitkäaikaissäilytyksessä käsittelevä tutkimus. Lisäksi digitaalisen tiedon

pitkäaikaissäilytyksen liittyy monia muita kartoittamattomia alueita, kuten säilymisen varmistamisesta syntyvät kustannukset ja niiden jakautuminen.

LÄHDELUETTELO

Kaikki verkkolähteet on tarkistettu 7.3.2003, jolloin ne olivat toimivia ja samanlaisia kuin varsinaisella viittaushetkellä.

Arkistolaitos. 2001. Sähköisten tietojärjestelmien ja aineistojen käsittely [online]. Määräys ja ohje 126/40/01, 22.5.2001. Helsinki: Arkistolaitos. Saatavilla www-muodossa <<http://www.narc.fi/sahk/>>.

Bantin, P. 2000. Electronic Records Management - A Review of the Work of a Decade and a Reflection on Future Directions [online]. Encyclopedia of Library and Information Science, Volume 71, 47-81. Saatavilla doc-muodossa <<http://www.indiana.edu/~libarch/ER/>>.

Beagrie, N., Greenstein D., Pressler C. 2001. A Strategic Policy Framework for Creating and Preserving Digital Collections [online]. Arts and Humanities Data Service. Saatavilla pdf-muodossa <<http://ahds.ac.uk/strategic.pdf>>.

Bearman, D. 1999. Reality and Chimeras in the Preservation of Electronic Records. D-Lib Magazine [online], (5) 4. Saatavilla www-muodossa: <<http://www.dlib.org/dlib/april99/bearman/04bearman.html>>.

Bullock, A. 1999. Preservation of Digital Information: Issues and Current Status [online]. Network Notes 60. National Library of Canada. Saatavilla www-muodossa <<http://www.nlc-bnc.ca/publications/1/p1-259-e.html>>.

Carrol B.C., Hodge, G. 1999. Digital Electronic Archiving: The State of the Art and The State of the Practice. International Council for Scientific and Technical Information Policy Committee.

Conway, P. 1996. Preservation in the Digital World [online]. Council on Library and Information Resources, CLIR Reports. Saatavilla [www-muodossa <http://www.clir.org/pubs/reports/conway2/index.html>](http://www.muodossa.org/CLIR/CLIR%20Reports/CLIR%20Reports%20Index.html).

Day, M. 1999. Metadata for Digital Preservation: an Update. Ariadne [online], Issue 22. Saatavilla [www-muodossa <http://www.ariadne.ac.uk/issue22/metadata>](http://www.muodossa.org/Ariadne/Ariadne%20Issue%2022/Metadata).

Dempsey L., Heery R. 1998. Metadata: a Current View of Practice and Issues. Journal of Documentation, 54 (2), 145-172.

DLM-Forum. 1997. Guidelines on Best Practices for Using Electronic Information – How to Deal with Machine Readable Data and Electronic Documents. Luxembourg: Office for Official Publications of the European Communities.

Dollar, C.M. 1992. Archival Theory and Information Technologies, University of Maserata, Ancona.

Garrett, J., Waters D. 1996. Preserving Digital Information [online]. The Commission on Preservation and Access & The Research Libraries Group. Saatavilla [www-muodossa <http://www.rlg.org/ArchTF/tfadi.index.htm>](http://www.muodossa.org/RLG/ArchTF/tfadi.index.htm).

Granger, S. 2000. Emulation as a Digital Preservation Strategy. D-Lib Magazine [online], (6)10. Saatavilla [www-muodossa <http://www.dlib.org/dlib/october00/granger/10granger.html>](http://www.muodossa.org/DLib/October00/granger/10granger.html).

Hakala, J. 2001. Metadata for Referencing and Archival Usage. Proceedings of the Second ICSU-UNESCO International Conference on Electronic Publishing in Science. UNESCO. Saatavilla [www-muodossa <http://associnst.ox.ac.uk/~icsuinfo/hakalafin.htm>](http://www.muodossa.org/ICISU/ICISU%20Proceedings/Hakala)

Hedstrom, M. 1995. Digital Preservation: a Time Bomb for Digital Libraries [online]. Saatavilla [www-muodossa <http://www.uky.edu/%7Ekiernan/DL/hedstrom.html>](http://www.muodossa.org/UKY/UKY%20DL/Hedstrom).

Hedstorm, M., Montgomery, S. 1998. Digital Preservation Needs and Requirements in RLG Member Institutions. Research Libraries Group.

Hedstrom, M., Lampe C. 2001. Emulation vs. Migration: Do Users Care? RLG DigiNews (6) 5. Research Libraries Group.

Hirtle, P. 2000. Archival Authenticity in a Digital Age [online]. Teoksessa Authenticity in a Digital Environment. Council on Library and Information Resources, CLIR Reports. Saatavilla pdf-muodossa <<http://www.clir.org/pubs/reports/pub92/pub92.pdf>>

Hodge, G. 2000. Best Practices for Digital Archiving – an Information Life Cycle Approach. D-Lib Magazine [online], (6) 1 [viitattu 10.2.2003] Saatavilla www-muodossa <<http://www.dlib.org/dlib/january00/01hodge.html>>.

Koulopoulos, T. T., Frappaolo C. 1995. Document Management: What's the Problem? Teoksessa Electronic Document Management Systems: A Portable Consultant, New York: McGraw-Hill, 13-27.

Kuny, T. 1997. A Digital Dark Ages? Challenges in the Preservation of Electronic Information. 63RD IFLA Council and General Conference.

Landauer, T. K. 1986. How Much Do People Remember? Some Estimates of the Quantity of Learned Information in Long-term Memory. Cognitive Science, (4) 10.

Levy, D. 1998. Heroic Measures: Reflections on the Possibility and Purpose of Digital Preservation. Teoksessa Digital Libraries 98: The Third ACM Conference on Digital Libraries. ACM, Pittsburgh.

Liikearkistoyhdistys. 2001. Pureeko ajan hammas? Arkistointi ja asiakirjojen säilytysajat. Liikearkistoyhdistys ry:n julkaisu 17. Helsinki: Liikearkistoyhdistys ry.

Lyman P., Varian H. R. 2000. How Much Information? [online]. Saatavilla www-muodossa <<http://www.sims.berkeley.edu/research/projects/how-much-info>>.

Mackenzie J.S., v.d. Walle J. 1996. Deposit Collections of Electronic Publications. European Commission, DG XIII-E/4. Libraries in the information society series. Luxembourg: Office for Official Publications of the European Communities.

McNurlin, B., Sprague, R. 2001. Information Systems management in Practice Prentice Hall.

Metsäranta, P. 1999. Rakenteisen tiedon säilyttäminen - XML-dokumentti OAIS-viitemallissa. Pro Gradu -tutkielma, Jyväskylän yliopisto.

Morris, R. J. 1998. Electronic Documents and the History of Late Twentieth Century: Black Holes or Warehouses? Teoksessa E. Higgs (toim.) History and Electronic Artefacts, Oxford: Clarendon, 31-49.

Päivärinta T., Tyrväinen P. 2001. Structuring Information by Genres to Bridge the Social and Technological in Information Resources Management – Leavitt's Framework Revis(IT)ed. Teoksessa Proceedings of the 24th Scandinavian Research Seminar on Information Systems (IRIS). Ulvik, Norway.

Päivärinta, T., Halttunen, V., Tyrväinen, P. 2001. A Genre-Based Method for Information Systems Planning. Teoksessa Information Modelling in the New Millennium. Hershey PA, Idea Group, 70-93.

Päivärinta T. 2000. Organisaation dokumenttien hallinta ja sen kehittäminen: periaatteet, osa-alueet ja lähtökohdat. METODI-raportti 008. Jyväskylän yliopisto. Tietotekniikan tutkimusinstituutti.

Päivärinta T., Tyrväinen P., Ylimäki T. 2002. Defining Organizational Document Metadata: a Case Beyond Standards. Proceedings of the Xth European Conference on Information Systems (ECIS). Gdansk, Poland.

Roos, C-M. 2001. Sähköisten asiakirjojen erityiskysymyksiä. Seminaariesitys. Dokumenttien hallinta –seminaari. Hotelli Kämp, Helsinki.

Rothenberg, J. 1999a. Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation [online]. Council on Library and Information Resources, CLIR Reports. Saatavilla pdf-muodossa <<http://www.clir.org/pubs/reports/rothenberg/pub77.pdf>>.

Rothenberg, J. 1999b. Ensuring the Longvity of Digital Information [online]. Council on Library and Information Resources, CLIR Reports. Saatavilla pdf-muodossa <<http://www.clir.org/pubs/archives/ensuring.pdf>>.

Rothenberg, J. 2000. An Experiment in Using Emulation to Preserve Digital Publications. The Koninklijke Bibliotheek, Den Haag.

Salminen, A., Lyytikäinen, V., Tiitinen P., Mustajärvi, O. 2001. Experiences of SGML Standardization: The Case of the Finnish Legislative Documents. HICSS-34, 34th Hawaii International Conference on System Sciences. IEEE Computer Society.

Shepherd, M., Watters C. 1998. The Evolution of Cybergenres. HICSS-31, 31th Hawaii International Conference on System Sciences. IEEE Computer Society.

Simpson H. 1998. The Management of Electronic Information Resources in a Corporate Environment. Teoksessa E. Higgs (toim.) History and Electronic Artefacts, Oxford: Clarendon, 87-101.

Stenvall, J. 2001. Metadata elektronisen julkaisujen pitkäaikaissäilytyksessä. Tampereen yliopisto, Informaatiotutkimuksen laitos, Pro gradu –tutkielma.

Tyrväinen, P. 2003. Estimating Applicability of New Mobile Content Formats to Organizational Use. HICSS-36, 36rd Hawaii International Conference on System Sciences. IEEE Computer Society.

Tyrväinen, P., Päivärinta, T. 2003. How Digital is Communication in Your Organization? A Metrics and an Analysis Method. ICEIS'03, 5th International Conference on Enterprise Information Systems. ICEIS Press.

Tyrväinen, P., Järvenpää, M., Sievänen A. 2003. On Estimating the Amount of Learning Materials: a Case Study. ICEIS'03, 5th International Conference on Enterprise Information Systems. ICEIS Press.

Yates J., Orlikowski W., Renneceker J. 1997. Collaborative Genres for Collaboration: Genre Systems in Digital Media. HICSS-30, 30th Hawaii International Conference on System Sciences. IEEE Computer Society.

Yates J., Orlikowski W. 1992. Genres of Organizational Communication: A Structural Approach to Studying Communication and Media. *Academy of Management Review*, 17, 2, 299-326.