

Opinnäytetyö (AMK)

Kirjasto- ja tietopalvelun koulutusohjelma

2011

Maiju Lammela

# INTERNETISSÄ OLEVA AINEISTO OSAKSI MAAKUNTAKOKOELMAA

– aineiston löytyminen ja pitkäaikaissäilytys



TURUN AMMATTIKORKEAKOULU  
TURKU UNIVERSITY OF APPLIED SCIENCES

Maiju Lammela

## INTERNETISSÄ OLEVA AINEISTO OSAKSI MAAKUNTAKOKOELMAA

### – aineiston löytyminen ja pitkäaikaissäilytys

Tämän opinnäytetyön tavoitteena oli tutkia internetissä olevien digitaalisten aineistojen löytymistä ja pitkäaikaissäilytystä. Opinnäytetyö tehtiin Porin kaupunginkirjasto – Satakunnan maakuntakirjastolle ja sen Satakunta-osastolle, joka ylläpitää Satakunnan maakuntakokoelmaa eli Satakuntakokoelmaa. Koska yhä enenevässä määrin aineistoja julkaistaan verkossa digitaalisessa muodossa perinteisen painetun sijaan, on maakuntakokoelman ylläpitäjän panostettava siihen, että se löytää aineistot internetistä. Lisäksi aineistojen pitkäaikaista säilytystä on mietittävä, jotta aineistot olisivat tulevaisuudessakin asiakkaiden ja muiden tiedontarvitsijoiden käytössä.

Tässä tutkimuksessa käytettiin tutkimusmenetelmänä puolistrukturoitua haastattelua. Tutkimusta varten haastateltiin seitsemän eri maakuntakirjaston edustajaa, ja heiltä kysyttiin oman kirjastonsa käytännöistä verkkotiedonhausta ja aineistojen pitkäaikaissäilytyksestä. Lisäksi tehtiin kaksi asiantuntijahaastattelua: Porin kaupungin ympäristöviraston johtajaa haastateltiin ympäristöviraston verkossa olevista julkaisuista ja Kansalliskirjaston kirjastoverkkopalvelujen pääsuunnittelijaa haastateltiin Kansallisen digitaalisen kirjaston eli KDK:n edustajana KDK:n tulevasta pitkäaikaissäilytysjärjestelmästä.

Haastatteluissa kävi ilmi, että maakuntakirjastot eivät juurikaan tee järjestelmällistä tiedonhakua maakuntakokoelmiin soveltuvista aineistoista joitakin poikkeuksia lukuun ottamatta, ja syynä tähän on enimmäkseen resurssipula. Pitkäaikaissäilytystä varten puolestaan ei ole tehty lopullisia suunnitelmia missään haastatelluista maakuntakirjastoista, vaan pääsääntöisesti niissä odotellaan Kansallisen digitaalisen kirjaston ratkaisuja.

Kaikkien haastattelujen perusteella pystyttiin tekemään joitakin ehdotuksia Satakuntakokoelman verkkotiedonhakua ja aineistojen löytymistä parantamaan. Näitä ehdotuksia ovat RSS-syötteiden sekä Google Alertin käyttöönotto sekä wikien ja blogien hyödyntäminen. Pitkäaikaissäilytystä varten esiteltiin KDK:n suosituksia pitkäaikaissäilytyksessä käytetyistä tiedostomuodoista ja metadatan tallennuksesta.

#### ASIASANAT:

maakuntakirjastot, paikalliskokoelmat, pitkäaikaissäilytys, tiedonhaku, verkkoaineisto

Maiju Lammela

## ONLINE MATERIALS IN REGIONAL COLLECTIONS

### – finding and preserving relevant materials

The objective of this Bachelor's thesis was to study how the digital material on the Internet can be collected and preserved in regional collections. The thesis was made for Pori Public Library – Regional Library of Satakunta and its Satakunta department, which is in charge of building and maintaining the regional Satakunta collection. An increasing amount of material is published merely online and that kind of material is not available in print. Hence, it is of vital importance that the libraries in charge of the regional collections focus on finding the relevant material on the Internet. Also the long-term preservation of the material must be considered, because the material must be accessible for the patrons and people seeking information in the future.

The research method used in this thesis was a half structured interview. For this study, seven persons representing different regional libraries were interviewed. They were asked about their practices in searching information on the Internet and the policies regarding the long-term preservation of online material. In addition, two experts were interviewed: the Director of the Environmental Protection Office in Pori was interviewed about the online publications of the Environmental Protection Office and a representative of National Digital Library project was interviewed about the forthcoming long-term preservation system.

The interviews showed that, apart from some exceptions, there is no systematic information retrieval system used in the regional libraries to find the relevant online material for the regional collections. The reason for this was mostly lack of resources. For the long-term preservation there are no plans made yet in any regional library, but they are waiting for the recommendations of the National Digital Library project.

According to the interviews made for this study there were some suggestions made for the Satakunta collection for improving the information retrieval in the Internet. These suggestions included the introduction of RSS feeds and Google Alert service and utilizing the blogs and wikis. For the long-term preservation there were introduced the recommended file formats and metadata formats by the National Digital Library project.

#### KEYWORDS:

regional libraries, regional collections, long-term preservation, information retrieval, digital materials

# SISÄLTÖ

<b>SANASTO</b>	<b>6</b>
<b>1 JOHDANTO</b>	<b>7</b>
<b>2 MAAKUNTAKIRJASTOT JA MAAKUNTAKOKOELMAT</b>	<b>9</b>
2.1 Maakuntakirjastot	9
2.2 Maakuntakokoelmat ja Satakuntakokoelma	10
<b>3 TIEDONHAKU INTERNETISTÄ</b>	<b>14</b>
3.1 Tiedonhaun prosessi	14
3.2 Internet tiedon kanavana	15
3.3 Hakukoneet	17
3.3.1 Sanahakukone Google	17
3.3.2 Muita sanahakukoneita	18
3.3.3 Metahakukoneet	19
3.4 Aihehakemistot	21
3.5 Viranomaistieto	22
3.6 Uuden tiedon automaattinen seuranta	23
<b>4 DIGITAALINEN PITKÄAIKAISSÄILYTYS</b>	<b>25</b>
4.1 Pitkäaikaissäilytyksen tavoitteet	25
4.1.1 Tiedon fyysinen elinkyky	26
4.1.2 Tiedon saaminen esitettäväksi	26
4.1.3 Tiedon ymmärrettävyys	26
4.1.4 Tiedon autenttisuus	27
4.1.5 Tiedon identifiointi	27
4.1.6 Tiedon saatavuus	27
4.1.7 Tiedon muuttumattomuus	28
4.2 Digitaalisen pitkäaikaissäilytyksen menetelmiä	28
4.2.1 Kopiointi	28
4.2.2 Migraatio ja konvertointi	30
4.2.3 Emulointi	31
4.3 Kansallinen digitaalinen kirjasto eli KDK	32
<b>5 TUTKIMUSHAASTATTELU</b>	<b>35</b>
5.1 Haastateltavat	35
5.2 Haastattelujen toteuttaminen	36
5.3 Haastattelukysymykset	37

5.4 Haastattelun tulokset	38
5.4.1 Tiedonhaku ja aineistojen löytyminen	39
5.4.2 Verkosta löytyneen aineiston luettelointi	41
5.4.3 Verkosta löytyneen aineiston pitkäaikaissäilytys	43
5.4.4 Muita haastatteluissa esille tulleita asioita	45
<b>6 VERKKOTIEDONHAKU JA PITKÄAIKAISSÄILYTYS</b>	
<b>SATAKUNTAKOKOELMASSA</b>	<b>47</b>
6.1 Ehdotuksia Satakuntakokoelman tiedonhakua varten	47
6.1.1 RSS-syötteiden käyttöönotto	47
6.1.2 Google Alertin käyttöönotto	49
6.1.3 Wikien ja blogien hyödyntäminen	50
6.2 Porin kaupungin ympäristöviraston julkaisut ja raportit	53
6.3 Pitkäaikaissäilytys ja Kansallinen digitaalinen kirjasto	55
6.3.1 Kansallisen digitaalisen kirjaston edustajan haastattelu	56
6.3.2 Kansallisen digitaalisen kirjaston suosittelemat tiedostomuodot	58
6.3.3 Kansallisen digitaalisen kirjaston suosittelemat metadataformaatit	61
<b>7 JOHTOPÄÄTÖKSET</b>	<b>65</b>
<b>LÄHTEET</b>	<b>68</b>

## **LIITTEET**

Liite 1. Haastattelukysymykset maakuntakirjastojen edustajille

Liite 2. Haastattelukysymykset Porin kaupungin ympäristöviraston johtajalle

Liite 3. Haastattelukysymykset Kansallisen digitaalisen kirjaston edustajalle

## SANASTO

Formaatti	Tiedostomuoto. Sovittu tapa tallentaa informaatiota tiettyyn digitaaliseen muotoon. Tällaisia muotoja ovat esimerkiksi Portable Document Format eli .pdf ja Tagged Image File Format eli .tiff. (Kansallinen digitaalinen kirjasto 2011a.)
PAS	Pitkäaikaissäilytys, pitkäaikaissäilyttäminen. Luonteeltaan pitkäaikaista säilyttämistä, jolloin aineistoa säilytetään vähintään 10 vuotta tai pysyvästi. (Kansallinen digitaalinen kirjasto 2011a.)
Standardi	Konsensusperiaatteella laadittu, standardisointijärjestön hyväksymä ja yleisesti saatavilla oleva asiakirja (Suomen standardisointiliitto SFS 2011a). Standardin tarkoitus on vähentää erilaisuutta tuotteiden välillä, jotta ne ovat sopivat toisiinsa, pystyvät toimimaan yhdessä ja ovat keskenään vaihdettavia (Suomen standardisointiliitto SFS 2011b).
Tesaurus	Asiasanasto, jossa asiasanojen väliset semanttiset suhteet on merkitty ja asiasanojen homonyymit ja synonyymit on tuotu esiin. Lisäksi asiasanojen ekvivalenssisuhteet, hierarkkiset suhteet ja assosiaatiosuhteet on ilmaistu esimerkiksi laajemmin, suppeammin ja rinnakkaistermein. (Blinnikka 1987, 146–147.)

# 1 JOHDANTO

Muiden muistiorganisaatioiden ohella myös kirjastoilla on velvollisuus etsiä ja säilyttää nykyaikana tuotettua tietoa jälkipolville. Painetun ja paperilla olevan tiedon aikakaudella tämä tehtävä oli vielä yksinkertaisempaa. Tiedon tuottajia oli vähemmän, ja näiden painamat julkaisut oli hankittavissa kirjaston kokoelmaan. Lisäksi tarvittiin vain riittävästi varastotilaa papereiden säilytykselle. Nykyään tehtävä on haasteellisempaa: Kuka tahansa voi tuottaa aineistoa verkkoon, mutta miten kaiken relevantin aineiston sieltä löytää? Entä missä digitaalinen aineisto pitäisi säilyttää ja missä muodossa, jotta se olisi vielä tulevaisuudessakin luettavissa ja käytettävissä?

Tämä opinnäytetyö on saanut lähtösäyöksensä Porin kaupunginkirjasto – Satakunnan maakuntakirjastolta, jonka Satakunta-osasto ylläpitää Satakunnan maakuntakokoelmaa eli Satakuntakokoelmaa. Sen tarkoituksena on kerätä ja säilyttää sekä kaunokirjallista että tieteellistä kirjallisuutta, joka käsittelee Satakuntaa tai sen henkilöitä tai joka on satakuntalaisen kirjoittamaa, että satakuntalaisia lehtiä, karttoja, pienpainatteita, aluetta koskevia opinnäytteitä ja av-aineistoa. (Porin kaupunginkirjasto – Satakunnan maakuntakirjasto 2011a.) Kokoelmaan pyritään siis hankkimaan kaikki Satakuntaan liittyvä aineisto, joten on varmistuttava siitä, että verkostakin saadaan talteen kaikki kokoelmaan soveltuva aineisto. Normaalisti yleisten kirjastojen tarkoituksena ei ole säilyttää aineistoja pitkään, vaan vanhentuneet tai muuten tarpeettomat aineistot poistetaan, mutta maakuntakokoelmasta ei poisteta mitään. Tämä luo omat vaatimuksensa aineiston säilytykselle, sillä sen on säilyttävä pitkään.

Tämä opinnäytetyö pyrkii vastaamaan kahteen eri kysymykseen. Ensimmäisenä pyritään löytämään vastaus siihen, miten aineistoa löytyy verkosta tehokkaasti ja erilaisia verkon tarjoamia palveluja hyödyntäen. Toisena pyritään vastaamaan siihen, miten kirjasto voisi valmistautua digitaalisen aineistonsa pitkäaikaiseen säilyttämiseen.

Kansallinen digitaalinen kirjasto eli KDK on opetus- ja kulttuuriministeriön hanke, jossa selvitetään kirjastojen, arkistojen, museoiden ja muiden kulttuuriperintöä säilyttävien organisaatioiden digitaalisten tietovarantojen hallintaa, jakelua ja pitkäaikaissäilytystä (Kansallinen digitaalinen kirjasto 2011b). Koska kaikki suomalaiset kirjastot käyttävät tulevaisuudessa hankkeessa aikaansaatuja digitaalisten aineistojen hallinta-, jakelu- ja pitkäaikaissäilytysratkaisuja, on perusteltua ensinnäkin tutustua hankkeeseen ja toiseksi miettiä yksittäisen kirjaston pitkäaikaissäilytysratkaisuja hankkeen vaatimusten mukaisesti. Tämän vuoksi tässä opinnäytetyössä pitkäaikaissäilytystä käsiteltäessä painotetaan KDK-hankkeen vaatimuksia.

Tässä tutkimuksessa käytettiin tutkimusmenetelmänä tutkimushaastattelua. Tutkimusta varten haastateltiin seitsemän eri maakuntakirjaston edustajaa, joilta kysyttiin heidän käytännöistään aineiston etsinnästä verkossa sekä digitaalisen aineiston pitkäaikaissäilytyksestä. Lisäksi haastateltiin Porin kaupungin ympäristöviraston johtajaa Matti Lankiniemeä ympäristöviraston verkkosivuilla olevista julkaisuista ja raporteista sekä Kansalliskirjaston kirjastoverkkopalvelujen pääsuunnittelijaa Esa-Pekka Keskitaloa, joka on työskennellyt KDK:n pitkäaikaissäilytysjaostossa. Keskitaloa haastateltiin yksittäisen maakuntakirjaston tilanteesta KDK:n pitkäaikaissäilytysratkaisussa.

Työn rakenne etenee teoreettisen taustan kautta käytännön kokemuksiin ja joidenkin ongelmien ratkaisuehdotuksiin. Ensin käsitellään maakuntakirjastoja ja maakuntakokoelmia yleisesti. Tämän jälkeen tutustutaan tiedonhakuun internetistä ja siihen, minkälaisia keinoja hakijalla on käytössään internetin sisältöjen etsintään ja seuraamiseen. Seuraavaksi käsitellään pitkäaikaissäilytystä yleisesti: Mitä tavoitteita pitkäaikaissäilytykselle asetetaan? Minkälaisia keinoja on olemassa aineiston säilymisen turvaamiseksi? Lisäksi tutustutaan Kansallinen digitaalinen kirjasto- eli KDK-hankkeeseen.

Käytännön kokemuksia esitellään tutkimushaastatteluosiossa, jossa kerrotaan maakuntakirjastojen edustajille tehdystä haastattelusta ja sen tuloksista. Tämän jälkeen esitellään joitakin käytännön ratkaisuehdotuksia Satakuntakokoelmalle sekä verkkotiedonhausta että pitkäaikaissäilytyksestä.



## 2 MAAKUNTAKIRJASTOT JA MAAKUNTAKOKOELMAT

Suomessa on monia eri asemassa olevia kirjastoja. Yleisten eli kunnallisten kirjastojen lisäksi Suomen kirjastoverkoston muodostavat tieteelliset ja ammattikorkeakoulukirjastot, erikoiskirjastot sekä koulujen ja oppilaitosten kirjastot. (Opetus- ja kulttuuriministeriö 2011a.) Yleisten kirjastojen toimipisteitä oli Suomessa vuonna 2010 yhteensä 1 495, joista 312 oli pääkirjastoja, 486 sivukirjastoja, 44 laitoskirjastoja, 495 muista palvelupisteitä ja 155 kirjastoautoja ja -veneitä (Suomen yleisten kirjastojen tilastot 2011).

### 2.1 Maakuntakirjastot

Jokaisessa Suomen maakunnassa on maakuntakirjasto. Opetus- ja kulttuuriministeriö on antanut 18 yleiselle kirjastolle tehtävän toimia maakuntakirjastona. Poikkeuksena on Maarianhamina, joka ei kuulu opetus- ja kulttuuriministeriön alaisuuteen. (Opetus- ja kulttuuriministeriö 2011a.) Maakuntakirjastojen tehtävistä on säädetty Kirjastoasetuksella, ja tehtävät ovat

- 1) tukea alueensa yleisten kirjastojen tieto- ja kaukopalvelua
- 2) kehittää toiminta-aluetta koskevaa tietopalvelua
- 3) perehdyttää toiminta-alueen kirjastojen henkilöstöä kirjastotyön uusiin toimintamuotoihin ja kehittämishankkeisiin sekä
- 4) suorittaa asianomaisen ministeriön antamat muut tehtävät.

(Kirjastoasetus 18.12.1998/1078)

Maakuntakirjastojen synty oli seurausta vuonna 1961 annetusta uudesta kirjastolaista, ja ensimmäisen maakuntakirjasto-oikeudet annettiin vuonna 1962 Joensuun kaupunginkirjastolle. Seuraavat maakuntakirjasto-oikeudet annettiin vasta 1960-luvun loppupuolella, vuonna 1966 Rovaniemelle, 1967 Kuopiolle ja 1968 Vaasalle ja Seinäjoelle. Seuraavien vuosien aikana oikeudet saivat vuonna 1970 Turku, 1971 Kouvola, 1972 Tampere, 1973 Lappeenranta ja Oulu, 1974 Mikkeli, 1975 Pori ja Lahti, 1977 Espoo ja 1978 Jyväskylä. (Mäkinen 2009,

411–412.) Myöhemmin oikeudet ovat saaneet vuonna 1988 Kajaani, vuonna 1995 Hämeenlinna ja Kokkola sekä vuonna 2001 Porvoo. (Helsingin kaupunginkirjasto – yleisten kirjastojen keskuskirjasto 1990, 6; Hämeenlinnan kaupunginkirjasto 2011; Kokkolan kaupunginkirjasto 2011; Porvoon kaupunginkirjasto 2011). Vuoden 2011 alusta maakuntakirjastojen alueisiin tuli muutos, kun Espoo luopui Uudenmaan maakuntakirjastotehtävistä ja Porvoosta tuli koko Uudenmaan maakuntakirjasto entisen Itä-Uudenmaan sijaan (Kirjastot.fi 2011b).

Vuoden 1962 kirjastolaki ja -asetus määrsivät, että maakuntakirjastoksi soveltuvalla kirjastolla tuli olla riittävät kokoelmat niin kooltaan kuin laadultaankin, pätevä johtaja ja sopiva huoneisto. Lisäksi asetus nimesi maakuntakirjastojen erityistehtäviksi kirjallisuuden ja audiovisuaalisten välineiden lainaamisen omista kokoelmista, kaukolainojen välittämisen muille oman alueensa kirjastoille sekä maakuntakokoelman keräämisen. (Mäkinen 2009, 412, 414.)

Maakuntakirjastojen työ on muuttunut sitten 1960- ja 1970-lukujen, varsinkin kaukolainauksen osuus on vähentynyt. Siitä todistaa osaltaan Espoon kaupunginkirjaston esittämät perustelut esityksessä Espoon maakuntakirjaston tehtävistä luopumisesta (Espoon kaupunki 2010). Perusteluissa sanotaan, että nykyisin mistä tahansa kirjastosta pystytään paikantamaan aineisto tosiaikaisesti, kun vielä 1970- ja 1980-luvuilla kirjastot ympäri Suomea tarvitsivat paikantamisessa Espoon kaupunginkirjaston apua ja lähettivät sinne kirjatoiveitaan. Myös virallisten julkaisujen välittäjänä ei maakuntakirjastoa enää tarvita, sillä kaiken aineiston voi löytää verkosta.

## 2.2 Maakuntakokoelmat ja Satakuntakokoelma

Kirjastoasetus antaa maakuntakirjastoille veloitteen ”kehittää toiminta-alueita koskevaa tietopalvelua” (Kirjastoasetus 18.12.1998/1078). Maakuntakirjastot keräävätkin paikallisia maakuntakokoelmia, Porissa kokoelma on nimeltään Satakuntakokoelma. Satakuntaan kuuluu tällä hetkellä (vuonna 2011) 21 kuntaa: Eura, Eurajoki, Harjavalta, Honkajoki, Huittinen, Jämijärvi, Kankaanpää,

Karvia, Kiikoinen, Kokemäki, Köyliö, Lavia, Luvia, Merikarvia, Nakkila, Pomarkku, Pori, Rauma, Siikainen, Säskylä ja Ulvila (Satakuntaliitto 2011). Monet satakuntalaiset kirjastot keräävät lisäksi omaan paikkakuntaan keskittyntä kotiseutu- tai paikalliskokoelmaa, esimerkiksi Rauma, Harjavalta, Huittinen, Lavia ja Pomarkku (Rauman kaupunginkirjasto 2011; Harjavallan kaupunginkirjasto 2011; Huittisten kaupunginkirjasto 2011; Lavian kunnankirjasto 2011; Pomarkun kunnankirjasto 2011).

Satakuntakokoelmaa on kerätty Porissa jo 1930-luvulta alkaen (Hursti 1993, 3). Aluksi kirjamäärät olivat pieniä, esimerkiksi vuoden 1945 lopussa kokoelmaan kuului 173 kirjaa, kun vertailun vuoksi kirjaston yleisellä osastolla oli samaan aikaan lähes 26 000 kirjaa (Porin kaupunginkirjasto 1946, 15). Vuoden 1960 lopussa kokoelmaan kuului 890 kirjaa (yleisellä lainausosastolla 44 702 kirjaa) ja vuoden 1970 lopussa 1485 kirjaa (yleisellä lainausosastolla 75 232 kirjaa) (Porin kaupunginkirjasto 1961, 7; Porin kaupunginkirjasto 1971, 4).

Maakuntakokoelmien keräämistä varten on laadittu yhteiset ohjeet, sillä alkujaan maakuntakokoelmia on kerätty vaihtelevasti. Yhteisiä ohjeita pidettiin tärkeänä myös siksi, että vältetään päällekkäisyyksiltä ja saadaan aineisto järkevästi rajatuksi. Ohjeet on laadittu vuonna 1975 maakuntakirjastoasiain neuvottelukokouksessa asetetun työryhmän ehdotuksen perusteella. (Helsingin kaupunginkirjasto – yleisten kirjastojen keskuskirjasto 1990, 8–9.) Esimerkiksi Satakuntakokoelman keräämisessä ohjeet ovat yhä edelleen lähes sellaisenaan käytössä. Satakuntakokoelman keruuperiaatteet ovat seuraavat:

#### 1. Kaunokirjallisuus

a) joka käsittelee keruualuetta ts. tapahtumat on sijoitettu sinne.

b) joka on alueen kirjailijoiden kirjoittamaa. Kotiseutukirjailijana pidetään kirjailijaa, joka on syntynyt ja sen lisäksi viettänyt siellä huomattavan osan elämästään tai on viettänyt alueella merkittävän osan elämästään. Kirjailijan tuotannosta voidaan hankkia vain Satakunnassa asutun ajan tuotanto.

#### 2. Tietokirjallisuus

a) joka käsittelee alueen historiaa, luontoa, kulttuurielämää, talouselämää jne.

b) joka käsittelee alueen henkilöitä

### 3. Lehdet

jotka ilmestyvät maakuntakirjaston toiminta-alueella. Sanomalehdet on jaettu siten, että ilmestymispaikan kirjasto säilyttää.

### 4. Kartat, jotka koskevat aluetta (sekä maa- että merikartat)

### 5. Pienpainatteet lukuun ottamatta mainoksia tms.

### 6. Opinnäytteet, mikäli ne koskevat aluetta.

### 7. Av-aineisto

#### a) kotiseutuhenkilöiden tuottama musiikki

b) dioja, valokuvia ja julisteita näytteenomaisesti alueen nähtävyyksistä, merkkihenkilöistä, tapahtumista yms.

c) alueen kirkonarkistojen mikrofilmit yms. mikrofilmattua aineistoa (mm. alueen sanomalehdet)

### 8. Alueeseen liittyvät lehtileikkeet

(Porin kaupunginkirjasto – Satakunnan maakuntakirjasto 2011a.) Erona työryhmän tekemään ohjeistukseen on suhtautuminen käsikirjoituksiin: Satakuntakokoelman keruuperiaatteissa ei mainita käsikirjoituksia lainkaan, mutta työryhmän ohjeissa käsikirjoitukset ovat saaneet oman kohtansa. Sen mukaan kokoelmaan kerätään kotiseutukirjailijoiden ja alueen pöytälaatikkokirjailijoiden käsikirjoituksia sekä kotiseutukirjailijoiden ja kotiseutuhenkilöiden kirjeitä. Myös opinnäytteet lasketaan työryhmän ohjeissa käsikirjoituksiin, kun taas Satakuntakokoelman periaatteissa niille on oma kohtansa.

Ensimmäisessä maakuntakirjastoja ja -kokoelmia koskevassa kirjastoasetuksessa eli vuoden 1961 kirjastoasetuksessa korostettiin maakuntakokoelman aineiston säilyttämistä. Seuraavassa, vuoden 1986 kirjastoasetuksessa puhutaan jo aineiston hankkimisesta ja säilyttämisestä alueellista tietopalvelua varten. (Helsingin kaupunginkirjasto – yleisten kirjastojen keskuskirjasto 1990, 7–8.) Myös Jukka Peltonen (1987, 72–73) korostaa artikkelissaan aluekokoelmien merkitystä paikallisen tiedon keskuksena, ei vain kirjamuseoina, vaan aineiston sisältämä tieto on saatettava

palvelemaan sitä tarvitsevia. Maakuntakirjastojen tehtävät voivat kuitenkin muuttua tarvittaessa, sillä tehtäviä on arvioitava uudelleen, jotta ne vastaisivat kirjastojen muuttuvia tarpeita (Opetus- ja kulttuuriministeriö 2011b). Koko kirjastokentän muutoksen ja kehityksen vuoksi nykyään onkin vaikeaa erottaa toisistaan maakuntakirjastotoimintaa ja muuta yleistä kirjastotoimintaa (Porin kaupunginkirjasto – Satakunnan maakuntakirjasto 2011b, 2).

### 3 TIEDONHAKU INTERNETISTÄ

Nykyään yhä suurempi osa kaikesta tuotetusta tiedosta on löydettävissä internetistä. Tässä luvussa tarkastellaan eri tapoja etsiä tietoa verkosta.

#### 3.1 Tiedonhaun prosessi

Tiedonhakua kuvataan usein prosessina. Prosessiin vaikuttavat esimerkiksi tiedonhaun aihe, tiedonhakijan taidot, tiedonhakijan taustatiedot aiheesta sekä tiedonhakumahdollisuudet, jotka ovat tiedonhaussa käytettävissä. Prosessi voidaan jakaa karkeasti kahteen erilaiseen osaan riippuen siitä, haluaako tiedonhakija nopeasti jonkin yksittäisen tiedon vai haluaako hän tehdä systemaattisen tiedonhaun ja perehtyä syvällisesti valitsemaansa aiheeseen. (Heikkinen ym. 2005, 23.)

Haasio (2007, 13) korostaa, että tiedonhaku on suunnitelmallista toimintaa. Tiedonhaun prosessi alkaakin tiedontarpeen pohtimisella. Tiedonhakijan pitää miettiä vastauksia muutamiin kysymyksiin, jotka auttavat tiedontarpeen määrittämisessä. Näitä ovat esimerkiksi seuraavat:

Mikä on tiedonhaun aihe?

Mikä on näkökulma, josta aihetta lähestytään?

Mitä ennakkotietoja tiedonhakijalla on aiheesta?

Tarvitaanko aiheesta yleistietoa vai tieteellistä tietoa?

Kuinka vanhaa tai uutta tietoa tarvitaan?

Mihin tarkoitukseen tietoa tarvitaan?

Kuinka suppeasti tai kattavasti tietoa tarvitaan?

(Alaterä & Halttunen 2002, 84, Heikkinen ym. 2005, 24.)

Nopealla tiedonhaulla haetaan esimerkiksi yksittäistä teosta tai faktatietoa, uusinta tietoa jostakin aiheesta tai yleistietoa jostakin alasta ja sen terminologiasta. Tällöin tiedonhaun prosessi alkaa aiheen määrittelyllä eli päätetään se aihe tai asia, josta tarvitaan tietoa. Seuraavaksi siirrytään suoraan hakuun. Haku voidaan tehdä esimerkiksi internetin hakukoneella, mutta myös

internetin uutispalveluista ja aihehakemistoista tai perinteisemmin sanomalehdistä, aikakauslehdistä ja uutuuskirjoista. (Heikkinen ym. 2005, 23–24.)

Systemaattista tiedonhakuja tarvitaan esimerkiksi tehtäessä opinnäytetyötä tai tutkielmaa eli silloin, kun tarvitaan tieteellistä tietoa. Tällöin hakuprosessi alkaa valmistelevalle vaiheella, joka sisältää aiheeseen tutustumisen, hakusanojen miettimisen, hakusanojen valinnan esimerkiksi asiasanastoista sekä hakulausekkeiden muodostamisen. Aiheeseen voi tutustua vaikkapa sanakirjojen, hakuteosten, käsikirjojen ja sanakirjojen avulla, jolloin saa tuntumaa aiheen terminologiaan ja käsitteisiin, joita voi käyttää hakusanoina. Kun hakusanoja on mietitty, valitaan haun kohteena olevien tietokantojen asiasanastoista ja tesauksista hakuun sopivat hakusanat. Hakulausekkeita muodostetaan valittuja hakusanoja yhdistelemällä, vaikka toki haku voidaan suorittaa vain yhdelläkin hakusanalla. Valmistelevalle vaiheen jälkeen tehdään varsinainen haku. Hakea voidaan esimerkiksi kirjastojen kokoelmatietokannoista, verkkolehdistä, aihehakemistoista ja eri tieteenalakohtaisista tietokannoista. (Heikkinen ym. 2005, 23–31.)

Teki hakua sitten nopeasti tai laajemmin systemaattisesti, on hakutulosta tärkeää arvioida. Käytettyjä hakusanoja tai hakulausekettä voi olla tarpeen muuttaa, jos hakutuloksia ei ole tarpeeksi tai jos niitä on liikaa. Myös käytettyä tietokantaa tai muuta tiedonlähdeä voi joutua vaihtamaan. (Heikkinen ym. 2005, 31.) Jos hakutuloksiksi saatuja dokumentteja on liikaa, on syytä tarkentaa hakusanoja ja rajata hakuja enemmän esimerkiksi ajallisesti tai tiettyyn kieleen. Jos taas hakutuloksia on liian vähän tai ei ollenkaan, voi olla aiheellista vaihtaa hakusanoja, katkaista niitä tai vähentää hakusanojen määrää. (Haasio 2007, 16.)

### 3.2 Internet tiedon kanavana

Internet on maailmanlaajuinen tietoverkko. (Heikkinen ym. 2005, 99). Se on valtava tietovarasto, josta etsitään tietoa nykyään yhä enemmän ja enemmän. Internet ei kuitenkaan ole tietolähde, sillä se ei ole varsinainen lähde, vaan

enemmänkin kanava, jonka kautta haluttu tieto löydetään. Internetin kautta löydetyt dokumentit puolestaan ovat tiedonlähteitä. (Haasio 2009b, 85.) Iso osa internetistä löytyvästä tiedosta on www-sivuja ja niillä olevia tietoja, mutta verkosta löytyy myös esimerkiksi pdf-muotoisia dokumentteja, ääni- ja videotiedostoja sekä tietokoneohjelmia (Heikkinen ym. 2005, 100).

Alaterä & Halttunen (2002, 111) erottavat verkkotiedonhausta kaksi eri tapaa:

Tietoa haetaan perinteisistä tietokannoista, luetteloista ja tekstikannoista internetin kautta.

Tietoa haetaan internetistä itsestään eli etsitään, mitä palveluita, dokumentteja ja resursseja on saatavilla.

Verkon kautta löytyneen tiedon ominaispiirteitä ovat esimerkiksi sen ajantasaisuus, jopa reaaliaikaisuus, ja globaalius. Globaali tieto on löydettävissä ja käytettävissä ajasta ja paikasta riippumatta ympäri maailman. Siinä missä painettu lähde voi olla jo ilmestyessään vanhentunut, on verkkotieto yleensä päivitettyä ja ajantasaista. (Haasio 2009b, 86). Toisaalta verkossa on paljon myös vanhentuneita sivuja. Tiedon laatu myös vaihtelee, sillä verkossa on helppo levittää monenlaista tietoa. (Heikkinen ym. 2005, 100.)

Internetistä löytyy paljon sellaista tietoa, joka on kaikille käyttäjille avointa, mutta monet tietokannat ja julkaisukokoelmat vaativat käyttäjätunnuksen ja salasanan (Heikkinen ym. 2005, 99). Tämä salattu sisältö ei yleensä löydy hakukoneilla, minkä vuoksi sitä kutsutaankin syväksi webiksi (*deep web*) tai näkymättömäksi webiksi (*invisible web*) (Haasio 2009b, 86).

Internetin käytön hakuvälineenä tekee haastavaksi sen jatkuva muuttuminen. Internet kasvaa ja muuttuu koko ajan, ja samanlaisena tehdyn tiedonhaun tulokset vaihtelevat päivittäin. Jotkin tuloksiksi tulleet sivut saattavat kadota myöhemmin tai niiden osoitteet tai sisällöt saattavat muuttua. Toisaalta samaa tietoa voi löytyä useasta eri paikasta, sillä digitaalisessa muodossa olevan aineiston kopioiminen on helppoa. (Heikkinen ym. 2005, 99–100.)



### 3.3 Hakukoneet

Koska internet on avoin verkko, jota ei ole alun perin luotu tiedonhakuun varten, tarvitaan verkossa olevan tiedonhakuun apua (Alaterä & Halttunen 2002, 16). Tähän tehtävään on tarkoitettu erilaiset hakukoneet.

Hakukoneet toimivat siten, että niillä on hakurobotteja, jotka keräävät tietoa verkossa olevista dokumenteista ja niiden sisällöstä joko kaikkialta internetistä tai tietyistä ennalta määrätystä aiheesta tai tietyltä alueelta. Näistä kerätyistä tiedosta muodostetaan tietokanta, jonka sisällöstä hakukoneen käyttäjät voivat tehdä hakuja. Hakurobotit voivat indeksoida kustakin dokumentista koko sen tekstin, mutta yleisimpiä sanoja ei indeksoida. Sen sijaan indeksoinnissa painotetaan esimerkiksi otsikkotietoja, useimmin esiintyviä sanoja ja fraaseja sekä kuvailevaa metadataa. (Heikkinen ym. 2005, 103.)

#### 3.3.1 Sanahakukone Google

Google on suomalaisten eniten käyttämä hakukone, sillä arviolta 90 prosenttia kaikista suomalaisten tekemistä hauista tehdään Googlella. Muualla maailmassa Googlen osuus on arviolta 50 prosenttia. (Haasio 2009a, 46.)

Googlen toiminta perustuu vuonna 1998 kehitettyyn Pagerank-algoritmiin, joka analysoi toisiinsa linkitettyjä dokumentteja ja antaa niille linkityksen perusteella eri painoarvoja. Mitä enemmän sivulle linkitetään, sitä suuremman painoarvon se saa. Algoritmin idea pohjautuu jo 1950-luvulle, Eugene Garfieldin Pennsylvanian yliopistossa kehittämään sitaattianalyysiin. Sen perusajatus on se, että laadukkaita julkaisuja lainataan huonolaatuisia enemmän. (Tähtinen 2008, 40.)

Google käyttää hauissaan toki muitakin algoritmeja, mutta niistä ei juurikaan löydy tietoa, sillä ne ovat yrityssalaisuuksia. Oletettavasti sillä on kuitenkin käytössä myös sivujen sisältöä analysoivia työkaluja, joiden avulla hakutuloksiksi saadaan relevantteja sivuja. (Tähtinen 2008, 41.)

Yleisesti Googlea käytetään hakemalla tietoa sen perushaulla, joka löytyy osoitteesta <http://www.google.fi/>. Tällainen nk. likainen haku on yleensä riittävä silloin, kun halutaan nopeasti hakea jokin tieto eikä hakusanoja ole kuin pari. Googlen suomenkielisessä versiossa on kuitenkin käytettävissä myös tarkennettu haku ja sen lisäksi kuvahaku, karttahaku ja esimerkiksi blogihaku. (Haasio 2009, 46–47.)

Suomeksi löytyy nykyisin myös Google Scholar (osoitteessa <http://scholar.google.fi/>), jonka avulla voi hakea tieteellistä kirjallisuutta, esimerkiksi yliopistojen, akateemisten julkaisijoiden tai muiden tieteellisten järjestöjen tekemiä tutkielmia, esitelmiä, kirjoja tai artikkeleita (Google 2011b). Google Scholaria on hyvä käyttää silloin, kun tarvitaan tieteellisiä artikkeleita, mutta ei ole mahdollisuutta käyttää maksullisia tietokantoja (Haasio 2009, 59). Google Books (osoitteessa <http://books.google.com/>) puolestaan hakee tietoa digitoituista kirjoista. Kirja voi olla luettavissa verkon kautta kokonaan tai sitä voi katsella rajoitetusti vain katkelmina. Joistakin kirjoista saa nähtäväksi vain teoksen tiedot riippuen sen tekijänoikeuksista. (Google 2011c.)

Googllella on Alerts-palvelu, josta on olemassa myös suomenkielinen Ilmoitukset-beta-versio. Palvelun idea on ilmoittaa käyttäjälle uusista häntä kiinnostavista hakutuloksista käyttäjän valitsemasta aiheesta. Käyttäjä siis luo jonkin hakulausekkeen, jolla Googlen hakukoneet hakevat tietoa, ja aina kun uusi hakua vastaava sivu löytyy, palvelu ilmoittaa siitä käyttäjälle. Ilmoituksen uudesta sivusta saa joko heti sen löydyttyä tai kootusti kerran päivässä tai kerran viikossa. (Google 2011a.)

### 3.3.2 Muita sanahakukoneita

Yahoo-hakukoneen suomenkielinen käyttöliittymä löytyy osoitteesta <http://fi.search.yahoo.com/>. Yahoon historia ulottuu vuoteen 1994, jolloin sen toiminta alkoi järjestettynä aihehakemistona. Vuonna 2002 Yahoo muuttui hakukoneeksi, mutta hakupalvelut se osti Googelta vuoteen 2004 asti. Nykyisin Yahoo käyttää omaa hakutekniikkaansa. Yritystosten kautta Yagoon piiriin

kuuluu myös muita, ennen itsenäisinä toimineita hakukoneita, kuten AllTheWeb ja Altavista. (Search Engine Watch 2007.)

Yahoota voi käyttää samalla tavalla kuin muitakin hakukoneita tekemällä sanahaun, joka hakee tietoa kaikkialta internetistä. Yahooolla on edelleen käytössä myös alkuperäinen hakemisto, josta tietoa voi hakea. Hakemiston ovat koonneet käsin Yagoon työntekijät, ja se on järjestetty eri kategorioihin ja alakategorioihin. Hakemistoa voi siis käyttää hakemiseen silloin, kun ei aivan täysin tunne haettavaa asiaa tai halua ymmärtää hakemaansa asiaa paremmin. (Yahoo 2011a.)

Internet- ja sanahaun lisäksi Yagoon avulla voi hakea kuvia ja uutisia. Laajennettu haku mahdollistaa hakutulosten rajaamisen esimerkiksi tiedostomuodon, sivuston iän ja kielen perusteella.

Ask-hakukoneella ei ole erillistä suomenkielistä käyttöliittymää, mutta se löytyy englanninkielisenä osoitteesta <http://www.ask.com/>. Ask-hakukone aloitti toimintansa 1990-luvulla nimellä Ask Jeeves, ja se tuli tunnetuksi hakukoneena, jonka kautta voi kysyä kysymyksiä ja saada niihin vastauksia. Tämä palvelu toteutettiin kuitenkin pääsääntöisesti ilman tekniikkaa, ja suosituimpiin kysymyksiin etsivät vastauksia palvelun työntekijät. (Search Engine Watch 2007.)

Tehdyn haun tuloksista Ask-hakukone nostaa esiin hakutulossivun ylälaitaan yhden tuloksen, jonka se arvelee olevan haun kannalta relevantein. Hakukone pyrkii esittämään vastauksen hakijan kysymykseen suoraan hakutulossivulla ilman, että hakijan pitää erikseen siirtyä löydetylle sivustolle. (Ask 2011.) Ask-hakukoneella voi hakea myös esimerkiksi kuvia, uutisia, videoita ja karttoja. Hakukone mahdollistaa myös tarkennetun haun, jolloin hakua voi rajata esimerkiksi hakutermin sivustolla sijainnin, kielen tai maan perusteella.

### 3.3.3 Metahakukoneet

Kuten jo aikaisemmin todettiin, sanahakukoneet keräävät internetistä tietoa omaan tietokantaansa, johon hakukoneella tehdyt haut kohdistuvat. Haut eivät

siis kohdistu suoraan verkkoon. Metahakukoneet eli monihakukoneet ovat hakukoneita, jotka etsivät tietoa samanaikaisesti monen eri hakukoneen ja aihehakemiston tietokannoista, mutta omaa tietokantaa niillä ei ole. Metahakukoneet saavat siis kattavampia hakutuloksia, kun niiden haut kohdistuvat moneen eri tietokantaan. (Haasio 2009a, 68.) Ne säästävät myös hakijan aikaa, sillä eri sanahakukoneilla haettaessa saadaan erilaisia tuloksia samoilla hakusanoilla, mutta metahakukonetta käytettäessä hakua ei tarvitse suorittaa moneen kertaan, vaan yksi kerta riittää (Haasio 2008, 81).

Metahakukoneet suodattavat saamiaan hakutuloksia eri algoritmeilla niin, että tulosten määrä pienenee ja laatu kasvaa. Siinä missä sanahakukoneet, esimerkiksi Google, saattavat antaa tuhansia tai jopa miljoonia hakutuloksia, on metahakukoneiden tulosjoukko joitakin kymmeniä tai satoja. Hakija pystyy siis käymään läpi kaikkikin tulokset, toisin kuin sanahakukoneiden tuloksia. (Haasio 2009a, 68.)

Metahakukone Dogpile löytyy osoitteesta <http://www.dogpile.com/>. Se hyödyntää hauissaan Google-, Yahoo- ja Bing-nimisiä hakukoneita (Dogpile 2011). Testihaku hakusanalla *pitkäaikaissäilytys* tuotti Dogpile-haussa noin 300 tulosta, kun samalla sanalla Googlessa haettuna tulokseksi tuli 33 400 tulosta. Haku tehtiin 13.10.2011 molempien hakukoneiden perushaulla. Dogpile siis karsii tuloksia. Jokaisen löytyneen tuloksen yhteydessä se näyttää, minkä hakukoneen tai hakukoneiden tietokannasta kyseinen tulos on löytynyt.

Dogpilella on mahdollista perussanahaun lisäksi kohdistaa haku erityisesti esimerkiksi kuviin, videoihin ja uutisiin. Dogpilen uutishaku on suppeampi kuin esimerkiksi Googlen uutishaku (Haasio 2009a, 69). Lisäksi Dogpilesta löytyy tarkennettu haku, jossa hakua pääsee rajaamaan tarkemmin erilaisilla hakulauseilla sekä kielen ja sen palvelimen mukaan, jolla aineisto sijaitsee.

Metahakukone Yippy, joka aikaisemmin on tunnettu myös nimillä Vivísimo ja Clusty, on löydettävissä osoitteessa <http://search.yippy.com/>. Yippy on klusteroiva hakukone, mikä tarkoittaa sitä, että se lajittelee hakutuloksia ryhmiin automaattisesti aihepiirien mukaan. (Haasio 2009a, 69.) Hakukone itse kutsuu

ryhmiä pilviksi. Sen mukaan pilvet auttavat nostamaan esille sellaisia hakutuloksia, jotka muuten jäisivät tulosjoukon häntäpäähän sekä näkemään hakutulosten välillä yhteyksiä, joita ei välttämättä ilman pilviä näkisi. (Yippy 2011.)

### 3.4 Aihehakemistot

Aiuehakemistot ovat linkkilistoja, joissa linkit on järjestetty aihealueittain (Haasio 2008, 91). Aiuehakemistoja selaamalla on mahdollista saada hyvä yleiskuva jonkin tietyn aiheen keskeisistä verkkosivustoista (Haasio 2009a, 73). Ne soveltuvatkin parhaiten orientoivan tiedon hankintaan eli kun halutaan tietää, mitä verkosta on löydettävissä jostakin tietystä aiheesta, esimerkiksi koirista ja niiden hankinnasta. Aiuehakemistoissa on myös listattuna syvään webiin eli näkymättömään webiin kuuluvia verkkosivuja, esimerkiksi tietokantoja, joihin hakukoneet eivät pääse käsiksi. (Haasio 2008, 91.) Hyviä hakemistot ovat myös silloin, jos alan terminologia on outoa eikä tiedä, mitä hakusanoja haussa käyttäisi. Lisäksi hakemistojen sisältö on valikoitua, joten löydettyjä verkkosivustoja voi pitää kohtuullisen luotettavina. Toisaalta kaikkein uusimman sivut eivät heti päädy hakemiston listoille, vaan ne voi löytää hakukoneella. (Haasio 2008, 89–90).

Aiuehakemistoista voidaan erottaa kaksi pääryhmää, yleishakemistot ja erikoishakemistot. Yleishakemistoissa on linkkejä monista erilaisista aihepiireistä mahdollisimman kattavasti. Niistä on hyvä etsiä tietoa arkipäiväisiin asioihin. Erikoishakemistot ovat puolestaan keskittyneet joko yhteen tai useampaan erikoisalaan, ja ne ovat hyödyllisiä etsittäessä tietoa kyseisestä aihealueesta tieteellisiin tarkoituksiin, mutta myös yleistietoa. Aiuehakemistoissa on yleensä sisäinen hakukone, joten tietoa voi hakea sen avulla, tai sitten selailemalla kategorioita ja alakategorioita. (Haasio 2008, 91.)

Suomalaisia aihehakemistoja ovat Makupalat (osoitteessa <http://www.makupalat.fi/>) sekä Kirjastot.fi-palvelun Linkkikirjasto (osoitteessa <http://www.kirjastot.fi/linkkikirjasto>). Nämä palvelut tulevan yhdistymään, mutta uuden palvelun julkaisuajankohtaa ei tällä hetkellä (lokakuussa 2011) vielä

tiedetä. Makupaloja on koottu jo vuodesta 1995 lähtien, ja se sisältääkin yhteensä noin 30 000 linkkiä, joiden sisältöä on myös kuvailtu. Makupalojen toimittamisesta vastaa Hämeenlinnan kaupunginkirjasto. (Makupalat 2011b.)

Makupalojen linkit on jaoteltu kahdeksaan kategoriaan: vapaa-aika, koti, ympäristö, yhteiskunta, viestintä, tekniikka, taiteet sekä eri tieteenalat. Nämä kategoriat on edelleen jaoteltu vielä alakategorioihin. (Makupalat 2011a.) Kirjastot.fi-palvelun Linkkikirjasto puolestaan on jaoteltu Yleisten kirjastojen luokitusjärjestelmän mukaan. Luokittelussa on käytetty sekä pääluokkia että niiden alaluokkia. Lisäksi linkeistä löytyvät myös luokitus- ja kuvailutiedot. (Kirjastot.fi 2011a.)

### 3.5 Viranomaistieto

Nykyään sekä valtion viranomaiset että kaikki kunnat tarjoavat tietoa ja palvelujaan verkossa. Tämä parantaa kansalaisten mahdollisuuksia toteuttaa demokraattisia arvoja. Verkosta löytyy niin pöytäkirjoja ja esityslistoja kuin lomakkeita ja neuvontaakin. Osa kaupungeista ja kunnista tarjoaa myös sähköisiä asiointipalveluja, esimerkiksi terveyskeskusajan voi varata tai lapselle hakea päivähoitopaikkaa verkossa. (Haasio 2008, 228.)

Keskitetysti viranomaisten tuottamaa tietoa ja palveluita on koottu Suomi.fi-portaaliin. Se löytyy osoitteesta <http://www.suomi.fi/>. Portaalia toimittaa Valtiokonttori. Portaalissa on mahdollista asioida verkossa, tulostaa ja täyttää lomakkeita, etsiä tietoa palveluista ja laeista sekä lukea uutisia näistä aiheista. Portaalin piiriin kuuluvia palveluita tuottavat julkishallinnon organisaatiot, kunnat ja niiden toimintaa täydentävät järjestöt. Lisäksi portaaliin sisältyy palvelukartta, josta näkee julkishallinnon palvelupisteiden sijainnin sekä kansalaisen asiointitili, jonka kautta voi vastaanottaa viranomaisten lähettämää postia. (Suomi.fi 2011a.)

Viranomaistietoa löytyy myös kuntien ja kaupunkien omilta verkkosivuilta, ja sellaiset jokaiselta kunnalta ja kaupungilta nykyään löytyykin. Verkkosivuille pääsee kätevästi esimerkiksi Suomi.fi-portaalin Valtio ja kunnat osion Kunnat ja

maakunnat alaosiosta, johon on listattu maakunnittain linkkejä kuntien ja kaupunkien verkkosivuille (Suomi.fi 2011b). Kuntien ja kaupunkien verkkosivuja löytää usein myös loogisesti päättelemällä, esimerkiksi Porin kaupungin verkkosivujen osoite on <http://www.pori.fi/> tai Turun kaupungin <http://www.turku.fi/>. Logiikassa on luonnollisesti myös poikkeuksia, esimerkiksi <http://www.nokia.fi/> viekin Nokia-nimisen yrityksen verkkosivuille. (Heikkinen ym. 2005, 112.)

Kuntien ja kaupunkien verkkosivuilta löytyy kaikenlaista tietoa esimerkiksi kunnan tai kaupungin ajankohtaisista asioista, päätöksenteosta tai palveluista. Esimerkiksi Porin kaupungin verkkosivulla on mm. ajankohtaisia uutisia kaupungin asioista, tapahtumakalenteri, palveluhakemisto ja kysymystori, jossa kaupunkilaiset voivat esittää kysymyksiä Porin kaupunkia koskevista asioista. Lisäksi jokaisella kaupungin hallintokunnalla, virastolla ja liikelaitoksella on sivustolla oma osionsa. (Porin kaupunki 2011a.)

### 3.6 Uuden tiedon automaattinen seuranta

Kuten edellä on kerrottu, tarvitsemansa tiedon luokse pääsee eri teitä hakukoneita ja hakemistoja käyttämällä. Niiden käyttö vaatii kuitenkin käyttäjältään aktiivisuutta, sillä jos haluaa pysyä jostakin tietystä asiasta ajan tasalla, on tasaisin väliajoin tehtävä haku hakukoneella tai käytävä selaamassa aiheeseen liittyvät linkit hakemistossa. On myös olemassa tapoja, joilla verkkoon ilmestyvästä kiinnostavasta uudesta tiedosta voi saada itselleen ilmoituksen eikä erillisiä hakuja tarvitse enää käydä tekemässä.

RSS (lyhenne sanoista *Rich Site Summary* tai *Really Simple Syndication*) on jatkuvasti päivittyvä syöte, joka kertoo, kun uutta sisältöä on päivitetty seuratulle verkkosivustolle. Käyttäjä voi tilata syötteen joko erilliseen RSS-ohjelmaan, verkkoselaimeensa tai erilliselle syötesivulle. (Haasio 2009a, 108.) RSS-syötteen avulla käyttäjä voi saada tietoa esimerkiksi uusista uutis- ja blogiotsikoista, mutta myös lehdistö- ja pörssitiedotteista, tuote- ja versiointiuutuuksista sekä vaikkapa neuvotteluista tai valokuvista. RSS-syötteestä on hyötyä sekä tiedon tuottajalle että sen käyttäjälle: tuottaja saa

levitettyä tietoa tehokkaasti, kun taas käyttäjä saa haluamaansa tietoa suoraan omalle koneelleen, eikä aikaa kulu eri verkkopalveluita kierrellessä. (Hintikka 2007, 26.)

RSS-syötteet eivät ole syötteiden ainoa formaatti, vaan syötteitä on mahdollista tilata myös Atom-muodossa. Yhteistä formaateille on se, että sivun tuottajan on mahdollistettava syötteiden tilaaminen. (AtomEnabled 2007.) Syötteen mahdollistaneet sivustot tunnistaa siitä kertovasta logosta (23 asiaa 2009a).

On olemassa myös verkkosivustoja, jotka koostavat uudesta tiedosta, esimerkiksi uutisista, valmiita kokonaisuuksia. Tällaiset palvelut hyödyntävät eri uutistarjoajia ja koostavat näiden tarjonnasta kokoelman omalle verkkosivustolleen. Maksuton, kotimainen uutispalvelu on esimerkiksi Ampparit, joka löytyy osoitteesta <http://www.ampparit.com/>. Ampparit seuraa yli kahta sataa lähdettä, joista se poimii uutiset omalle sivustolleen. Lähteiden joukossa on kotimaisia sanomalehtiä, aikakaus- ja harrastelehtiä, uutissivustoja ja muita vastaavia lähteitä (Ampparit 2011).

Ampparit-sivuston uutiset on jaettu useaan kategoriaan ja nämä edelleen alakategorioihin. Etusivulla on esillä uusimmat uutiset, ja se päivittyy jatkuvasti. Ampparit-sivustolle on mahdollista luoda oma profiili, jonka avulla uutisnäkyä voi personoida haluamakseen. Personoinnin avulla käyttäjä voi seurata uutisia vain tietyistä kategorioista tai medioista sekä suodattaa pois tiettyä aihetta käsitteleviä uutisia. Ampparit-palvelusta käyttäjä voi myös tilata RSS-syötteen, joka voi koskea kaikkia uutisia tai vain tiettyä kategoriaa. (Haasio 2008, 239–240.)

Uutta tietoa voi automaattisesti seurata myös Googlen Alert-palvelulla, jota käsiteltiin jo luvussa 3.3.1. Kun RSS-syötettä varten seurattavan sivuston tarvitsee mahdollistaa syötteet, voi Google Alertin avulla seurata minkälaisia sivustoja tahansa. Google Alertin ja RSS-syötteet yhdistävä kombinaatio on myös mahdollinen, sillä Google Alertin löytämät tulokset saa tilattua itselleen myös RSS-syötteenä. (Google 2011d.)



## 4 DIGITAALINEN PITKÄAIKAISSÄILYTYS

Digitaalisen aineiston pitkäaikaissäilytyksellä tarkoitetaan toimia, jotka tähtäävät siihen digitaalisessa muodossa olevien aineistojen käytettävyyteen vielä tulevaisuudessakin. Aineiston käytettävissä olemisen lisäksi aineistoon pitää olla myös pääsy, vaikka jotkut kustantajat ja julkaisijat mielellään rajoittaisivatkin julkaisemiinsa aineistoihin pääsyä. Kuitenkaan mikään aineisto ei ole toden teolla käytettävissä, jos aineistoon ei ole pääsyä, joten nämä molemmat näkökannat tulee ottaa huomioon suunniteltaessa pitkäaikaissäilytystä. (Caplan 2008a, 7.)

Toisinaan termejä digitaalinen pitkäaikaissäilytys (*digital preservation*) ja digitointi säilytyksen vuoksi (*digitization for preservation*) erehdytään sekoittamaan toisiinsa. Vanhoja ja hauraita kirjoja ja lehtiä on viime vuosina digitoitu, jotta alkuperäisiä aineistoja ei tarvitsisi enää käyttää, mutta niiden sisältämä informaatio olisi silti tutkijoiden käytössä. Digitointiprosessi tuottaa siis digitaalisessa muodossa olevaa aineistoa, jota myös tarvitsee säilyttää pitkäaikaisesti. (Caplan 2008a, 7.) Digitaalista pitkäaikaissäilytystä voidaan siis pitää ikään kuin yläkäsitteenä, joka kattaa myös digitoidun aineiston pitkäaikaisen säilytyksen.

### 4.1 Pitkäaikaissäilytyksen tavoitteet

Kuten edellä määriteltiin, digitaalinen pitkäaikaissäilytys muodostuu erilaisista toimenpiteistä, joka mahdollistavat aineiston pitkäaikaisen säilyvyyden. Näillä toimenpiteillä on tiettyjä tavoitteita, jotka pyrkivät turvaamaan tiedon säilymisen kaikilta osin. Robin Wendler (2006, 61–64) on listannut viisi säilytystavoitetta: tiedon elinkyky (*viability*), tiedon saaminen esitettäväksi (*renderability*), tiedon ymmärrettävyys (*understandability*), tiedon autenttisuus (*authenticity*) ja tiedon identifiointi (*identification*). Priscilla Caplan (2008a, 7–9) on lisännyt näihin tavoitteisiin vielä kaksi: tiedon saatavuus (*availability*) ja tiedon muuttumattomuus (*fixity*). Seuraavaksi esitellään lyhyesti jokainen seitsemästä tavoitteesta.

#### 4.1.1 Tiedon fyysinen elinkyky

Tiedon fyysisen elinkyvyn säilyttämisellä tarkoitetaan sitä, aineisto pidetään turvassa ja vahingoittumattomana. Elinkykyisyyden turvaaminen on siis yleensä varsin mekaaninen ja automaattisesti hoidettava prosessi, joka koostuu mm. varmuuskopioiden tekemisestä ja tarkistuksista. (Wendler 2006, 61–62.) Aineiston tallentamisessa käytettyä mediaa on siis pystyttävä lukemaan. Aineiston elinkykyisyyttä uhkaavat niin käytetyn median huonokuntoisuus kuin median vanhentuminenkin. Tämän vuoksi tiedostoja pitää varmuuskopioiden lisäksi kopioida uusiin laitteisiin. Varmuuskopioita pitää säilyttää alkuperäisistä tiedostoista erillään. (Caplan 2008a, 8.)

#### 4.1.2 Tiedon saaminen esitettäväksi

Tiedon saamista esitettäväksi voidaan pitää koko pitkäaikaissäilytysprosessin ydinkysymyksenä, muuten koko säilytysprosessi on ollut turhaa (Caplan 2008a, 9). Sillä tarkoitetaan mahdollisuutta avata, soittaa, näyttää, suorittaa tai muuten päästä käsiksi digitaaliseen aineistoon. Vaatimuksena on siis aineiston muodon ja sekä saatavilla olevan laitteiston että ohjelmiston yhteensopivuus. Tämä yhteensopivuus voidaan saavuttaa esimerkiksi muuttamalla aineiston muotoa (ks. luku 4.2.2) tai jäljittelemällä alkuperäistä ohjelmistoympäristöä (ks. luku 4.2.3). (Wendler 2006, 62.) Tiedostoformaattien vanheneminen, tiedostoformaattien korvautuminen jollakin toisella formaatilla tai ohjelmistojen vanheneminen voivat estää tiedostojen käytön (Caplan 2008a, 9).

#### 4.1.3 Tiedon ymmärrettävyys

Jotta aineisto olisi hyödyllistä vielä tulevaisuudessakin, sen on oltava elinkykyisyyden ja esitettävyyden lisäksi merkityksellistä käyttäjälle. Tiedon on siis oltava ymmärrettävää. (Wendler 2006, 63.) Tämän vuoksi aineiston säilytyksestä vastaavan organisaation pitää säilyttää varsinaisen aineiston lisäksi myös metadataa, dokumentaatioita ja muita aineistoja, jotta tulevaisuuden käyttäjä pystyy tulkitsemaan ja ymmärtämään aineiston sisällön (Caplan 2008a, 8). Pitkäaikaissäilytyksen organisoijan pitää silti miettiä, voiko

joissakin tapauksissa luottaa yhteisön tietoon asiasta. Wendler nostaa esimerkiksi saksankielisen tekstin, jonka mukana ei tarvitse tallentaa saksan sanakirjaa, vaan voidaan luottaa siihen, että tulevaisuudessakin on saksan kielen osaajia tulkitsemassa tekstiä. Toisaalta esimerkiksi taulukoiden tulkitsemiseen taas tarvitaan mukana olevaa selvitystä, jotta kaikille riveille ja sarakkeille saadaan merkitys. (Wendler 2006, 63.)

#### 4.1.4 Tiedon autenttisuus

Digitaalisen tiedon pelätään usein muuttuvan säilytyksen aikana, jolloin sen luotettavuus kärsii (Wendler 2006, 63). Tiedon autenttisuudella tarkoitetaan sitä, että sekä lähteen että itse tiedon eheys pystytään todistamaan. Säilytyksen aikana aineistot eivät saa tuhoutua eikä niihin saa tulla muutoksia, joiden alkuperää ei pystytä todistamaan. Kaikkien aineistoihin tehtyjen muutosten tulee näkyä kunkin tiedoston historiassa ja metadatassa. (Caplan 2008a, 8.) Tulevaisuuden käyttäjän on siis nähtävä, onko aineistoa muutettu, ja jos on, miten ja minkä prosessin seurauksena, milloin ja kuka muutokset on tehnyt (Wendler 2006, 63).

#### 4.1.5 Tiedon identifiointi

Edellä kuvatut tavoitteet voidaan toteuttaa tietämättä mitään itse tiedoston sisällöstä, esimerkiksi sen nimestä tai tekijästä. Kuitenkin yleisesti ollaan sitä mieltä, että tietoa täytyy identifioida, jotta sen säilyttämisessä on järkeä. (Wendler 2006, 64.) Tiedon identifioinnilla tarkoitetaan siis aineiston kuvailua tarpeeksi tarkasti, jotta tulevaisuudessa aineiston löytyminen ja käyttö on mahdollista. Aineisto tarvitsee tätä varten metadatan, joka voi joko olla sisäänkirjoitettuna aineistossa itsessään tai se voi olla ulkoista. (Caplan 2008a, 8.)

#### 4.1.6 Tiedon saatavuus

Saatavuudella tarkoitetaan säilytettävän aineiston fyysistä hallinnointia, sillä organisaatio ei voi säilyttää pitkäaikaisesti sellaista aineistoa, johon sillä ei ole

pääsyä. Itse tuottamaansa aineistoon organisaatiolla on yleensä pääsy helpostikin, mutta muiden tuottajien kanssa on hyvä tehdä erilaisia sopimuksia esimerkiksi arkistokappaleiden saamisesta, jotta aineiston saatavuus on turvattu. (Caplan 2008a, 8.)

#### 4.1.7 Tiedon muuttumattomuus

Tiedon muuttumattomuudella tarkoitetaan sitä, että pitkäaikaissäilytysjärjestelmän pitää suojata digitaalista aineistoa muutoksilta. Ohjelmistoteollisuuden tuottamat turvallisuusohjelmistot ovat tähän tarkoitukseen hyviä, sillä niihin sisältyy virustorjunta, palomuuuri, käyttäjien tunnistaminen, tunkeutujien havaitseminen sekä välitön puuttuminen mahdollisiin häilytyksiin. Toisaalta myös fyysisen varastoinnin pitää olla kunnossa, esimerkiksi varastotilojen ilmastointiin pitää kiinnittää huomiota ja tietoja pitää kopioida silloin tällöin uusiin laitteisiin, jotta laitteiden vanhentuminen ei vahingoita tiedostoja. (Caplan 2008a, 8.)

#### 4.2 Digitaalisen pitkäaikaissäilytyksen menetelmiä

Digitaalisen aineiston pitkäaikaissäilytykseen ei ole vielä löydetty yhtä toimivaa ja laajasti hyväksyttyä menetelmää, mutta erilaisia ehdotuksia ja ajatuksia on esitetty. Aineiston tulostamisen paperille lienevät jo kaikki hylänneet, vaikka luotettavuudessaan ja säilyvyydessään se onkin omaa luokkaansa. Tulostaminen kuitenkin soveltuu vain tekstille ja stillkuvalle, joten paljon erilaista aineistoa jäisi tämän säilytysmenetelmän ulkopuolelle. (Hakala & Stenvall 2002, 84–85.) Digitaalisen aineiston pitkäaikaissäilytykseen liittyy paljon riskejä, joista osa on vielä täysin tuntemattomia (Kansallinen digitaalinen kirjasto 2010, 56). On kuitenkin joitakin teorioita menetelmiksi, joita voitaisiin hyödyntää. Tässä luvussa esittelen niistä kolme.

##### 4.2.1 Kopiointi

Digitaalisen aineiston säilyttämisen perustyötä on bittien säilyttäminen ja niiden siirtäminen muistivälineeltä toiselle. Digitaalisen aineiston kopioinnissa hyvänä

puolena on se, että aineistoa pystyy kopioimaan häviöttä lähes rajattomasti toisin kuin analogista aineistoa. Ei kopiointikaan silti aina ole aivan yksinkertaista: Hakala ja Stenvall ottavat artikkelissaan ongelmalliseksi esimerkiksi 5 tuuman levykkeet, joita lukevia levykeasemia ei välttämättä enää löydy. (Hakala & Stenvall 2002, 85–86.) Vaikka levykkeellä oleva tieto olisikin luettavaa ja käyttökelpoista, ei siitä ole hyötyä, jos levykettä ei voida käyttää.

Hakala ja Stenvall nostavat esiin myös taloudelliset näkökohdat: Tekniset edellytykset aineiston lukemiselle ja talteen saamiselle voivat olla kunnossa, mutta usein työ on hyvin hidasta, työvoimavaltaista ja edellyttää erityisammattitaitoa. Tällöin varsinkin suurten aineistomäärien yhteydessä taloudelliset rajat tulevat vastaan. (Hakala & Stenvall 2002, 85–86.)

Ongelmallista pelkässä datan kopioinnissa on myös se, että pelkät bitit eivät kerro siitä, miten data pitäisi tulkita, vaan tulkintaan tarvitaan ohjelmisto, jonka avulla data muuttuu informaatioksi. Ongelma ei ratkea edes tallentamalla aineiston tuottamisessa ja käytössä tarvittava ohjelma (tai ohjelmat), sillä tulevaisuudessa ei välttämättä ole enää käyttöympäristöä, jossa ohjelma toimisi. (Hakala & Stenvall 2002, 86.)

Kansallisen digitaalisen kirjaston pitkäaikaissäilytysjaosto on loppuraportissaan määritellyt erikseen vielä säilyttävän kopioinnin. Tällöin digitaalinen informaatio siirretään tallennusalustalta toiselle kahdesta syystä: joko halutaan säilyttää tieto teknisesti muuttuvassa ympäristössä tai halutaan hajauttaa sen säilytys. Säilyttävällä kopioinnilla on kolme tekijää, jotka erottavat sen tavallisesta kopioinnista:

informaatio sisältö säilyy täydellisenä  
uusi kopio korvaa vanhan  
säilyttävän kopioinnin täydellinen hallinta ja vastuu on PAS-järjestelmällä.

(Kansallinen digitaalinen kirjasto 2010b, 13.)

#### 4.2.2 Migraatio ja konvertointi

Migraatiolla ja konvertoinnilla tarkoitetaan alan kirjallisuudessa samaa asiaa, ja tässä työssä olen valinnut käytettäväksi pääsääntöisesti termiä migraatio. Migraatiolla eli konvertoinnilla tarkoitetaan sitä, että aineistot muunnetaan eli konvertoidaan tarvittaessa uuteen laitteisto- ja ohjelmistoympäristöön sopiviksi. Esimerkiksi kun vanha tekstinkäsittelyohjelma poistuu käytöstä, muunnetaan sillä tehdyt dokumentit uudelle, korvaavalle tekstinkäsittelyohjelmalle soveltuviksi. Muutaman vuoden kuluttua sama operaatio joudutaan toistamaan, sillä uusi, korvaava tekstinkäsittelyohjelma on saanut tilalleen jälleen uuden ohjelman. (Hakala & Stenvall 2002, 86.)

Migraatio pitkäaikaissäilytyksen menetelmänä perustuu vahvasti uskoon standardeihin. Uskotaan siis, että standardeja hyödynnettäessä migraatiota ei tarvitse tehdä niin usein ja että migraatiot ovat helpompia. Jossain määrin tämä varmasti pitääkin paikkansa, sillä standardeja käytettäessä migraation tarve on yleismaailmallinen, jolloin tarvittavia välineitä on todennäköisesti hyvin saatavilla. Kuitenkaan standardien merkitystä ei saisi yliarvioida. Tekniikan kehitys on nykyään niin nopeaa, että standardien elinkaari voi olla vain muutamia vuosia. Lisäksi monilta aloilta standardi puuttuu kokonaan, esimerkkinä tästä Hakala ja Stenvall mainitsevat stillkuvaformaattit, joiden suuresta joukosta ei ole löytynyt ISO-standardia tai edes kansallista standardia. (Hakala & Stenvall 2002, 86–87.)

Ongelmana migraatiossa on myös sen tavoite: halutaanko dokumentti säilyttää täsmälleen samanlaisena vai pyritäänkö vain intellektuaalisen sisällön säilyttämiseen? Yksinkertaisesta formaatista monipuolisempaan konvertoitaessa tämä tavoite toteutuu melko varmasti, mutta päinvastaisessa migraatiossa voidaan menettää muotoilutietojen lisäksi myös oleellista tietoa. (Hakala & Stenvall 2002, 87.) Olisikin hyvä, jos alkuperäinen tiedosto voitaisiin säilyttää ja konvertoida aina alkuperäisestä tiedostosta uusimpaan formaattiin eikä monien välissä olevien formaattien kautta (Caplan 2008b, 11). Lisäksi migraatio voi vaatia tuloksen tarkistamista manuaalisesti, jotta tulosta ei ole pilattu käyttökelvottomaksi (Hakala & Stenvall 2002, 87). Toisaalta konversiot

on tehtävä automaattisesti taloudellisista syistä (Liikearkistoyhdistys ry 2009, 51). On myös formaatteja, jotka eivät sovellu konvertoitaviksi, tällaisia ovat esimerkiksi ohjelmistot (Hakala & Stenvall 2002, 87).

Jos pitkäaikaissäilytysstrategiaksi valitaan migraatio, kannattaa erilaisia tiedostomuotoja normalisoida. Normalisointi perustuu ajatukseen, että toiset tiedostomuodot ovat paremmin säilytettävissä kuin toiset. Normalisoinnissa alkuperäinen tiedosto muutetaan johonkin toiseen tiedostomuotoon, joka kuitenkin säilyttää alkuperäisen tiedoston oleelliset ominaisuudet. Esimerkiksi eri tekstinkäsittelyohjelmilla tuotetut tiedostot voidaan kaikki muuttaa Rich Text Format- eli RTF-muotoon. Tällöin tulevaisuudessa pitkäaikaissäilytysjärjestelmässä on vähemmän erilaisia ylläpidettäviä tiedostomuotoja. Huono puoli normalisoinnissa on sen aiheuttama mahdollinen tietojen häviäminen, sillä alkuperäisten tiedostojen kaikki ominaisuudet eivät säily tiedostomuodon muutoksessa. Tällöin on erityisen tärkeää miettiä, mitkä ominaisuudet ovat säilyttämisen arvoisia. (Caplan 2008b, 11.)

#### 4.2.3 Emulointi

Emuloinnilla tarkoitetaan alkuperäisen käyttöjärjestelmän jäljittelyä. Tällöin tallennetaan dokumentti ja sen käyttöön tarvittavat ohjelmistot, jonka jälkeen rakennetaan sovellus eli emulaattori, joka jäljittelee alkuperäistä dokumenttien käyttöympäristöä. Näin vanhoja dokumentteja ja ohjelmia pystytään käyttämään uudessa laitteisto- ja ohjelmistoympäristössä. Koska kyseessä on aineiston säilyttäminen nimenomaan pitkäaikaisesti, saatetaan päätyä tilanteeseen, jossa emulaattoreita joudutaan käyttämään useita toinen toisensa päällä. Hakala ja Stenvall kertovat artikkelissaan 4–5 emulaattorin päällekkäisestä käytöstä, eikä tämän määrän ole vielä todettu aiheuttavan ongelmia. Emuloinnissa tärkeää on se, että resurssit, siis tallenteet ja käyttöympäristö, kuvaillaan tarkasti, jotta ne voidaan tarvittaessa rekonstruoida. (Hakala & Stenvall 2002, 87–88.)

Caplan (2008b, 10) pitää emulointia pitkäaikaissäilytyksen strategiana yhä varsin kokeellisena. Tietokoneiden koko ajan muuttuva tekniikka tekee emuloinnista hankalaa, sillä emulaattoreita täytyy päivittää ja kirjoittaa uudelleen

jokaiselle uudelle käyttöjärjestelmälle. Caplan esittelee Universal Virtual Machinen eli UVM:n, jonka tarkoituksena on olla välittäjänä emulaattorin ja käytössä olevan käyttöjärjestelmän välillä. Tällöin emulaattori pystytään erottamaan sen hetkisestä käyttöjärjestelmästä ja siinä tapahtuvista muutoksista. Vaikka UVM todennäköisesti vaatii päivittämistä, on Caplanin mielestä helpompaa päivittää yhtä UVM:ää kuin monia emulaattoreita. (Caplan 2008b, 10–11.)

Emulointia on kritisoitu sen monimutkaisesta tekniikasta. Lisäksi arvostelua on herättänyt se, että päästäkseen käsiksi vanhoihin dokumentteihin käyttäjän pitää opetella käyttämään vuosien tai vuosikymmenten takaisia ohjelmistoja. Konvertoinnillahan nämä dokumentit olisivat käytettävissä nykyisillä ohjelmistoilla. Onkin esitetty ajatuksia erityisten katseluohjelmien rakentamisesta. Näiden katseluohjelmien avulla aineistoon päästäisiin käsiksi ilman vanhojen ohjelmistojen hallintaa. (Hakala & Stenvall 2002, 88–89.) Emulaatiota pidetään lähinnä lyhyen ajan säilyttämiseen soveltuvana tekniikkana kuin pitkäaikaiseen säilyttämiseen tai edes keskipitkään säilyttämiseen soveltuvana (Liikearkistoyhdistys ry 2009, 52).

#### 4.3 Kansallinen digitaalinen kirjasto eli KDK

Kansallinen digitaalinen kirjasto eli KDK on opetus- ja kulttuuriministeriön hanke, jossa toimijoina ovat kirjastot, arkistot ja museot sekä muut kulttuuriperintöaineistoa säilyttävät organisaatiot. Hanke tähtää kulttuurin ja tieteen digitaalisten aineistojen tehokkaaseen ja laadukkaaseen hallintaan, jakeluun ja pitkäaikaissäilytykseen sekä niiden digitoinnin edistämiseen. Hankkeessa Kansalliskirjasto vastaa KDK:n asiakasliittymän ylläpitämisestä ja Tieteen tietotekniikan keskus CSC vastaa keskitetyn pitkäaikaissäilytysjärjestelmän suunnittelu- ja toteuttamishankkeesta. (Kansallinen digitaalinen kirjasto 2011b.) Hanke on perustettu toukokuussa 2008 (Kansallinen digitaalinen kirjasto 2010a, 8).



KDK:n painopistealueet ovat:

- kirjastojen, arkistojen ja museoiden yhteisen asiakasliittymän ylläpito ja kehittäminen sekä aineistojen välittäminen Euroopan digitaaliseen kirjastoon Europeanaan
- kulttuuriperintö- ja asiakirja-aineistojen digitointi
- digitaalisten kulttuuriperintö- ja asiakirja-aineistojen keskitetyn pitkäaikaissäilytyspalvelun suunnittelu ja toteuttaminen
- tietohallinnon yhteentoimivuuden edistäminen.

(Kansallinen digitaalinen kirjasto 2011b.)

Opetus- ja kulttuuriministeriö nimitti KDK-hankkeelle seuranta- ja ohjausryhmän. Ohjausryhmän alaisina toimivat saatavuus- ja pitkäaikaissäilytysjaostot ja tekninen asiantuntijaryhmä. Näistä pitkäaikaissäilytysjaoston eli PAS-jaoston vastuulla oli digitaalisessa muodossa olevien kulttuuriperintöaineistojen pitkäaikaissäilyttämisen kehittämishankkeen ohjaus. Pas-jaoston päättyi toukokuussa 2010. (Kansallinen digitaalinen kirjasto 2010a, 8.) Myöhemmin, elokuussa 2010, käynnistyi pitkäaikaissäilytyksen jatkohanke, joka kestää vuoden 2013 loppuun saakka (Kansallinen digitaalinen kirjasto 2011c).

Keskitetylle pitkäaikaissäilytysjärjestelmälle on tarvetta, sillä viranomaisasiakirjat, tieteelliset ja muut julkaisut, musiikki- ja muut äänitteet sekä radio- ja televisio-ohjelmat tuotetaan yhä enemmän ja enemmän alkujaankin digitaalisesti, ja aineistoja myös digitoidaan koko ajan. Monilla laitoksilla ja virastoilla on lakisääteinen velvollisuus säilyttää kulttuuriperintöä, oli se sitten digitaalisessa tai analogisessa muodossa. Tällaisia tahoja ovat esimerkiksi arkistolaitos, Kansalliskirjasto, Museovirasto ja Valtion taidemuseo. (Opetusministeriö 2008, 5.)

PAS-jaosto organisoii työnsä niin, että vetovastuu oli Kansallisarkistolla, mutta projektissa oli mukana henkilöitä mm. myös Tieteen tietotekniikan keskukselta CSC:stä, Museovirastosta, Kansalliskirjastosta sekä monista muista museoista ja arkistoista. (Kansallinen digitaalinen kirjasto 2010a, 10–11.) Jatkohankkeessa vetovastuun kantaa Tieteen tietotekniikan keskus CSC, mutta mukana on

edelleen henkilöitä esimerkiksi Kansalliskirjastosta, Kansallisarkistosta ja Museovirastosta (Kansallinen digitaalinen kirjasto 2011c).

PAS-jaosto on laatinut tai tilannut erilaisia raportteja ja selvityksiä koskien pitkäaikaissäilytyksen eri osa-alueita. Keskeisimpiä ovat esimerkiksi raportit PAS-järjestelmän vaatimuksista, PAS-palvelun kuvaus, kuvaus PAS-järjestelmän organisatorisesta rakenteesta ja vastuista, raportti TAPAS-määrityksistä eli hallinnollisista metatiedoista, katsaus kansainvälisiin PAS-hankkeisiin sekä selvitys PAS-hankkeen hyödyistä organisaatioille ja näiden valmiuksien kehittämistä. (Kansallinen digitaalinen kirjasto 2010a, 11.) PAS-jatkohankkeen tarkoituksena puolestaan on tuottaa yksityiskohtainen suunnitelma, jonka mukaan pitkäaikaissäilytysjärjestelmä voidaan toteuttaa sekä edistää yhteistoimintaa pitkäaikaissäilytyksen tavoitteiden ja painopisteiden hyväksi. (Kansallinen digitaalinen kirjasto 2011c.)

## 5 TUTKIMUSHAASTATTELU

Tätä tutkimusta varten haastateltiin suomalaisia maakuntakirjastojen edustajia, jotka ovat työnsä puolesta tekemisissä maakuntakokoelmien kanssa. Tarkoituksena oli saada tietoa muiden maakuntakirjastojen tämänhetkisistä menetelmistä ja tulevaisuudensuunnitelmista verkossa olevan aineiston löytymisestä ja pitkäaikaissäilyksestä. Ajatuksena oli siis löytää hyväksi havaittuja malleja, joista Satakuntakokoelmassakin voitaisiin ottaa opiksi.

### 5.1 Haastateltavat

Tutkimuksessa haastateltiin seitsemän eri maakuntakirjaston edustajaa. Kaiken kaikkiaanhan Suomessa on 19 maakuntakirjastoa, eikä kaikkien maakuntakirjastojen edustajien haastattelemisen ollut mielekäästä. Hirsjärvi & Hurme (2004, 58–59) puhuvat harkinnanvaraisesta näytteestä, joka valitaan silloin, kun halutaan saada syvällistä ymmärrystä käsiteltävään asiaan tilastollisten yleistysten sijaan. Tällöin haastattelemalla jo muutamaakin henkilöä voidaan saada merkittävää tietoa.

Tutkimukseen valitut maakuntakirjastot valittiin osittain Porin kirjastotoimen apulaisjohtaja Paula Kauppilan ehdotuksen perusteella (P. Kauppila, sähköposti 10.5.2011), osittain tutkijan oman mielenkiinnon perusteella. Kauppilan ehdotuksesta mukaan otettiin seuraavat (suluissa haastateltu kyseisen kirjaston edustaja): Tampereen kaupunginkirjasto - Pirkanmaan maakuntakirjasto (vs. kokoelmapäällikkö Maaret Lehtovirta), Hämeenlinnan kaupunginkirjasto-maakuntakirjasto (informaatikko Pirjo Lehtilä), Oulun kaupunginkirjasto-maakuntakirjasto (kirjastonhoitaja Jaana Märsynaho), Joensuun seutukirjasto – Pohjois-Karjalan maakuntakirjasto (erikoiskirjastonhoitaja Seija Salminen) sekä Kouvolan kaupunginkirjasto-maakuntakirjasto (informaatikko Tuomas Kunttu). Tutkijan omia valintoja olivat Seinäjoen kaupunginkirjasto-maakuntakirjasto (kirjastotoimen apulaisjohtaja Sinikka Ijäs) ja Jyväskylän kaupunginkirjasto - Keski-Suomen maakuntakirjasto (informaatikko Tiina Teijonsalo). Näiden lisäksi haastattelua pyydettiin vielä kahdelta maakuntakirjastolta, mutta molempien

edustajat kieltäytyivät vedoten siihen, että kirjastossa ei ollut vielä käsitelty pitkäaikaissäilytysasioita ollenkaan.

Haastateltavien yhteystiedot löytyivät jokaisen maakuntakirjaston verkkosivuilta. Joidenkin haastateltavien kohdalla otettiin aluksi yhteyttä eri henkilöön verkkosivujen yhteystietojen ja vastuualueiden perusteella. Tällöin kyseiset henkilöt ohjasivat haastattelijan eteenpäin toiselle henkilölle, jolla on enemmän tietoa kysymysten aihealueista.

## 5.2 Haastattelujen toteuttaminen

Haastattelut toteutettiin pääsääntöisesti sähköpostin välityksellä, vaikka aluksi olikin tarkoitus haastatella kaikkia puhelimitse. Haastateltaviin otettiin ensin yhteyttä puhelimitse ja tiedusteltiin halukkuutta osallistua haastatteluun. Varsinainen puhelinhaastattelu-aika oli tarkoitus sopia tämän ensimmäisen puhelun aikana, jotta haastateltavalle jää aikaa valmistautua haastatteluun (Hirsjärvi & Hurme, 2004, 64). Tämän jälkeen haastateltavat saivat haastattelukysymykset sähköpostitse, ja tämän jälkeen suurin osa haastateltavista halusi myös vastata kysymyksiin sähköpostilla. Vain yksi haastattelu toteutettiin kokonaan puhelimitse. Tämä puhelinhaastattelu nauhoitettiin haastateltavan suostumuksella, jotta haastattelun myöhempi analysointi ei olisi vain haastattelijan muistin ja muistiinpanojen varassa.

Virtuaalisen haastattelun etuna on se, että haastattelijan ja haastateltavan ei tarvitse olla samassa paikassa samaan aikaan, jolloin säästetään matkakustannuksissa. Osallistujat voivat myös kirjoittaa vastauksensa silloin, kun heille sopii. Virtuaalisessa viestinnässä haastateltavien saattaa olla vaikea arvioida haastattelijan luotettavuutta, jolloin luottamus on rakennettava tiedottamalla avoimesti tutkimuksen tarkoituksesta ja järjestämällä kontaktin ottaminen teknisesti helpoiksi. (Tiittula, Rastas & Ruusuvuori 2005, 266–269.) Myös Hirsjärvi & Hurme (2005, 84–85) korostavat henkilökohtaisen kontaktin merkitystä varsinaisen haastattelun aloittamisen helpottajana ja haastateltavan motivoijana. Juuri luottamuksen vuoksi haastatteluissa otettiin ensimmäinen kontakti puhelimitse, jotta haastateltavilla oli mahdollisuus perehtyä

tutkimukseen ja kysyä epäselviksi jääviä asioita. Teknistä helppoutta taas pyrittiin edistämään sillä, että haastateltavat saivat valita itselleen parhaiten sopivan vastaustavan.

### 5.3 Haastattelukysymykset

Tässä tutkimuksessa haastattelumenetelmänä oli puolistrukturoitu haastattelu, jota voidaan kutsua myös teemahaastatteluksi (Hirsjärvi & Hurme 2004, 47). Haastatteluthan voidaan karkeasti jakaa strukturoituihin ja strukturoimattomiin haastatteluihin, joita nimitetään joskus myös standardoiduksi ja standardoimattomaksi haastatteluksi. Eron strukturoidun ja strukturoimattoman haastattelun välillä tekevät kysymysten valmius ja sitovuus: strukturoidussa haastattelussa kysymykset ja vastausvaihtoehdot ovat valmiita, kun taas strukturoimattomassa haastattelussa ne muotoutuvat haastateltavan mukaan haastattelun kuluessa. (Tiittula & Ruusuvuori 2005, 11.)

Puolistrukturoitu haastattelu eli teemahaastattelu sijoittuu siis näiden kahden ääripään väliin. Kysymysten muoto ja järjestys voivat vaihdella, eikä vastausvaihtoehtoja ole määrätty ennalta, kuten strukturoidussa haastattelussa. (Hirsjärvi & Hurme 2004, 47; Tiittula & Ruusuvuori 2005, 11.)

Asiantuntijahaastatteluissa käy usein niin, että haastatteluaineiston keruu ja aineiston analyysi kietoutuvat yhteen, eikä analyysiä voi erottaa täysin omaksi työvaiheekseen. Haastattelijan on perehdyttävä aiheeseen etukäteen, jotta hän saa kunnon haastattelurungon, osaa tarvittaessa esittää tarkentavia kysymyksiä haastateltavalle sekä osaa auttaa haastateltavaa palauttamaan mieleensä tarpeellisia asioita. (Alastalo & Åkerman 2010, 373, 378–379.)

Tässäkin tutkimuksessa haastattelukysymysten muotoilu lähti liikkeelle tutustumisella tiedonhankinnan ja pitkäaikaissäilytyksen teoriaan. Koska tarkoitus oli selvittää, kuinka maakuntakokoelman hoitaja löytää tietoa ja aineistoa kokoelmaansa internetistä, piti tutkijan tutustua erilaisiin tiedonhankintatapoihin ja verkossa oleviin tiedonlähteisiin. Tässä apuna olivat erilaiset internetiä käsittelevät tiedonhaun oppaat, kuten Haasion (2009a)

Johdon assistentin tiedonhankinnan opas ja Heikkisen, Aganderin, Ijäksen & Laitisen (2005) Tiedonhakijan teho-opas. Kysymyksissä haluttiin esitellä mahdollisimman monta erilaista tiedonhankinnan tapaa, jotta haastateltava muistaisi haastatteluhetkellä kaikki erilaiset keinot, joita hän tiedonhaussa käyttää.

Myös pitkäaikaissäilytyksen osalta hyödynnettiin teorialähteitä, etenkin Kansallisen digitaalisen kirjaston (2010a) eli KDK:n Pitkäaikaissäilytysjaoston loppuraportti auttoi jäsentämään pitkäaikaissäilytyksen nykytilaa. Pitkäaikaissäilytystä käsittelevät kysymykset muotoutuivat kuitenkin paljolti toimeksiantajan eli Porin kaupunginkirjasto – Satakunnan maakuntakirjaston edustajien kanssa käydyistä keskusteluista ja niissä esiin nousseista käytännön kysymyksistä. Myös kysymykset verkkoaineiston luetteloinnista olivat lähtöisin käytännön tarpeesta.

Kysymysten suunnittelussa pyrittiin huomioimaan se seikka, että haastattelua ei toteuteta kasvokkain, vaan puhelimitse tai sähköpostitse. Hirsjärvi & Hurme (2004, 65) toteavatkin, että puhelinhaastattelussa kysymysten pitäisi olla kasvokkaishaastatteluun verrattuna lyhyempiä. Virtuaalihaastatteluissa kysymysten sanamuotojen olisi hyvä olla mahdollisimman eksakteja, sillä haastateltavalla ei ole mahdollisuutta tarkistaa kysymyksen merkitystä välittömästi, sillä haastateltava ei ole välittömässä vuorovaikutuksessa haastattelijan (Tiittula ym. 2005, 265–266).

#### 5.4 Haastattelun tulokset

Seuraavaksi käydään läpi haastatteluissa esille nousseita asioita osa-alueittain. Tuloksia luettaessa on huomattava se, että kaikki haastateltavat eivät vastanneet jokaiseen kysymykseen. Joku jätti vastaamatta kokonaan jonkin osa-alueen kysymykseen, joku vain muutamaan kysymykseen. Kysymyksiin jätettiin vastaamatta siksi, ettei vastaaja tuntenut tietävänsä asiasta tarpeeksi, käsitellyt asiat eivät kuuluneet hänen työnkuvaansa tai vastaajan organisaatiossa ei kyseessä olevia asioita ollut vielä mietitty.

#### 5.4.1 Tiedonhaku ja aineistojen löytyminen

Haastattelun ensimmäinen osio käsitteli tiedonhakua verkossa ja aineiston löytymistä verkosta. Tarkoitus oli siis selvittää, miten kirjastonhoitaja etsii kokoelmaan sopivaa aineistoa verkosta, mitä menetelmiä ja apuvälineitä hän käyttää, mitä ja miksi hän verkosta etsii sekä saako hän tyydyttäviä tuloksia.

Aluksi haastateltavilta tiedusteltiin, onko heillä tiedonhakua varten jokin tietty suunnitelma tai kaava, jota noudattamalla tietoa haetaan verkosta. Kenelläkään vastaajista ei tällaista kaavaa ollut, vaan tieto löytyy verkosta enemmänkin sattumalta. Kun kirjastonhoitaja seuraa alaa ja eri medioita, esimerkiksi paikallislehtiä, hän voi saada vinkkejä kiinnostavista verkkosivuista. Myös muun tiedonhaun yhteydessä voi sattumalta törmätä kiinnostavaan aineistoon. Myös järjestelmällistä tiedonhakua on tehty, esimerkiksi Hämeenlinnan Lydiaa varten informaattikko on käynyt systemaattisesti läpi digitaalisia kokoelmia, tällaisia olivat esimerkiksi Kansalliskirjaston, Åbo Akademin ja Kansallisarkiston kokoelmat. (Lydia on Hämeenlinnan kaupunginkirjaston digitaalinen aineistopankki, ks. <http://kirjasto.hameenlinna.fi/kirjasto/lydia/>.) Myös muita vastaavia yksittäisiä projekteja tai hankkeita on ollut eri kirjastoilla.

Lähes kaikilla haastatelluilla on joitakin verkkosivuja, joilla he vierailevat säännöllisesti tai melkein säännöllisesti. Tällaisia verkkosivuja ovat esimerkiksi Kansalliskirjaston digitoinnit, korkeakoulujen julkaisuja ja opinnäytteitä sisältävät sivut, Kirjasampo, Kirjastot.fi-palvelun Linkkikirjasto ja oman kaupungin julkaisuja sisältävät sivut. Artikkelitietokantoja ei käytetä yhtä säännöllisesti, mutta esimerkiksi Aleksia käytetään silloin tällöin. Samoin tallennettaessa tietoa itse artikkelitietokantaan saatetaan vilkaista, löytyisikö jotain omaan kokoelmaan sopivaa.

Haastattelussa haluttiin tietää, minkä tyyppistä materiaalia verkosta etsitään. Haastateltavien vastaukset vaihtelivat jonkin verran. Osa ei hae verkosta varsinaisesti kokoelman täydennystä, vaan vain esimerkiksi opinnäytteitä tai sellaista tietoa, jota ei löydy painettuna. Osa taas etsii monenlaista materiaalia: rakennusten historiaan tai rakennuskulttuuriin liittyviä selvityksiä, muita

historiikkeja, kotiseutuhenkilöihin liittyvää aineistoa, kulttuuriorganisaatioiden verkkosivuja tai vanhaa, digitoitua aineistoa. Erityyppisten aineistojen haut eivät välttämättä eroa juurikaan toisistaan. Tiedonlähteinä olevat verkkosivut vaihtelevat luonnollisesti etsittävän aineiston mukaan, esimerkiksi karttoja etsitään arkistolaitoksesta tai vanhakartta.fi-verkkosivustolta, kun taas kaunokirjallisuutta etsitään Kansalliskirjaston digitoinneista, Gutenberg- ja Runeberg-projektien sivuilta tai painetusta materiaalista.

Kaikki käyttävät tiedonhakuun Googlea hakukoneena ainakin jonkin verran. Kenellekään se ei ole paras työkalu verkkotiedonhaussa, sillä – kuten eräs haastateltava toteaa – sen avulla ei löydy kovinkaan relevantteja viitteitä. Kuitenkin sen löytämien tulosten avulla voi päästä oikean aineiston jäljille. Parhaiten Googlen käyttäminen sujuu silloin, kun tietää etukäteen, mitä aineistoa hakee. Myös tietyn, nimetyn aineiston, kuten kotisivujen etsintään, se soveltuu erityisen hyvin. Metahakukoneita ei käyttänyt juuri kukaan haastateltavista. Ainoastaan yksi kertoi käyttävänsä Dogpile-metahakukonetta, mutta sitäkin vain satunnaisesti, sillä hän ei aina muista sen olemassaoloa.

Neljä haastateltavaa kertoi hyödyntävänsä Wikipediaa tai sosiaalista mediaa tiedonhaussa. Wikipediasta löytyy esimerkiksi alueen kirjailijoiden blogien tai kotisivujen osoitteita tai muita alueeseen liittyviä verkkolähteitä tai sieltä voi tarkastaa jonkin tiedon paikkansapitävyyden. Sosiaalisesta mediasta mainittiin Facebook, jota pidettiin hyvänä kanavana saada vinkkejä kiinnostavista verkkosivuista. Facebookin todettiin auttavan myös silloin, kun tietoa ei löydy painetuista lähteistä.

Informaaleista lähteistä saavat lähes kaikki haastateltavat apua tiedonhaussa ainakin joskus. Enimmäkseen apua saadaan omilta työkavereilta, harvemmin asiakkailta. Yksi haastateltava mainitsi saaneensa juuri vinkin eräästä aluettaan koskevasta pro gradu -työstä tutulta historiantutkijalta.

Joillakin kirjastoilla on perinteisesti ollut sopimus eri julkaisijoiden, yritysten tai yhteisöjen kanssa uusien painettujen julkaisujen toimittamisesta kirjastoon tai ainakin niiden ilmestymisen ilmoittamisesta kirjastolle. Sähköisten julkaisujen



osalta tällaista sopimusta ei monellakaan haastateltavista ole, vain kaksi haastateltavaa saa säännöllisesti tiedon uusista julkaisuista. Sopimuksia on tehty niin paikallisen yliopiston kirjaston kuin Kelan ja muutamien yhdistystenkin kanssa. Paikallisen yliopiston kirjasto ilmoittaa tiedot niistä opinnäytteistä, jotka koskevat maakunta-aluetta, ja sama sopimus on suunnitteilla myös kyseisen kaupungin ammattikorkeakoulun kanssa.

Haastateltavilta kysyttiin myös tiedonhaun lähtökohtia. Lähtevätkö tiedonhaut siis enemmänkin asiakkaiden tarpeista vai kokoelman kartuttamisesta? Lähes kaikki haastateltavat pitivät kokoelman kartuttamista ja puutteiden paikkaamista hakujen pääasiallisena syynä, mutta yksi haastateltava kertoi hakujen lähtevän lähes aina asiakkaiden tarpeista. Monilla haastateltavilla hakujen syinä olivat sekä asiakkaiden että kokoelman tarpeet, ja yksi haastateltava totesikin aineiston kokoamisen tarpeiden olevan toivottavasti myös asiakkaiden tarpeita. Näitä ei siis voi täysin erottaa toisistaan.

Kysyttäessä vielä muita mahdollisia tapoja tiedonhakuun yksi haastateltava nosti esille uutena mahdollisena työkaluna Google Alertin. Se lähettää sähköpostiin uusimpia osuvia hakutuloksia itse määritellystä aiheesta itse valitulla ilmoitustiheydellä (Google 2011a). Toinen haastateltava taas kertoi käyvänsä läpi kaikki alueensa sanomalehdet, joista hän saa vinkkejä kiinnostavista verkkosivuista ja julkaisuista.

Kehittämis- tai muutostoiveita haastateltavat eivät osanneet sen enempää nimetä, yksi haastateltava toivoi tulevaisuudessa eri tahojen tuottamien aineistojen kokoamista yhteen palveluun. Ongelmana taas yksi haastateltava piti verkkoaineiston keräämiseen kuluvaan aikaa, sillä se vie työaikaa fyysisen aineiston kokoelmatyöltä. Haastateltava joutuukin välillä priorisoimaan tehtäviä, sillä kaikelle ei ole aikaa.

#### 5.4.2 Verkosta löytyneen aineiston luettelointi

Joillakin maakuntakirjastoilla on aineistotietokantojen lisäksi myös omat aluetietokantansa. Tällaisia ovat esimerkiksi (suluissa julkaiseva kirjasto) Arvi-

tietokanta (Hämeenlinna), Koivikko-alue-tietokanta Pohjois-Karjalassa (Joensuu), Keski-alue-tietokanta Keski-Suomessa (Jyväskylä), Kymenlaakso-tietokanta (Kouvola), Ostrobotnia-tietokanta Pohjois-Pohjanmaalla (Oulu) sekä Porstua-portaali (Kokkola, Seinäjoki ja Vaasa). Näihin aluetietokantoihin luetteloidaan maakuntaa koskevaa aineistoa, ja joissakin kirjastoissa maakunta-aineistoa luetteloidaan myös kirjaston aineistotietokantaan. Kuitenkin esimerkiksi Jyväskylässä ja Kouvolassa maakunta-aineisto luetteloidaan pelkästään aluetietokantaan.

Pääsääntöisesti verkosta löytynyt aineisto luetteloidaan samalla tavalla kuin muukin kirjaston aineisto eli valtakunnallisten luettelointisääntöjen mukaan. Oulun kaupunginkirjasto-maakuntakirjasto ylläpitää Pakkala-kirjailijahakemistoa (ks. <http://oukasrv6.ouka.fi:8003/Intro?formid=kirqu>), jonne tallennettavalla aineistolla on oma tallennuspohjansa, joka on muokattu Finmarc-formaatista. Useampi haastateltavat kertoi, että asiasanoina käytetään YSA:n asiasanoja ja aineistolajina voi olla verkkoaineisto, verkkolehti, verkkoartikkeli, verkkokartta tai PDF-tiedosto. Aineistoa kuvaillaan pääasiassa asiasanoin kuten painettuakin aineistoa. Yksi haastateltava kertoi, että ulkoasun kohdalle kirjoitetaan ”www-sivusto” tai ”PDF-tiedosto”, lisäksi aineiston verkko-osoite kirjoitetaan omaan kenttäänsä.

Kysymys siitä, tallennetaanko verkossa olevasta aineistosta jotakin metadataa luetteloitaessa, sai melko vaihtelevia vastauksia. Kysymys oli mitä luultavimmin epäselvästi määritelty, eikä metadatan käsite ollut näin ollen kaikille haastateltaville selvä. Osa haastateltavista kertoi, että metadataa ei tallenneta. Yksi haastateltavista kertoi, että luetteloinnin yhteydessä tallennetaan sellaista metadataa, joka istuu Finmarc-kenttiin, ja toinen haastateltavista kertoi tietueisiin tallennettavan asiakkaille näkymätöntä tietoa, kuten luetteloinnin taso, kansallisuuskoodi ja kohderyhmäkoodi. Tampereen kaupunginkirjasto – Pirkanmaan maakuntakirjasto ylläpitää kirjailijoista kertovaa Pirkanmaalaisia nykykirjailijoita -verkkosivustoa (ks. <http://www.tampere.fi/kirjasto/pirkanmaankirjailijat/nykykirjailijat.htm>), jonne

puolestaan on tallennettu metadataa, jotta sivusto olisi mahdollisimman helposti hakukoneiden löydettävissä verkosta.

#### 5.4.3 Verkosta löytyneen aineiston pitkäaikaissäilytys

Maakuntakokoelmissa aineistoja säilytetään ikuisesti sekä painetussa että digitaalisessa muodossa. Tämä on yleisperiaate, joka koskee kaikkia maakuntakokoelmia. Tämän periaatteen lähes jokainen haastateltava nosti esille kysyttäessä aineiston säilytysajasta. Yksi haastateltava totesikin, että lyhytaikaista materiaalia ei ole maakuntakokoelmassa lainkaan, vaan kaikki on pysyvästi säilytettävää. Esimerkkinä hän mainitsee oman maakuntansa kirjailijatietokannan, jota ollaan nyt hiljalleen ajamassa alas Kirjasampoon siirtymisen vuoksi. Tietosisältö silti itsessään säilyy pysyvästi, vaikka sen ulkoinen olemus muuttuukin omasta tietokannasta osaksi Kirjasampoa.

Pitkäaikaissäilytykseen ei kellään haastatellulla ollut vielä varsinaista ratkaisua. Yksi haastateltu totesi, että ei heillä vielä ole kokoelmassa sentyyppistä aineistoa, jonka vuoksi olisi jouduttu miettimään pitkäaikaissäilytystä. Kansallisen digitaalisen kirjaston eli KDK-hankkeen tuloksia odotetaan myös. Eräs haastateltava kertoi, että tällä hetkellä ainoa teko pitkäaikaissäilytyksen hyväksi on paperikopioiden tulostaminen joistakin verkkoartikkeleista, mutta muuten säilytyksen on ajateltu olevan Kansalliskirjaston ja KDK-hankkeen vastuulla. Toinen haastateltava kertoi kirjastonsa seuraavan mielenkiinnolla KDK-hanketta ja siihen liittyvää PAS-hanketta, ja toivoi näiden hankkeiden löytävän kansallisia ratkaisuja ja tallennuspaikkoja paikalliselle aineistolle, jota eri maakunnissa digitoidaan. Hän toivoisi Kansalliskirjaston tarjoavan tallennusalustan ja yhteiset ohjeistukset ja käytännöt näille materiaaleille.

Aineistoa säilytetään pitkäaikaisesti eri kirjastoissa eri tavoin. Yksi haastateltava kertoi aineiston olevan palvelimella, ja jos palvelin vaihtuu, siirtyvät aineistotkin uudelle palvelimelle ja lisäksi kirjastolla on aineistosta kopiot myös CD-ROM-levyllä. Toisen haastateltavan mukaan aineistoa säilytetään palvelimella ja tietokoneen kovalevyllä, osa aineistosta on myös Kansalliskirjastolla. Kolmas haastateltava kertoi kirjastonsa aineistojen olleen kaupungin palvelimella, jolla

oli myös muunlaista materiaalia. Aineisto on nyt kuitenkin siirretty kaupungin mediapalvelimelle, joka on tarkoitettu nimenomaan kuvien, videoiden ja muun vastaavan materiaalin käyttöä ja tallennusta varten. Kirjastossa on myös mietitty hallintoon liittyvien dokumenttien pitkäaikaissäilytystä, ja kaupungilla onkin sitä varten oma sähköinen arkistonsa.

Aineistojen pitkäaikaissäilytyksestä vastaa pääasiassa oma organisaatio, eikä säilytyspalveluja ole juurikaan ulkoistettu. Kirjastot kuuluvat hallinnollisesti sijaintikaupunkinsa organisaatioon, ja kaupungin organisaatioon kuuluvat yksiköt, esimerkiksi tietohallinto tai tietotekninen liikelaitos, vastaavat säilytyksestä ja varmuuskopioinnista. Kirjastolla itsellään on luonnollisesti vastuu tietojen oikeellisuudesta.

Suosituin tiedostomuoto säilytyksessä on itse tuotetuissa dokumenteissa ehdottomasti PDF-muoto. Kuvat ovat esimerkiksi JPG-muodossa. Verkossa oleva aineisto, johon tehdään linkki tietokannasta, on puolestaan siinä muodossa, jossa se on verkossa alun perinkin. Jos linkitetyn aineiston verkko-osoite muuttuu, monen kirjaston kirjastojärjestelmä ilmoittaa toimimattomasta linkistä. Järjestelmä tarkistaa linkkien toimivuuden viikoittain. Tällöin yritetään etsiä uusi korvaava osoite, joka muutetaan luettelointitietoihin. Joskus käy niin, että uutta osoitetta ei löydy, jolloin luettelointitiedot on poistettava. Yksi haastateltava kertoi toimimattomien linkkien löytymisen vaikeutuneen viime aikoina, sillä kirjastonjärjestelmän vaihdon yhteydessä linkkientarkistusominaisuus poistui, eikä uusi järjestelmä ilmoita enää toimimattomista linkeistä.

Kysyttäessä tulevaisuuden suunnitelmista pitkäaikaissäilytyksen suhteen kolme haastateltava toteaa vain seuraavansa KDK-hankkeen ratkaisuja ja kansallista kehitystä. Yksi haastateltava nosti esille vapaakappalekirjastoissa tehtävän verkkoaineistojen tallennuksen. Häntä mietitytti, kuinka nämä tallennetut aineistot ovat tulevaisuudessa maakuntakirjastojen löydettävissä. Lähes kaikki verkossa oleva aineisto kuitenkin tallennetaan, joten maakuntakirjastojen kannalta olisi olennaista, että aineistot myös löytyisivät, jotta niistä olisi hyötyä. Hän kertoi myös, että suunnitelmissa on konvertoida oma aluetietokanta

MARC21-formaattiin samalla, kun kirjaston aineistotietokantakin konvertoidaan. Kirjastoon on tulossa myös uusi kirjastojärjestelmä, ja linkkien tarkistuksen ja yleisestikin verkkoaineistojen toimiminen siinä huoletti haastateltavaa.

#### 5.4.4 Muita haastatteluissa esille tulleita asioita

Haastatteluissa tuli esille myös joitakin asioita kysymysten ulkopuolelta. Niiden läpikäymisen voi katsoa olevan perusteltua, sillä ne tuovat esiin sellaisia näkökulmia, joita haastattelija ei ole osannut kysyä, mutta jotka kuitenkin ovat tärkeitä ja käsittelyä ansaitsevia asioita.

Yksi haastateltavista toi esiin ajanpuutteen ja resurssien riittämättömyyden. Hänen työnkuvansa on sen verran laaja, että hän ei ole ehtinyt kartoittaa maakunnan verkkoaineistoa ollenkaan. Resurssit eivät tahdo riittää edes omaan kaupunkiin liittyvän verkkoaineiston kartoitukseen, eikä verkkoaineistoa ole pystytty järjestelmällisesti seulomaan tai käymään läpi. Suunnitelmaa verkkoaineiston kartoituksesta ja luetteloinnista ei ole kyseisessä kirjastossa vielä tehty, mutta verkkostrategia on ollut jo mietinnässä.

Toinen haastateltava kiinnitti huomion tekijänoikeuksiin. Tällä hetkellä tekijänoikeuksien huomioiminen on suuri ongelma, sillä kirjastot eivät voi kopioida mitään aineistoa verkosta. Tekijänoikeuslain mukaan kyseessä on kappaleen valmistaminen, ja se on kiellettyä, sallittua on vain kappaleen valmistaminen yksityiskäyttöä varten. Kopioimiseen tarvittaisiin aina lupa tekijänoikeuden haltijoilta, mutta lupien kysyminen on työlästä. Tilanne hankaloituu entisestään, kun verkkosivujen tekijöitä on useita tai tekijänä on yhteisö. Myös opinnäytteiden kohdalla luvan kysyminen on hankalaa, sillä valmistuneita opiskelijoita on vaikeaa tavoittaa. Haastateltava on joidenkin opinnäytteiden kohdalla ratkaissut tekijänoikeusongelman tallentamalla aluetietokantaan opinnäytteestä vain viitetiedot, joissa on linkki opinnäytteen sijaintiin verkossa. Tällöin tosin ongelmaksi muodostuu opinnäytteen pitkäaikaissäilytys, sillä tiedosto ei ole kirjaston hallinnassa. Tekijänoikeudesta vapaata aineistoakin verkosta löytyy, ja se on luonnollisesti vapaasti käytössä.

Tällaiset aineistot ovat kuitenkin vanhoja, mikään uusi aineisto ei ole tekijänoikeudesta vapaata.

Sama haastateltava pitää ongelmana myös digitaalisen aineiston tallentamiseen liittyvää sisällönkuvailua. Hän toivoisi, että digitaalista aineistoa voisi tallentaa niin, että myös sisällönkuvailu onnistuisi. Valitettavasti kirjastojärjestelmät eivät siihen sovellu tällä hetkellä, vaan tallennukseen tarvittaisiin muita ohjelmistoja.

Lopuksi haastateltava kertoo oman visionsa ja toiveensa siitä, millainen KDK:n PAS-hankkeen tuottama ratkaisu pitkäaikaissäilytykseen olisi hänen mielestään hyvä. Hän kuvailee ympäristön, johon kirjastot voisivat itse lisätä aineistoa ja johon voisi myös kertyä automaattisesti digitaalista aineistoa verkosta ja suljetuistakin lähteistä. Näihin aineistoihin liittyvät tekijänoikeudelliset ongelmat olisi ratkaistu kansallisesti, jotta yksittäisen kirjaston ei niistä tarvitsisi huolehtia. Kirjastot voisivat lisätä aineistoon metatietoa, jota voisi tekstimuotoisesta aineistosta tuottaa myös automaattinen ohjelmisto. Tähän kerätyn ja tallennetun aineiston käyttöä varten kirjastot voisivat tehdä omia käyttöliittymiään, ja esimerkiksi maakunnallinen käyttöliittymä voisi olla sellainen. Kun samaan käyttöliittymään kirjasto yhdistäisi vielä aluetietokannan ja kirjaston aineistotietokannan, oli tiedonhakijalle tarjolla laaja paketti alueellista materiaalia, joka kaiken lisäksi säilyisi varmasti.

## 6 VERKKOTIEDONHAKU JA PITKÄAIKAISSÄILYTYS SATAKUNTAKOKOELMASSA

### 6.1 Ehdotuksia Satakuntakokoelman tiedonhakua varten

Tässä luvussa on esitetään joitakin ehdotuksia, joiden avulla Satakuntakokoelmasta vastaava ja siihen aineistoa keräävä kirjastonhoitaja voi löytää paremmin relevanttia aineistoa kokoelmaan. Lista ei ole läheskään täydellinen, vaan siinä on tuotu esiin joitakin tutkimuksen aikana esille nousseita verkkotiedonhaun apuvälineitä ja mahdollisuuksia. Lisäksi listaa koottaessa on oletettu, että tiedonhakijalla on jo hallussaan perushakutoiminnot, kuten hakukoneiden käyttö, joten niitä ei siksi näissä ehdotuksissa käsitellä.

#### 6.1.1 RSS-syötteiden käyttöönotto

Tutkimusta varten haastatellut eri maakuntakirjastojen edustajat nostivat esille jatkuvan ajanpuutteen ja sen, että verkossa olevan aineiston läpikäymiseen ei riitä aikaa. Verkossa oleva tieto täytyy siis laittaa virtaamaan automaattisesti tiedonhakijan luokse, jotta tämän ei tarvitse etsiä sitä joka kerta uudelleen.

RSS-syötteitä ja niiden tekniikkaa on käsitelty jo aiemmin tässä tutkimuksessa luvussa 3.6., joten tässä luvussa keskitytään pääasiassa niiden käyttöönottoon. RSS-syötteethän ovat ilmoituksia, joita verkkosivusto lähettää käyttäjälle, kun sivuston sisällössä tapahtuu muutoksia ja päivityksiä (23 asiaa 2009a).

RSS-syötteitä voi lukea erillisellä tietokoneelle asennettavalla lukuohjelmalla, verkossa toimivalla palvelulla tai selaimen avulla. Tietokoneelle asennettava lukuohjelma ja selaimen syötteenlukija soveltuvat sellaiselle käyttäjälle, joka työskentelee koko ajan samalla koneella, kun taas verkossa toimiva palvelu on käytössä tietokoneesta ja työpisteestä riippumatta. (23 asiaa 2009a.) Helppointa RSS-syötteiden käyttö on aloittaa selaimessa valmiina olevalla lukuohjelmalla, sillä silloin ei tarvitse erikseen kirjautua eikä luoda tiliä mihinkään verkossa

toimivaan palveluun eikä liioin asentaa uutta ohjelmaa omalle tietokoneelle. Syötteiden lukutapa kannattaa siis valita oman työskentelytavan mukaan.

Syötteiden tilaaminen selaimen kautta aloitetaan siirtymällä sille sivustolle, josta syötteitä halutaan. Esimerkkiselaimina tässä käytetään Mozilla Firefoxia (versio 7.0.1) ja Internet Exploreria (versio 9) sekä esimerkiverkkosivuna Yleisradion verkkosivua osoitteessa <http://www.yle.fi/>, josta navigoidaan Satakuntaa käsittelevien uutisten osioon (polkua etusivu > Alueet > Satakunta, suora osoite on <http://yle.fi/alueet/satakunta/>). Kun haluttu sivu on löytynyt, sieltä etsitään RSS-syötteen tilaamismahdollisuudesta kertova logo tai muu ilmoitus. Tällä Satakunta-sivulla ilmoitus on lähes sivun alalaidassa (tilanne lokakuussa 2011).

Ilmoitusta napsauttamalla voidaan tilata syöte omaan selaimeen. Sivun saattaa vielä tässä vaiheessa kysyä, minkälaisia uutisia syötteenä haluaa. Ylenkin verkkosivu kysyy, haluaako käyttäjä tilata syötteitä uutissähkeistä, videoista vai radiojutuista. Haluttu syötteiden tyyppi valitaan napsauttamalla sitä, jonka jälkeen selain avaa vielä uuden ikkunan tai välilehden syötteen tilaamisen vahvistamista varten. Tilaus vahvistetaan ja syötteen voi nimetä vielä haluamallaan tavalla sekä valita kirjanmerkkivalikosta paikan, johon web-selain uusia syötteitä tallentaa.

Tilaamisen jälkeen uudet syötteet ovat luettavissa selaimen kirjanmerkkivalikossa. selain päivittää uudet syötteet valikkoon automaattisesti. Uudet syötteet erottaa luetuista syötteen edessä olevasta pienestä ikonista: Luetuissa ikoni on erilainen kuin lukemattomissa.

Omalle tietokoneelle asennettavia ilmaisia RSS-syötteenlukuohjelmia ovat esimerkiksi FeedReader (<http://www.feedReader.com/>), SharpReader (<http://www.sharpreader.net/>) ja RssReader (<http://www.rssreader.com/>). Verkossa toimivia palveluita ovat puolestaan esimerkiksi iGoogle (<http://www.google.com/ig>) ja Netvibes (<http://www.netvibes.com/fi>). Ohjelmien ja palveluiden käyttöohjeista selviää, kuinka kyseisiin ohjelmiin tai palveluihin saadaan tilattua syötteet.



### 6.1.2 Google Alertin käyttöönotto

Maakuntakirjastojen edustajien haastatteluissa nousi myös esille palvelu nimeltä Google Alert. Siitä on kerrottu myös tämän tutkimuksen luvussa 3.3.1, jossa käsitellään tiedonhakua internetistä sanahakukone Googlea käyttäen. Google Alert on siis hakukone Googlen tarjoama palvelu.

Google Alertin ideana on etsiä tietoa käyttäjän muotoilemaa hakulausetta käyttäen, ja raportoida käyttäjälle uusista löytyneistä haku vastaavista verkkosivuista. Google Alertin etuna on, että sen avulla voi seurata mitä vain verkossa julkaistavaa tietoa mediasta ja muodosta riippumatta. (Google 2011d.) Edellisessä luvussa käsitellyt RSS-syötteethän vaativat seurattavalta sivustolta RSS-palvelua, mutta Google Alertilla voi seurata mitä tahansa sivustoa.

Google Alertin käyttö aloitetaan muotoilemalla hakusana tai hakulause asiasta, josta tietoa haetaan. Muokkauslomake löytyy osoitteesta <http://www.google.com/alerts>. Hakusanoiksi ei kannata laittaa liian yleisiä sanoja, joilla tuloksia löytyy liikaa ja koko ajan, vaan haku kannattaa rajata jotenkin. Mahdollisimman tarkat hakusanat auttavat saamaan relevantteja tuloksia. Hakulausekkeen muotoilussa voi käyttää seuraavia merkkejä tarkkuutta lisäämään:

lainausmerkit fraasin ympärillä, esimerkiksi *"pori jazz"*  
 miinusmerkki pois jätettävän sanan edellä, esimerkiksi *pori -jazz*  
 plusmerkki sanan edessä merkitsee täsmällistä kirjoitusasua, jolloin synonyymit ja vaihtoehtoiset kirjoitusasut jäävät pois, esimerkiksi *+Mari Kunnas*, jotta Google ei tarjoa tuloksia Mauri Kunnaksesta  
 site:-merkintä osoittaa, että tuloksia halutaan vain tietyltä sivustolta tai tietyn maatunnuksen alueelta, esimerkiksi *"pori jazz" site:.fi* tai *"pori jazz" site:hs.fi*

(Google 2011e.)

Kun hakulauseke on muotoiltu, tehdään muut hakuun tarvittavat asetukset. Ensinnäkin pitää valita, minkätyyppistä tietoa Google Alertin haluaa etsivän

verkosta. Vaihtoehtoina ovat kaikenlaisen tiedon etsiminen tai etsinnän rajaaminen vain blogeihin, videoihin tai keskusteluihin. Seuraavaksi valitaan ilmoitusten tiheys. Ilmoitukset saa sähköpostiinsa joko heti uusien tulosten löydyttyä, kerran päivässä tai kerran viikossa. Tämän jälkeen valitaan vielä se, valikoiko Google Alert tuloksista parhaat ja mielestään relevantteimmat, vai lähettääkö se tuloksina kaikki löytyneet sivut. (Google 2011a.)

Lopuksi annetaan sähköpostiosoite, johon Google Alert lähettää tulokset. Haasio (2009a, 64) suosittelee, että Google Alertin ilmoituksia varten varataan erillinen sähköpostiosoite. Oletettavasti hän ehdottaa tätä siksi, että tavallisesti käytetty sähköpostilaatikko ei täytyisi ilmoituksista, kun uutta tietoa on löytynyt. Tähän tarkoitukseen soveltuu varmasti jokin ilmainen sähköpostitili, joita ovat niin Googlen Gmail ([www.gmail.com](http://www.gmail.com)) kuin Microsoftin Hotmail ([www.hotmail.com](http://www.hotmail.com)), Lycos Mail (<http://www.mail.lycos.com/>) ja MTV3:n Luukkukin (<http://www.luukku.com/luukku>).

### 6.1.3 Wikien ja blogien hyödyntäminen

Maakuntakirjastojen edustajien haastatteluissa eräs haastateltava kertoi hyödyntävänsä Wikipediaa etsimällä sen kautta tietoa maakunnan kirjailijoista, esimerkiksi heidän pitämistään blogeista. Muuten Wikipedia ei aina ole lähteistä parhain, etenkin etsittäessä objektiivista ja luotettavaa informaatiota (Haasio 2008, 118).

Suomenkielisen Wikipedian löytää osoitteesta <http://fi.wikipedia.org/> ja englanninkielisen Wikipedian osoitteesta <http://en.wikipedia.org/>. Wikipedia on yhteisöllinen ensyklopedia, johon kuka tahansa voi kirjoittaa artikkeleita mistä aiheesta tahansa tai täydentää jo kirjoitettuja artikkeleita. Ideana on siis, että ihmiset jakavat tietämystään toisilleen niistä aiheista, jotka heitä kiinnostavat ja joista heillä on tietoa. (Haasio 2009a, 35.) Wikipedian vahvuus on sen nopeus ja avoimuus, jolla se mahdollistaa esimerkiksi ajantasaisen kansalaisjournalismin (Tuominen 2008, 74). Samat ominaisuudet ovat myös Wikipedian heikkouksia. Koska kuka tahansa voi kirjoittaa sinne mitä tahansa,

on artikkeleiden taso vaihtelevaa ja niiden luotettavuus välillä kyseenalaista. (Haasio 2009a, 35.)

Wikipediaa voi kuitenkin hyödyntää esimerkiksi linkkien etsimisessä (Haasio 2009a, 36). Monessa Wikipedian artikkelissa on ilmoitettu artikkelin kirjoittamisessa lähteenä käytettyjä verkkosivuja, joiden kautta pääsee tutustumaan aiheeseen laajemmin. Lisäksi monen artikkelin lopussa on Aiheesta muualla -osio, jossa on listattuna aiheesta kertovia verkkosivuja ja esimerkiksi blogeja. Esimerkiksi kirjailija Kirsti Ellilästä kertovassa Wikipedian artikkelissa (osoitteessa [http://fi.wikipedia.org/wiki/Kirsti\\_Ellilä](http://fi.wikipedia.org/wiki/Kirsti_Ellilä)) on lopussa lueteltu linkkejä häneen liittyen. Joukossa on erilaisia kirjailijaesittelyjen artikkeleita ja lehdissä julkaistuja artikkeleita, yksi linkeistä vie kirjailijan omaan blogiin.

Wikipedia ei suinkaan ole ainoa wiki-muotoista tietoa sisältävä sivusto. Wikit ovat yhteisöllisesti eli kollaboratiivisesti tuotettuja sivustoja, jotka voivat olla monialaisia kuten Wikipedia tai keskittyä yhteen tiettyyn teemaan (Tuominen 2008, 134). Wikeissä uusien sivujen luominen sekä olemassa olevien sivujen muokkaaminen ja linkittäminen keskenään on tehty mahdollisimman helpoksi, ja wikit ovatkin jatkuvassa muutoksessa olevia sivustoja. Sivujen sisältöjen muokkaaminen voi olla vapaata kaikille tai muokkaus-oikeutta voi olla rajoitettu esimerkiksi tietyille käyttäjätunnuksille. (23 asiaa 2009b.) Wikipedian sivulta Luettelo suomenkielisistä wikeistä löytyy lista eriaiheisista suomenkielisistä wikeistä, kuten eri maakuntia ja harrastuksia käsitteleviä wikejä (osoitteessa [http://fi.wikipedia.org/wiki/Malline:Luettelo\\_suomenkielisistä\\_wikeistä](http://fi.wikipedia.org/wiki/Malline:Luettelo_suomenkielisistä_wikeistä)).

Blogi on verkkopäiväkirja, jonka ideana on vuorovaikutteisuus. Blogin lukijat voivat kommentoida kirjoituksia, ja myös kommentteja voi kommentoida. Näin blogeista syntyy parhaimmillaan vilkkaita keskustelufoorumeita. (Haasio 2008, 162.) Erona wikiin blogi on kuitenkin henkilökohtainen eikä kollaboratiivinen eikä blogiin kirjoitettuja asioita muokata jälkeenpäin. Blogin teksti on keskustelunomaisempaa kuin wikin, ja se esitetään kronologisessa järjestyksessä. (23 asiaa 2009b.)

Blogeja pitävät niin yksityishenkilöt kuin erilaiset yhteisötkin, esimerkiksi monilla verkkolehdistillä, televisiokanavilla, yrityksillä tai kirjastoilla on omia blogeja (Haasio 2008, 162). Blogien pitäjät eli bloggaajat muodostavat laajan verkoston, ja he seuraavat toistensa blogeja, kommentoivat niiden kirjoituksia ja saavat ideoita omiin kirjoituksiinsa toisten blogeista. Monissa blogeissa on myös listattuna ja linkitettyinä blogeja, joita kirjoittaja seuraa. Näiden linkkien kautta löytää helposti lisää kiinnostavia blogeja, sillä kiinnostavan blogin kirjoittaja todennäköisesti seuraa muita kiinnostavia blogeja. Blogisivulla voi näkyä myös linkkejä kyseistä blogia lukevien ja seuraavien bloggaajien omiin blogeihin. (23 asiaa 2009b.)

Aiemmin tässä luvussa mainitun Kirsti Ellilän Wikipedian kautta löydetty blogi on nimeltään Kirjailijan häiriökliniikka ja sen osoite on <http://kirstiellila.blogspot.com/>. Blogissa kirjailija kertoo arjestaan ja kirjoistaan. Blogissa on myös linkit blogeihin, joita kirjoittaja seuraa. Seurattavia blogeja on noin sata kappaletta, joukossa monia kirjallisuusaiheisia blogeja, mutta myös kaikenlaisiin erilaisiin aiheisiin keskittyneitä blogeja. Kirjallisuusaiheisia blogeja on sekä kirjailijoiden pitämiä että aktiivisten lukijoiden pitämiä kirjoista kertovia blogeja.

Eri blogien kautta pysyy hyvin perillä alalla tapahtuvista ja puhuttavista asioista ja muutoksista. Kuten muitakin verkkosivuja, myös blogien uusia tekstejä voi tilata RSS-syötteenä. Näin kaikkia kiinnostavia blogeja ei tarvitse muistaa käydä lukemassa, vaan uusista kirjoituksista saa aina ilmoituksen. On myös olemassa valmiita syötelistoja, jotka tilaamalla saa samalla tiedot useasta eri blogista kerralla. Tällainen palvelu on esimerkiksi Yahoos Pipes-palvelu. Palvelulla voi yhdistellä eri verkkolähteistä saatavaa materiaalia, ja sinne on luotu Kirjastovirta-niminen pipe. Kirjastovirta on yhdistelmä suomalaisista kirjastoaiheisista blogeista (tällä hetkellä lokakuussa 2011 59 lähdeä), ja tilaamalla Kirjastovirran RSS-syötteenä saa tiedon aina, kun jotakin näistä blogeista päivitetään. (Yahoo 2011b).

Suomalaisia blogeja eri aiheista voi etsiä Blogilista-palvelusta osoitteesta <http://www.blogilista.fi/>. Palvelussa blogeja on jaoteltu eri asiasanojen mukaan.

Palveluun voi lisätä blogeja kuka tahansa, ja kunkin blogin on asiansanoittanut sen lisääjä. Blogeja voi hakea sanahauulla, tai etsiä niitä asiansanojen mukaan. (Blogilista.fi 2011.) Esimerkiksi asiansanalla *satakunta* palvelusta löytyy 22 blogia, jotka edustavat hyvin vaihtelevia aiheita. Asiansanalla *kirjallisuus ja kirjoittaminen* puolestaan löytyy jopa yli 5000 blogia, mutta blogien esittelyjä pikaisesti silmäillessä huomaa, että aivan kaikki eivät suoranaisesti käsittele kirjallisuutta, tai eivät ainakaan ole pelkkiä kirjallisuusaiheisia blogeja.

## 6.2 Porin kaupungin ympäristöviraston julkaisut ja raportit

Tutkimuksessa haluttiin tarkastella tarkemmin jonkin viranomaisen tuottamien julkaisujen löytymistä ja säilyttämistä. Tällaiseksi tahoksi valikoitui Porin kaupungin ympäristövirasto. Valinta tapahtui Satakuntaosaston tietopalveluvirkailija Merja Mäkelän ehdotuksen perusteella.

Porin kaupungin ympäristöviraston tehtäviin kuuluu huolehtia ympäristöterveyteen, ympäristönsuojeluun ja rakennusvalvontaan liittyvistä asioista. Lisäksi sen yhteydessä toimii erillisenä taseyksikkönä Porilab-laboratorio, joka on Porin kaupungin elintarvike- ja ympäristölaboratorio. Myös jätehuolto ja jäteneuvonta kuuluvat sen toimintaan. (Porin kaupunki 2011b.)

Porin kaupungin ympäristöviraston verkossa luettavissa olevia julkaisuja on listattu Ympäristöviraston julkaisuja ja raportteja -verkkosivulla osoitteessa <http://www.pori.fi/ymparistovirasto/julkaisutjaraportit.html>. Sivun sisältää esimerkiksi ilmanlaadun vuosiraportteja, erilaisia tutkimusraportteja ja erilaisia toimintaohjelmia, kuten ilmasto-ohjelma ja kestävän kehityksen toimintaohjelma. (Porin kaupunki 2011c.)

Ympäristöviraston julkaisuista haastateltiin sähköpostitse ympäristöjohtaja Matti Lankiniemeä, jonka vastuulla on ympäristöviraston johtamisen lisäksi myös viraston tiedotus (Porin kaupunki 2011d). Haastattelun tarkoituksena oli selvittää ympäristöviraston julkaisupolitiikkaa julkaisujen valinnan, säilyttämisen ja tekijänoikeuksien sekä julkaisuista tiedottamisen suhteen.

Ensimmäinen haastattelukysymys koski ympäristöviraston verkkosivulle laitettavien julkaisujen valintaa. Haastateltavalta kysyttiin siis, millä perusteella verkkosivuille laitettavat julkaisut ja raportit valitaan, laitetaanko sinne kaikki julkaisut vai vain osa niistä. Vastauksena kysymykseen oli, että nykyisin verkkosivuille laitetaan kaikki raportit. Lisäksi suuri osa esitteistä laitetaan verkkosivuille myös.

Seuraava haastattelukysymys koski verkkosivuille laitettujen raporttien ja julkaisujen pysyvyyttä ja säilyttämistä. Haluttiin siis tietää, miten kauan julkaisut pysyvät esillä verkkosivuilla ja vaihtuvatko verkkosivulla olevat julkaisut vai pysyvätkö kerran sinne laitettut julkaisut siellä aina. Haastateltava kertoi, että julkaisujen pysyvyys harkitaan aina tapauskohtaisesti. Jos jonkin raportin uskotaan olevan mielenkiintoinen ja kysytty, sitä pidetään esillä verkkosivulle pidemmän aikaa. Pidemmän aikaa säilytettäviä ovat olleet esimerkiksi kasvillisuus- ja linnustoselvitykset. Toisaalta ympäristövirastolla on myös vuosittain toistuvia raportteja, joista ilmestyy uusi versio joka vuosi. Näitä pidetään esillä kahdesta kolmeen vuotta, ja vanhempia raportteja poistetaan aina uuden ilmestyessä. Vuosittain toistuvia raportteja ovat esimerkiksi ilmanlaaturaportit.

Seuraavaksi haastateltavalta kysyttiin, miten uusista julkaisuista ja raporteista tiedotetaan. Tiedotus hoituu haastateltavan mukaan niin, että uudet julkaisut ja raportit viedään valmistuttuaan tiedoksi ympäristölautakunnalle, ja näin ne tulevat julkisiksi. Joistakin julkaisuista voidaan järjestää erillisiä tiedotustilaisuuksia, joissa samalla kerrotaan raportin olevan luettavissa ympäristöviraston verkkosivuilla. Näitä julkaisuja ovat olleet esimerkiksi Kuntalaisen ympäristöopas sekä ilmastoraportti. Uusista julkaisuista ilmoitetaan myös aina ympäristöviraston verkkosivun uutispalstalla. Kaikki esitteet eivät ole kuitenkaan ympäristöviraston Julkaisut ja raportit -verkkosivulla, vaan joitakin esitteitä pidetään esillä niiden aiheeseen liittyvissä osioissa ympäristöviraston verkkosivuilla. Tällaisia ovat esimerkiksi luonnonsuojeluun ja luonnon virkistyskäyttöön liittyvät esitteet.

Lopuksi haastateltavalta kysyttiin vielä julkaisujen ja raporttien tekijänoikeuksista. Haluttiin tietää, ovatko julkaisujen ja raporttien tekijänoikeudet ympäristövirastolla vai niiden tekijöillä. Haastateltava kertoi, että tekijänoikeudet ovat pääsääntöisesti julkaisujen ja raporttien tekijöillä. Haastateltava ei kuitenkaan yksilöinyt tarkemmin, millaisissa tapauksissa tekijänoikeudet ovat ympäristövirastolla ja millaisissa tapauksissa tekijöillä.

Koska haastateltava kertoo, että kaikista uusista ympäristöviraston julkaisuista tiedotetaan aina ympäristöviraston verkkosivujen uutispalstalla, tätä palstaa seuraamalla voidaan saada tieto kaikista ilmestyvistä julkaisuista. Helpoiten näiden uutisten seuraaminen käy tilaamalla uutissivusta RSS-syöte, joka kertoo tilaajalle uusista uutisista aina niiden ilmestyttyä. Tilaaminen onnistuu menemällä ympäristöviraston verkkosivulle (osoitteeseen <http://www.pori.fi/ymparistovirasto/>), jossa heti etusivulla on ajankohtaista-osio ja RSS-syötteiden tilaamismahdollisuudesta kertova oranssi RSS-logo. Logoa napsauttamalla saa syötteet tilattua itselleen.

### 6.3 Pitkäaikaissäilytys ja Kansallinen digitaalinen kirjasto

Kansallinen digitaalinen kirjasto eli KDK on opetus- ja kulttuuriministeriön hanke, joka tähtää kulttuurin ja tieteen digitaalisten aineistojen tehokkaaseen ja laadukkaaseen hallintaan, jakeluun ja pitkäaikaissäilytykseen sekä niiden digitoinnin edistämiseen (Kansallinen digitaalinen kirjasto 2011b). KDK:sta on kerrottu enemmän tämän tutkimuksen luvussa 4.3.

Tavoitteena on, että KDK:n pitkäaikaissäilytysjärjestelmän ensimmäinen toteutusvaihe päättyisi vuoden 2014 lopussa, ja toinen toteutusvaihe alkaisi vuonna 2015. Toinen toteutusvaihe pitää sisällään pysyvän järjestelmän ylläpidon ja järjestelmän täysmittaisen käyttöönoton. Koko pitkäaikaissäilytysjärjestelmä olisi siis käytössä aikaisintaan vuonna 2016. (Kansallinen digitaalinen kirjasto 2011d.) KDK:n pitkäaikaissäilytysjärjestelmä on siis käytössä jo muutaman vuoden kuluttua tämän tutkimuksen suorittamisesta, joten enää tässä vaiheessa ei ole perusteltua suunnitella yksittäiselle kirjastolle omaa pitkäaikaissäilytysjärjestelmää. Tärkeää sen sijaan

on jo tässä vaiheessa tallentaa ja kuvailla aineisto tulevaa KDK:n järjestelmää silmällä pitäen, jotta aineisto saadaan sujuvasti siirrettyä KDK:n pitkäaikaissäilytysjärjestelmään.

### 6.3.1 Kansallisen digitaalisen kirjaston edustajan haastattelu

Tätä tutkimusta varten haastateltiin sähköpostitse digitaalisen aineiston pitkäaikaissäilyttämisestä maakuntakirjastojen näkökulmasta Kansalliskirjaston kirjastoverkkopalvelujen pääsuunnittelijaa Esa-Pekka Keskitaloa, joka on työskennellyt KDK:n pitkäaikaissäilytysjaostossa.

Haastattelun aluksi haastateltava halusi korostaa maakuntakirjastojen näkökulman tärkeyttä, sillä hänen mielestään yleiset kirjastot ovat olleen hieman sivussa pitkäaikaissäilytysasiassa ja muutenkin digitaalisten aineistojen hallinnan ja säilytyksen kehittämisessä. Hän korosti myös vastauksensa liittyvän nimenomaan kirjastojen omassa hallinnassa oleviin aineistoihin, sillä lisensoitujen aineistojen tilanne on toinen.

Ensimmäinen kysymys koski yksittäisen maakuntakirjaston roolia pitkäaikaissäilytyksen järjestämisessä. Haluttiin tietää, voiko yksittäinen kirjasto tehdä asian eteen jotain vai tullaanko palveluja tarjoamaan keskitetysti. Haastateltava kertoi, että tulevaisuudessa palveluja tarjotaan keskitetysti ja pitkäaikaissäilytykseen liittyvät asiat tullaan hoitamaan myös keskitetysti. Aineiston omistus ja vastuu siitä sen sijaan säilyvät edelleen aineiston ylläpitäjällä. Ylläpitäjälle tulee jäämään silti vielä paljon työtä, jota haastateltava kutsuu laadunvarmistukseksi.

Haastateltava kertoi, että tulevaisuuden pitkäaikaissäilytyspalvelut tulevat sisältämään turvallisen bittitason säilytyksen. Se tarkoittaa sitä, että huolehditaan itse datan säilymisestä korruptoimattomana. Sen lisäksi säilytykseen kuuluu käytettävyyden varmistamiseen liittyviä palveluita. Näihin voisivat kuulua esimerkiksi aineiston migraatio uuteen tiedostomuotoon sellaisessa tapauksessa, että aineistossa käytössä olevan tiedostomuodon,



esimerkiksi PDF:n, versio muuttuu tai jos koko PDF tiedostomuotona jää pois käytöstä.

Haastateltava haluaa korostaa, että kirjastoillekin jää vielä paljon tehtävää. Niiden pitäisi huolehtia siitä, että digitaalinen aineisto on säilyttämiskelpoista. Tämä tarkoittaa ensinnäkin sitä, että aineiston tulee olla tarpeeksi laadukasta. Jos kirjasto esimerkiksi digitoi itse materiaalia, tulee lopputuloksen laadukkuus ottaa huomioon jo digitoinnin suunnitteluvaiheessa. Toiseksi aineiston säilyttämiskelpoisuus tarkoittaa sitä, että aineistossa on oltava riittävästi kuvailutietoja sekä teknistä ja hallinnollista metadataa. Näitä olisi tarpeen kerätä jo riittävän varhaisesta vaiheesta alkaen. Kolmanneksi kirjaston tulisi määritellä säilytyspolitiikkansa: Mitä säilytetään? Millä aikavälillä?

Näiden tehtävien lisäksi kirjastolta vaaditaan myös kykyä käyttää keskitettyjä palveluita. Tämä pitää sisällään aineiston siirtämisen pitkäaikaissäilytysjärjestelmään sekä niitä koskevien säilyttämispäätösten ja toimeksiantojen tekemisen. Keskitettyjen palveluiden käyttö saattaa asettaa joitakin vaatimuksia tietojärjestelmille tai edellyttää esimerkiksi formaattikonversioita. Haastateltava kuitenkin huomauttaa, että näihin toimintoihin on suunnitteilla tukipalveluita pienemmille organisaatioille.

Haastateltava haluaa erityisesti tähdentää kokoelman ylläpitäjälle sitä, että pitkäaikaissäilytys pyrkii turvaamaan aineistojen säilymisen merkittävien muutosten, mullistusten ja vuosisatojen vaihtumisen yli. Silti hän toivoo jokaisen organisaation miettivän ja suunnittelevan arkipäiväiset järjestelmänsä niin, että aineiston säilyvyyttä ja käytettävyyttä mietittäisiin esimerkiksi viiden vuoden horisontilla. Haastateltava toivoisi organisaatioiden miettivän esimerkiksi aineistojen hallinnoin ja käyttöön asettamisen tapaa ja sitä, onko se skaalautuva, aineiston siirtämistä seuraavaan tietojärjestelmään ja miten se tehdään sekä sitä, onko metadata riittävää.

Haastateltava selventää vielä pitkäaikaissäilytysmallia käyttämällä käsitteitä *data provider* ja *service provider*, vapaasti suomennettuna datan tarjoaja ja palveluntarjoaja. Hän tarkoittaa näillä käsitteillä sitä, että pitää erottaa toisistaan

data (voidaan nimittää myös informaatioksi tai kokoelmaksi) ja siihen perustuva palvelu. Datan pitää olla huolellisesti hallittuna ja järjestelmällisesti kuvattuna omalla paikallaan, kun taas palveluja voidaan rakentaa eri aikoina ja eri tarpeita varten erilaisia. Hän vertaa tätä mallia KDK:n asiakasliittymään, jossa myös on erotettuna taustajärjestelmä (eli *data provider*) ja asiakasliittymä (eli *service provider*). Haastateltava tähdentää, että kokoelmanhallinta osana erilaisia palveluita, esimerkiksi kirjaston verkkosivuja, ei ole järkevää.

Lopuksi vastauksena maakuntakirjaston konkreettisten toimenpiteiden tekemisen mahdollisuudelle jo nyt haastateltava toteaa, että konkreettisia teknisiä vaatimuksia ja määrittämiä ei vielä ole saatavilla. Laatu- ja metadatakysymyksiin on joitakin vastauksia syntynyt KDK:n PAS-hankkeessa, mutta vielä niitä ei ole puettu julkaistavaan asuun.

### 6.3.2 Kansallisen digitaalisen kirjaston suosittelemat tiedostomuodot

KDK-hankkeelle on laadittu vuonna 2010 kokonaisarkkitehtuurin kuvaus, jossa kuvataan eri elementtien toimintaa kokonaisuutena. Näitä elementtejä ovat esimerkiksi organisaatioyksiköt, ihmiset, toimintaprosessit, tiedot ja tietojärjestelmät. Kokonaisarkkitehtuurin kuvauksen liitteeksi on laadittu standardisalkku, jossa määritellään kokonaisarkkitehtuurin kannalta keskeistä tietoa, kuten hankkeessa käytettävät standardit. (Kansallinen digitaalinen kirjasto 2010c, 2.)

Standardisalkun ja standardien määrittämisen tavoitteena on varmistaa KDK-järjestelmäkokonaisuuden toimivuus, jonka vuoksi eri organisaatioiden tallentaman datan on oltava yhteismitallista. Lisäksi standardointi mahdollistaa tulevaisuudessa palveluiden mahdollisimman vaivattoman siirtämisen uusiin laitteisto- ja ohjelmistoympäristöihin. Standardisalkussa määritellyt standardit on jaettu kolmeen ryhmään: Ne ovat joko välttämättömiä, suositeltavia tai hyödyllisiä. (Kansallinen digitaalinen kirjasto 2010d, 2.)

On tärkeää, että tiedostot toimitetaan pitkäaikaissäilytysjärjestelmään alkuperäisissä tiedostomuodoissaan, jotta tietosisällön säilyminen

muuttumattomana voidaan varmistaa. Tällöin mahdollinen migraatio tehdään vasta pitkäaikaissäilytysjärjestelmässä muutoksen dokumentoitavuuden vuoksi. Koska yleisesti käytetyt tiedostomuodot muuttuvat nopeasti, ei standardisalkussa voida listata tallennuskelpoisia tiedostomuotoja kattavasti. Joitakin suositeltavia standardeja on kuitenkin listattu sekä tekstiaineistolle, kuvatiedostoille, liikkuvalla kuvalla että audiotallenteille. (Kansallinen digitaalinen kirjasto 2010d, 3.)

Sovellettavat standardit eri aineistolle ovat

tekstiaineistolle SGML, XML, TEI, HTML, XHTML sekä PDF/A  
kuvatiedostoille TIFF, JPG sekä SPIFF  
liikkuvalla kuvalla MPEG-1, MPEG-2 tai MPEG-4 [Technical] sekä Motion  
JPEG 2000  
audiotallenteille Broadcast Wave Format eli BWF.

(Kansallinen digitaalinen kirjasto 2010d, 3.)

Seuraavassa on käsitelty tarkemmin joitakin edellä mainituista standardeista. Käsitteilyyn on valittu tämän tutkimuksen aihepiirin kannalta olennaisia standardeja.

PDF-tiedostomuoto (*Portable Document Format*) on Adobe Systemsin kehittämä tiedostomuoto. Sen keskeisin ominaisuus on sen eheys, sillä PDF-muotoiset tiedostot näyttävät samanlaisilta kuin alkuperäiset dokumentit riippumatta luomiseen käytetystä sovelluksesta. (Adobe 2011a.) PDF-standardeja on useampia, kuten PDF/A, PDF/E ja PDF/X. Erona standardeissa on niiden käyttötarkoitus: PDF/A:ta käytetään pitkäaikaiseen säilytykseen tarkoitetuissa dokumenteissa, PDF/E:tä tekniseen suunnitteluun liittyvissä dokumenteissa ja PDF/X:ää painotuotteiden valmistuksessa. (Adobe 2011b.) PDF/A-muodon keskeisiä ominaispiirteitä ovat esimerkiksi omavaraisuus ja dokumentoitavuus. Omavaraisuudella tarkoitetaan sitä, että kaikkien tiedoston tulostamiseen tai esittämiseen tarvittavien tietojen esimerkiksi fonteista tai väreistä on oltava tiedostossa itsessään mukana. Dokumentoitavuus

puolestaan tarkoittaa tiedostoa, joka kantaa metadataa eli tietoa dokumentista itsestään mukanaan. Lisäksi tiedoston salaaminen on kielletty. (Adobe 2011c.)

Tekstiaineiston standardeista SGML, XML, HTML ja XHTML liittyvät toisiinsa. SGML (*Standard Generalized Markup Language*) kehitettiin tekstin rakenteen erottamiseksi sisällöstä. SGML onkin merkkauskielten määrittelykieli. HTML (*HyperText Markup Language*) on SGML:llä määritelty merkkauskieli, jolla voidaan toteuttaa hypertekstidokumentteja. XML puolestaan on eräänlainen SGML:n karsittu versio, joka on kehitetty verkkokäyttöä varten poistamalla SGML:stä ominaisuuksia, jotka vaikeuttavat siirrettävyyttä verkossa ja ohjelmointia. (Parviainen 1999.) XHTML (*EXtensible HyperText Markup Language*) on uudelleenmuotoiltu HTML, joka toimii yhteistyössä XML:n kanssa (W3C 2002). Nämä kaikki standardit siis liittyvät dokumenttien esittämiseen verkkosivuilla.

TIFF (*Tagged Image File Format*) on tiedostomuoto, jota tyypillisesti käyttävät skannerit sekä piirto- ja kuvankäsittelyohjelmat (Adobe 1992, 4). TIFF-kuvat ovat bittikarttakuvia (The Digital Formats 2011). Bittikarttakuvissa kuva muodostuu pikseleistä, ja yksi pikseli vastaa yhtä kuvapistettä näytöllä. Bittikarttakuvia suurennettaessa niiden laatu yleensä heikkenee, sillä pikselien määrä ei kasva, vaan niiden koko suurenee. (Taideteollinen korkeakoulu 2011.) TIFF-tiedostopäätteitä ovat .tif ja .tiff (The Digital Format 2011b).

JPEG (*Joint Photographic Experts Group*) on häviöllisen kuvanpakkauksen menetelmä (W3C 2003.) Pakkaus pienentää kuvatiedoston kokoa, mutta samalla se heikentää kuvan laatua. Pakattaessa kuvasta poistetaan pieniä yksityiskohtia, joita ihmissilmä ei erota, kuten värien ja kirkkauden muutoksia. Pakkauksessa pyritään kuitenkin mahdollisimman alkuperäisen kaltaiseen lopputulokseen. (Leinonen 2010.) JPEG-muotoisen kuvan tiedostopäätte voi olla .jpg, .jpeg, .jpe, .jfi, .jfif, .j, jif tai .jmh. Yleisin näistä on .jpg-päätte. (The Digital Formats 2007.)

KDK:n pitkäaikaissäilytysjärjestelmään on suunnitteilla formaattikirjasto, joka sisältää tietoa tiedostomuodoista. Se on pitkäaikaissäilytysjärjestelmän

migraatio-ominaisuuksien perusta, sillä sen avulla voidaan selvittää migraation toteuttamiseen tarvittavat välineet ja siihen mahdollisesti liittyvät ongelmat. Tiedostomuodoista tallennetaan yleistä tietoa, kuten tiedostomuodon nimi, versio, tunniste, kuvaus sekä edeltävä tiedostomuoto ja seuraava tiedostomuoto. Yleisen tiedon lisäksi tallennetaan myös paikallista tietoa, kuten organisaation tiedostomuodon käyttölisenssin päättymispäivä. (Kansallinen digitaalinen kirjasto 2010e, 8.)

### 6.3.3 Kansallisen digitaalisen kirjaston suosittamat metadataformaattit

Metatieto eli metadata on tietoa tiedosta. Se on sekundaarista tietoa, jota käytetään primaaritiedon kuvailemiseen. Se on verkkodokumentin tai muunlaisen informaatioresurssin kuvaus, jonka avulla pyritään helpottamaan tiedon hakua ja paikallistamista sekä toisaalta löydettyjen dokumenttien arviointia: Ovatko ne relevantteja, käyttökelpoisia, luotettavia, laadukkaita, uskottavia, alkuperältään hyväksyttäviä ja saatavilla? Metadataa on sekä rakenteista että rakenteetonta. Rakenteettomalla metadatatalla ei ole ennalta määrättyä muotoa, kun taas rakenteinen metadata nojautuu ennalta sovittuun formaattiin. Tällaisia ovat esimerkiksi kirjastotietokannat, jotka perustuvat MARC-formaatteihin. Formaateissa on valmiiksi määritelty paikkansa erilaisille tiedoille, kuten tekijälle, kirjan nimelle, kustantajalle ja niin edelleen. Digitaalisissa dokumenteissa käytetään paljon Dublin Core -formaattia. (Tuominen 2008, 88–90.)

Pitkäaikaissäilytyksessä metadatatalla voidaan varmistaa tiedon oikeellisuus ja eheys. Sitä varten metadata jaetaan kuvailevaan ja hallinnolliseen metadataan. Hallinnolliseen metadataan kuuluvat esimerkiksi aineiston tekninen metadata, pitkäaikaissäilytyksen metadata sekä käyttöoikeuksiin tai -rajoituksiin liittyvät tiedot. KDK:n pitkäaikaissäilytysjärjestelmän puitteissa metadataan liittyy joitakin ongelmia. Eri organisaatioissa oleva digitaalinen tieto on varustettu metadatatalla hyvin vaihtelevasti, sillä on olemassa eri metadatatamutoja ja vakiintuneita käytäntöjä. Läheskään kaikkea metadataa ei ole aina saatavilla, varsinkin jos on kyse vanhasta tiedosta. Lisäksi metadatan lisääminen

aineistoon on työlästä ja vie paljon aikaa ja resursseja. (Kansallinen digitaalinen kirjasto 2010, 19.)

KDK:n kokonaisarkkitehtuurin kuvauksen liitteenä olevassa standardisalkussa määritellään standardeja, joita pitkäaikaissäilytysjärjestelmään tuotavan metadatan on suositeltavaa noudattaa. Standardisalkussa metadata jaetaan kuvailevaan ja hallinnolliseen metadataan. (Kansallinen digitaalinen kirjasto 2010d, 3.)

Kuvailevan metadatan formaatteja on useampia, sillä eri sektoreilla käytetään ja tullaan käyttämään yhtä tai useampaa metadataformaattia. KDK:n asiakasliitymässä metadata tullaan muuttamaan sovelluksen sisäiseen tallennusmuotoon, minkä vuoksi on tärkeää, että kuvailuperiaatteet ja formaattien soveltamistavat ovat riittävän yhtenäiset. (Kansallinen digitaalinen kirjasto 2010d, 4.)

Seuraavat standardisalkussa esitetyistä kuvailevan metadatan formaateista soveltuvat kirjastojen aineistolle:

- MARC 21
- FINMARC (siirtymäkauden ajan)
- Dublin Core

(Kansallinen digitaalinen kirjasto 2010d, 4.)

MARC (*Machine Readable Cataloging*) on kehitetty Kongressin kirjastossa Yhdysvalloissa 1960-luvulla luettelokortteja lukevaa ohjelmaa varten. Aikojen kuluessa siitä on kehitetty erilaisia kansallisia versioita, joista Suomessa käytössä oleva FINMARC on yksi. MARC 21 on kansainvälinen formaatti, jonka nimessä oleva 21 tarkoittaa 21. vuosisadan formaattia. (Kansalliskirjasto 2011a.) Suomessa Voyager-kirjastot käyttävät jo MARC 21 -formaattia, ja yleisten kirjastojenkin pitäisi siirtyä käyttämään MARC 21 -formaattia vuoteen 2014 mennessä (Kansalliskirjasto 2011c.)

MARC-formaatin mukainen tietue koostuu nimiöstä, hakemistosta sekä kiinteä- ja vaihtuvamittaisista kentistä. MARC-formaatilla on sekä pysty- että

vaakasuora rakenne. Tämä tarkoittaa sitä, että pystysuorassa rakenteessa voidaan erottaa eri kenttäryhmiä, joiden kolminumeroinen kenttäkoodin ensimmäinen numero on sama, kun taas vaakasuorassa rakenteessa kenttäkoodin kaksi jälkimmäistä numeroa kertovat tarkemmin kentän sisältämästä tiedosta. (Kansalliskirjasto 2011b.)

Dublin Core on erityisesti digitaalisten julkaisujen kuvailuun kehitetty formaatti (Kansalliskirjasto 2011d.) Sen kehitystyö aloitettiin vuonna 1995 tarkoituksena tehostaa verkkotallenteiden hakua. Formaatin mukainen tietue koostuu 15 kentästä. Jokainen kenttä on vapaaehtoinen ja tarvittaessa toistettavissa. Lisäksi useimmilla kentillä on tarkenteita, joiden avulla kentän merkitystä pystytään tarkentamaan. Formaatin tavoitteina on kuvailutietojen luomisen ja ylläpidon yksinkertaisuus, yleisesti ymmärrettävä semantiikka, kansainvälisyys sekä laajennettavuus. Laajennettavuudella tarkoitetaan sitä, että vaikka formaatti halutaan perusteiltaan pitää yksinkertaisena, sitä on mahdollista laajentaa, jos halutaan spesifimpää kuvailua. (Stenvall 2002, 2–3.)

Sisällönkuvailussa käytettäviksi välineiksi standardisalkku suosittelee kotimaisia tesauuksia tai ontologioita. Tesauukset ja ontologiat voivat olla yleisluontoisia, kuten esimerkiksi YSA, YSO, Kaunokki ja MUSA, tai tieteenalakohtaisia, kuten FinMeSH, Eduskunnan kirjaston asiasanasto tai Kula – Kulttuurien tutkimuksen asiasanasto. Näiden ohella voidaan täydennykseksi soveltaa kansainvälisiä järjestelmiä, mikäli aihealue niin vaatii. (Kansallinen digitaalinen kirjasto 2010d, 5, 9–10.)

Hallinnollisen metadatan tuottamiseen ei vielä ole vakiintuneita käytäntöjä, sillä kyseessä on aivan uusi toiminta-alue. Hallinnolliseen metadataan kuuluvan teknisen metadatan kuvailuun standardisalkku suosittelee seuraavia formaatteja:

- kuville MIX (Metadata for images in XML)
- tekstille textMD (technical metadata for text)
- äänelle AudioMD
- liikkuvalla kuvalla VideoMD.

(Kansallinen digitaalinen kirjasto 2010d, 5.)

MIX (*Metadata for Images in XML*) on malli, jonka avulla saadaan talletettua kuviin liittyvää tärkeää tietoa, jota tarvitaan kuvien esittämisessä ja tulkitsemisessa sekä nykyhetkessä että tulevaisuudessa. Tekninen metadatahan on ei-kuvailevaa tietoa tiedoston teknisistä ominaisuuksista, ja kuvien yhteydessä sillä tarkoitetaan esimerkiksi tietoja kuvanottohetkestä, bittien määrästä pikseliä kohden, käytetystä väriavaruudesta, valotuksesta ja kuvaformaatin erikoispiirteistä. (Guenther 2007, 16.)

TextMD on puolestaan teknisen metadatan esittämisen malli tekstipohjaisille digitaalisessa muodossa oleville tiedoille. Sen avulla talletetaan yksityiskohtaista tietoa tiedoston ominaisuuksista, kuten esimerkiksi koodaustiedoista, käytetystä merkistöstä, kielestä, fonteista, merkkauksesta, tulostukseen ja katseluun liittyvistä teknisistä vaatimuksista sekä sivujärjestyksestä. (Library of Congress 2011.)



## 7 JOHTOPÄÄTÖKSET

Tämän työn tarkoituksena oli pyrkiä selventämään maakuntakokoelmien digitaalisten aineistojen hallintaa kahdesta eri näkökulmasta. Ensimmäisenä tavoitteena oli tutkia sitä, kuinka maakuntakokoelman hoitaja löytää verkossa olevaa aineistoa. Toisena tavoitteena oli tutkia sitä, kuinka tämän digitaalisessa muodossa olevan aineiston pitkäaikaissäilytystä voidaan edistää ja kuinka voidaan varautua Kansallisen digitaalisen kirjaston eli KDK:n tulevaan pitkäaikaissäilytysjärjestelmään.

Tutkimuksessa haastateltiin seitsemää maakuntakirjaston edustajaa heidän keinoistaan löytää maakuntakokoelmaan soveltuvaa aineistoa verkosta sekä heidän ajatuksistaan ja mahdollisista tehdyistä pitkäaikaissäilytystoimenpiteistä. Haastatteluissa kävi ilmi, että monessakaan maakuntakirjastossa ei tehdä laajempia systemaattisia hakuja maakuntakokoelmaan soveltuvan aineiston löytämiseksi, vaan kokoelmaan saadaan aineistoa enemmänkin sattumalta. Aineistoa löytyy myös muiden tiedonhakujen yhteydessä. Syynä tähän sattumanvaraisuuteen on pääasiassa resurssipula, sillä painetussa muodossa olevan aineiston käsittely vie oman aikansa, eikä aikaa digitaalisessa muodossa olevan aineiston etsintään ole.

Erilaisten verkossa saatavilla olevien palveluiden käyttäminen tiedon hankinnan ja seurannan apuvälineinä ei noussut maakuntakirjastojen haastatteluissa esille. Joitakin silti mainittiin, tällaisena esimerkkinä mainittakoon Google Alerts -palvelu, joka suorittaa säännöllisesti tiedonhakuja halutusta aiheesta ja ilmoittaa haun tuloksista sähköpostitse säännöllisin väliajoin. Vastaavia palveluita olisi kannattavaa käyttää enemmänkin, sillä ne helpottavat verkossa olevan aineiston kartoittamista ja säästävät kokoelman hoitajan aikaa.

Näiden maakuntakirjastojen edustajien haastattelujen lisäksi haastateltiin Porin kaupungin ympäristöviraston johtajaa Matti Lankiniemeä. Haastattelu käsitteli ympäristöviraston verkkosivuilla olevia julkaisuja ja raportteja. Haastattelun tarkoituksena oli selvittää, kuinka kirjasto pysyisi ajan tasalla ympäristöviraston julkaisuista, sillä ne kaikki ovat maakuntakokoelmaan soveltuvaa aineistoa.

Haastattelussa kävi ilmi, että julkaisuista ilmoitetaan ympäristöviraston verkkosivulla, joten sitä sivua seuraamalla voidaan pysyä perillä uusien julkaisujen ilmestymisestä.

Maakuntakirjastojen edustajien haastattelujen pitkäaikaissäilytystä käsittelevästä osiosta nousi selvästi esille se, että monet eivät ole vielä tehneet juurikaan ratkaisuja aineiston pitkäaikaissäilytykselle. Joitakin ratkaisuja ja mietintöjä toki on tehty, mutta pysyviä ratkaisuja ei ole. Monet ovat sen sijaan odottavalla kannalla ja ennen kaikkea KDK:n ratkaisuja odotellaan.

KDK:n edustajan Esa-Pekka Keskitalon haastattelussa korostui kirjaston rooli aineistojen laadun varmistajana. Vaikka KDK:n tuleva pitkäaikaissäilytysratkaisu on keskitetty, jää kirjastoillekin vielä tehtäviä. Niiden tulee huolehtia aineiston laadukkuudesta itse digitoitaessa sekä sen säilytuskelpoisuudesta eli esimerkiksi siitä, että aineistolle on tallennettu riittävästi metadataa. Tämän vuoksi tässä tutkimuksessa onkin kartoitettu KDK:n edellyttämiä tiedosto- ja metadatamuotoja, jotta kirjasto osaisi varautua jo nyt tulevaan pitkäaikaissäilytysratkaisuun.

Ongelmalliseksi tulevaan pitkäaikaissäilytysratkaisuun varautumisen tekee se, että tarkkoja tietoja ratkaisusta ei vielä ole saatavilla. KDK-hankkeen pitkäaikaissäilytyksen jatkohanke, jonka tarkoituksena on tuottaa pitkäaikaissäilytyksen yksityiskohtainen toteutussuunnitelma, saa työnsä päätökseen vasta vuonna 2013 (Kansallinen digitaalinen kirjasto 2011c). Vasta tällöin voidaan todella nähdä, mitä keskitetty pitkäaikaissäilytysratkaisu tulee kirjastoilta vaatimaan, siihen saakka ollaan vuonna 2010 toimintansa päättäneen pitkäaikaissäilytysjaoston loppuraportin ja sen liitteiden varassa (Kansallinen digitaalinen kirjasto 2011e).

Tämän tutkimuksen tekeminen osoitti, että pitkäaikaissäilytyksestä teoreettisessa mielessä on kyllä kirjoitettu melko paljon, mutta käytännön ratkaisuja on vaikea löytää. Varsinkin yksittäisen organisaation näkökulma, tällainen olisi esimerkiksi yksittäinen kirjasto, puuttuu kirjallisuudesta. Tämä johtunee suoraan käytännön kokemusten puutteesta, sillä koko digitaalisten

aineistojen maailma on vielä kuitenkin kovin nuori: sekä aineistojen löytyminen, niiden hallinta että pitkäaikaisesti säilyttäminen ovat uutta sekä kirjastolaisille että muillekin niiden kanssa työskenteleville. Onneksi varsinkin aineistojen löytymistä helpottamaan kehitellään jatkuvasti uusia palveluita ja keinoja, joiden kehitystä kannattaa seurata oman työn helpottamiseksi.

## LÄHTEET

- 23 asiaa 2009a. RSS-ohjelman käyttöönotto. Viitattu 15.10.2011.  
<http://www.23asiaa.net/asiat/2009/rss-lukuohjelman-kayttoonotto>
- 23 asiaa 2009b. Mikä on wiki? Viitattu 19.10.2011. <http://www.23asiaa.net/asiat/2009/mika-wiki>
- Adobe 1992. TIFF Revision 6.0. Mountain View: Adobe Systems Incorporated. Viitattu 20.10.2011. Saatavilla myös <http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf>
- Adobe 2011a. Adobe PDF:n historia. Viitattu 20.10.2011.  
<http://www.adobe.com/fi/products/acrobat/adobepdf.html>
- Adobe 2011b. Acrobat ja PDF-standardit. Viitattu 20.10.2011.  
<http://www.adobe.com/fi/products/acrobat/standards.html>
- Adobe 2011c. Adobe PDF/A archiving standard: Key documentation. Viitattu 20.10.2011.  
<http://www.adobe.com/enterprise/standards/pdfa/>
- Alatalo, M. & Åkerman, M. 2010. Asiantuntijahaastattelun analyysi: faktojen jäljillä. Teoksessa Ruusu vuori, J. & Nikander, P. & Hyvärinen, M. (toim.) Haastattelun analyysi. Tampere: Vastapaino. 372–392.
- Alaterä, A. & Halttunen, K. 2002. Tiedonhaun perusteet – osa lukutaitoa. Helsinki: BTJ Kirjastopalvelu.
- Ampparit 2011. Lähteet. Viitattu 15.10.2011. <http://www.ampparit.com/tietoa/lahteet>
- Ask 2011. Ask Search Technology. Viitattu 12.10.2011.  
[http://sp.ask.com/en/docs/about/ask\\_technology.shtml](http://sp.ask.com/en/docs/about/ask_technology.shtml)
- AtomEnabled 2007. AtomEnabled Everyone. Viitattu 15.10.2011.  
<http://www.atomenabled.org/everyone/>
- Blinnikka, V. 1987. Kirjastoaineiston luokitus ja luettelointi. Teoksessa Blinnikka, V., Holmberg, I. & Lehväslaiho, V. (toim.) Kirjastotyön perusteet. Helsinki: Suomen kirjastoseura.
- Blogilista.fi 2011. Ohjeita. Viitattu 20.10.2011. <http://www.blogilista.fi/sivu/ohjeita>
- Caplan, P. 2008a. What Is Digital Preservation? Library Technology Reports. February/March 2008, Vol. 44 Issue 2, s. 7–9. Viitattu 16.11.2011.  
<http://search.ebscohost.com.ezproxy.turkuamk.fi/login.aspx?direct=true&db=lxh&AN=29439252&site=ehost-live>
- Caplan, P. 2008b. Preservation Practices. Library Technology reports. February/March 2008, Vol 44. Issue 2, s. 10–13. Viitattu 16.11.2011.  
<http://search.ebscohost.com.ezproxy.turkuamk.fi/login.aspx?direct=true&db=lxh&AN=29439253&site=ehost-live>
- The Digital Formats 2007. Sustainability of Digital Formats: Planning for Library of Congress Collections: JFIF JPEG File Interchange Format. Viitattu 20.10.2011.  
<http://www.digitalpreservation.gov/formats/fdd/fdd000018.shtml>
- The Digital Formats 2011. Sustainability of Digital Formats: Planning for Library of Congress Collections: TIFF, Revision 6.0. Viitattu 20.10.2011.  
<http://www.digitalpreservation.gov/formats/fdd/fdd000022.shtml>
- Dogpile 2011. About Dogpile. Viitattu 13.10.2011.  
<http://www.dogpile.com/info.dogpl.t3.2/support/aboutus>

Espoon kaupunki 2010. Esitys opetus- ja kulttuuriministeriölle Espoon kaupunginkirjaston maakuntakirjastotehtävistä luopumiseksi. Viitattu 18.9.2011. [http://www.espoo.fi/asiakirja.asp?path=1;31;37423;37424;37425&id=DEC962FE089D7B4FC22577E6003CFC00&Kanta=Kunnari\intrakun\\_e.nsf](http://www.espoo.fi/asiakirja.asp?path=1;31;37423;37424;37425&id=DEC962FE089D7B4FC22577E6003CFC00&Kanta=Kunnari\intrakun_e.nsf)

Google 2011a. Google ilmoitukset. Viitattu 12.10.2011. <http://www.google.com/alerts>

Google 2011b. Tietoja Google Scholarista. Viitattu 12.10.2011. <http://scholar.google.fi/intl/fi/scholar/about.html>

Google 2011c. Mitä näet tehdessäsi hakuja Google-teoshaulla. Viitattu 12.10.2011. <http://books.google.com/intl/fi/googlebooks/screenshots.html>

Google 2011d. Google Alerts Help: Getting Started Guide. Viitattu 16.10.2011. <http://www.google.com/support/alerts/bin/static.py?page=guide.cs&guide=28413&topic=28417>

Google 2011e. Google Alerts Help: Getting Started Guide: Tips. Viitattu 17.10.2011. <http://www.google.com/support/alerts/bin/static.py?page=guide.cs&guide=28413&topic=28416&answer=175927>

Guenther, R. 2007. MIX. What it stands for: Metadata for Images in XML Schema. *Computers in Libraries*. April 2007, Vol. 27 Issue 4, s. 16. Viitattu 21.10.2011. <http://web.ebscohost.com.ezproxy.turkuamk.fi/ehost/pdfviewer/pdfviewer?sid=2c882431-12b0-46f1-9a5c-347f16b18d95%40sessionmgr115&vid=1&hid=113>

Haasio, A. 2007. Internet-tiedonhaun teho-opas. Helsinki: BTJ Finland.

Haasio, A. 2009a. Johdon assistentin tiedonhankinnan opas. Helsinki: WSOYpro.

Haasio, A. 2009b. Tiedon avain – opas tiedonlähteille. Helsinki: BTJ Kustannus.

Hakala, J. & Stenvall, J. 2002. Digitaalisten aineistojen säilyttäminen. Teoksessa Kekki, K. & Salminen, O. (toim.) *Digimaan kartta – Puheenvuoroja digitaaliseen tietohuoltoon*. Helsinki: Opetusministeriö. 84–93.

Harjavallan kaupunginkirjasto 2011. Kotiseutukokoelma. Viitattu 18.9.2011. <http://www.harjavalta.fi/palvelut/kirjasto/kotiseutukokoelma/>

Hirsjärvi, S. & Hurme, H. 2004. Tutkimushaastattelu – Teemahaastattelun teoria ja käytäntö. Helsinki: Yliopistopaino.

Heikkinen, R., Agander, A., Ijäs, E. & Laitinen, M. 2005. Tiedonhakijan teho-opas. 6. uudistettu painos. Jyväskylä: Docendo.

Helsingin kaupunginkirjasto – yleisten kirjastojen keskuskirjasto 1990. Maakuntakokoelmista alueellisiksi tietokeskuksiksi: Aluetietokantatyöryhmän raportti. Helsinki.

Hintikka, K. A. 2007. Web 2.0 – johdatus internetin uusiin liiketoimintamahdollisuuksiin. TIEKE Tietoyhteiskunnan kehittämiskeskus ry:n julkaisusarja 28. Helsinki: TIEKE Tietoyhteiskunnan kehittämiskeskus ry. Viitattu 15.10.2011. Saatavilla [http://www.tieke.fi/mp/db/file\\_library/x/IMG/20815/file/julkaisu\\_28.pdf](http://www.tieke.fi/mp/db/file_library/x/IMG/20815/file/julkaisu_28.pdf)

Huittisten kaupunginkirjasto 2011. Kotiseutukokoelmat. Viitattu 18.9.2011. <http://www.huittinen.fi/Default.aspx?id=564>

Hämeenlinnan kaupunginkirjasto 2011. Maakuntakirjastotoiminta. Viitattu 17.9.2011. <http://hameenlinna.verkkokirjasto.fi/web/arena/maakuntakirjasto>

Kansallinen digitaalinen kirjasto 2010a. Pitkäaikaissäilytys-jaosto: Loppuraportti. Helsinki. Viitattu 20.10.2011. Saatavilla myös [http://www.kdk2011.fi/images/stories/PAS\\_loppuraportti\\_v.1.0.pdf](http://www.kdk2011.fi/images/stories/PAS_loppuraportti_v.1.0.pdf)

Kansallinen digitaalinen kirjasto 2010b. Pitkäaikaissäilytys-jaoston loppuraportin liite: Pitkäaikaissäilytysjärjestelmän toiminnallisuus ja toteutus. Helsinki. Viitattu 20.10.2011. Saatavilla myös [http://www.kdk2011.fi/images/stories/KDK\\_PAS\\_Toiminnallisuus\\_ja\\_toteutus\\_v1.01.pdf](http://www.kdk2011.fi/images/stories/KDK_PAS_Toiminnallisuus_ja_toteutus_v1.01.pdf)

Kansallinen digitaalinen kirjasto 2010c. Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuri. Helsinki. Viitattu 20.10.2011. Saatavilla myös [http://www.kdk.fi/images/stories/KDK\\_kokonaisarkkitehtuuri.pdf](http://www.kdk.fi/images/stories/KDK_kokonaisarkkitehtuuri.pdf)

Kansallinen digitaalinen kirjasto 2010d. Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuri Liite B: Standardisalkku. Helsinki. Viitattu 20.10.2011. Saatavilla myös [http://www.kdk.fi/images/stories/KDK\\_standardisalkku.pdf](http://www.kdk.fi/images/stories/KDK_standardisalkku.pdf)

Kansallinen digitaalinen kirjasto 2010e. Pitkäaikaissäilytysjaosto: PAS-palvelukuvaus. Helsinki. Viitattu 21.10.2011. Saatavilla myös [http://www.kdk.fi/images/stories/KDK\\_PAS\\_palvelukuvaus\\_v1.0.pdf](http://www.kdk.fi/images/stories/KDK_PAS_palvelukuvaus_v1.0.pdf)

Kansallinen digitaalinen kirjasto 2011a. Sanasto. Viitattu 7.11.2011. <http://www.kdk.fi/fi/kokonaisarkkitehtuuri/sanasto>

Kansallinen digitaalinen kirjasto 2011b. Tietoa hankkeesta. Viitattu 22.9.2011. <http://www.kdk.fi/fi/tietoa-hankkeesta>

Kansallinen digitaalinen kirjasto 2011c. PAS-jatkohanke. Viitattu 22.9.2011. <http://www.kdk.fi/fi/pitkaaikaissailytys/pas-jatkohanke>

Kansallinen digitaalinen kirjasto 2011d. Mikä PAS? Kysymyksiä ja vastauksia. Viitattu 16.11.2011. <http://www.kdk.fi/fi/pitkaaikaissailytys/maeaerittely-ja-dokumentit/98-mikae-pas-10-kysymystae>

Kansallinen digitaalinen kirjasto 2011e. Loppuraportti. Viitattu 16.11.2011. <http://www.kdk.fi/fi/pitkaaikaissailytys/maeaerittely-ja-dokumentit/99-loppuraportti>

Kansalliskirjasto 2011a. MARCin ja FINMARCin historiaa. Viitattu 21.10.2011. <http://www.kansalliskirjasto.fi/kirjastoala/formaatit/historiaa.html>

Kansalliskirjasto 2011b. Mikä on MARC? Viitattu 21.10.2011. <http://www.kansalliskirjasto.fi/kirjastoala/formaatit/yleista.html>

Kansalliskirjasto 2011c. MARC 21 -formaattiin siirtyminen. Viitattu 21.10.2011. <http://www.kansalliskirjasto.fi/kirjastoala/formaatit/aikataulu.html>

Kansalliskirjasto 2011d. Dublin Core. Viitattu 21.10.2011. <http://www.kansalliskirjasto.fi/julkaisuuala/dublincore.html>

Kirjastoasetus 18.12.1998/1078. Viitattu 17.9.2011. Saatavilla myös <http://www.finlex.fi/fi/laki/ajantasa/1998/19981078>

Kirjastot.fi 2011a. Linkkikirjaston ohjeet. Viitattu 14.10.2011. <http://www.kirjastot.fi/fi-FI/linkkikirjasto/ohjeet/>

Kirjastot.fi 2011b. Tiedote: Muutos maakuntakirjastojen toiminta-alueissa Uudellamaalla. Viitattu 17.9.2011. <http://www.kirjastot.fi/fi-FI/ajankohtaista/tiedote/ajankohtaista/muutos-maakuntakirjastojen-toiminta-alueissa-uudellamaalla>

- Kokkolan kaupunginkirjasto 2011. Kokkolan kaupunginkirjaston historia. Viitattu 17.9.2011. <http://lib.kokkola.fi/historia/maakuntakirjasto.htm>
- Lavian kunnankirjasto 2011. Kotiseutukokoelma. Viitattu 18.9.2011. [http://kirjasto.lavia.fi/lavia\\_kirjasto/sivu.tmp?siivu\\_id=2005](http://kirjasto.lavia.fi/lavia_kirjasto/sivu.tmp?siivu_id=2005)
- Leinonen E. 2010. Digitaalisen median perusteet 2010. Viitattu 20.10.2011. <http://www.tol.oulu.fi/kurssit/dmp/eeva/bittikartta.html>
- Library of Congress 2011. textMD – Technical Metadata for Text Official Web Site. Viitattu 21.10.2011. <http://www.loc.gov/standards/textMD/>
- Liikearkistoyhdistys 2009. Vuodesta sataan - sähköisten asiakirjojen hallinta ja säilyttäminen. Liikearkistoyhdistys ry:n julkaisuja 18. Helsinki.
- Makupalat 2011a. Makupalat-etusivu. Viitattu 14.10.2011. <http://www.makupalat.fi/>
- Makupalat 2011b. Makupalat Info. Viitattu 14.10.2011. <http://www.makupalat.fi/fi-FI/Info/>
- Mäenpää, M. 2002. Sähköisen aineiston pitkäaikaissäilytys. Teoksessa Kekki, K. & Salminen, O. (toim.) Digimaan kartta – Puheenvuoroja digitaaliseen tietuhooltoon. Helsinki: Opetusministeriö. 94–100.
- Mäkinen, I. 2009. Yleinen kirjasto hyvinvointiyhteiskunnassa 1960-luvun alusta vuosituhannen vaihteeseen. Teoksessa Mäkinen, I. (toim.) Suomen yleisten kirjastojen historia. Helsinki: BTJ Kustannus.
- Opetus- ja kulttuuriministeriö 2011a. Kirjastojen tehtävät ja työnjako. Viitattu 17.9.2011. [http://www.minedu.fi/OPM/Kirjastot/kirjastoverkosto/tehtavat\\_ja\\_tyonjako/?lang=fi](http://www.minedu.fi/OPM/Kirjastot/kirjastoverkosto/tehtavat_ja_tyonjako/?lang=fi)
- Opetus- ja kulttuuriministeriö 2011b. Perustietopaketti yleisistä kirjastoista. Viitattu 28.9.2011. [http://www.okm.fi/export/sites/default/OPM/Kirjastot/kirjastoverkosto/Liitteet/Kirjastopaketti\\_2011.ppt](http://www.okm.fi/export/sites/default/OPM/Kirjastot/kirjastoverkosto/Liitteet/Kirjastopaketti_2011.ppt)
- Opetusministeriö 2008. Sähköisen aineiston pitkäaikaissäilytystä ja käyttöä koskevan työryhmän muistio. Opetusministeriön työryhmämuistioita ja selvityksiä 2008:2. Helsinki. Viitattu 22.9.2011. [http://www.minedu.fi/OPM/Julkaisut/2008/Sahkoisen\\_aineiston\\_pitkaaikaissailytys\\_ja\\_kaytto.html](http://www.minedu.fi/OPM/Julkaisut/2008/Sahkoisen_aineiston_pitkaaikaissailytys_ja_kaytto.html)
- Parviainen, J. 1999. SGML-, HTML- ja XML-kielten eroja ja yhtäläisyyksiä. Viitattu 20.10.2011. <http://www.tml.tkk.fi/Studies/Tik-110.300/1998/Essays/merkkkauskielet.html>
- Peltonen, J. 1987. Paikallis- ja aluekokoelmat. Teoksessa Blinnikka, V., Holmberg, I. & Lehtslaiho, V. (toim.) Kirjastotyön perusteet. Helsinki: Suomen kirjastoseura.
- Pomarkun kunnankirjasto 2011. Pomarkun kotiseutukokoelma. Viitattu 18.9.2011. <http://213.192.129.17/sivu.aspx?taso=2&id=73>
- Porin kaupunki 2011a. Tietoa sivustosta. Viitattu 16.10.2011. <http://www.pori.fi/tietoasivustosta.html>
- Porin kaupunki 2011b. Ympäristöviraston organisaatio. Viitattu 17.10.2011. <http://www.pori.fi/ymparistovirasto/organisaatio.html>
- Porin kaupunki 2011c. Ympäristöviraston julkaisuja ja raportteja. Viitattu 17.10.2011. <http://www.pori.fi/ymparistovirasto/julkaisutjaraportit.html>
- Porin kaupunki 2011d. Yhteystiedot, koko virasto. Viitattu 17.10.2011. <http://www.pori.fi/ymparistovirasto/yhteystiedot.html>

- Porin kaupunginkirjasto 1946. Kertomus toiminnasta vuodelta 1945. Pori.
- Porin kaupunginkirjasto 1961. Porin kaupunginkirjaston toimintakertomus vuodelta 1960. Pori.
- Porin kaupunginkirjasto 1971. Porin kaupunginkirjaston toimintakertomus v. 1970. Pori.
- Porin kaupunginkirjasto – Satakunnan maakuntakirjasto 2011a. Satakuntakokoelma. Viitattu 28.9.2011. <http://www.pori.fi/kirjasto/satakunta/satakuntakokoelma.html>
- Porin kaupunginkirjasto – Satakunnan maakuntakirjasto 2011b. Maakuntakirjasto: toimintakertomus 2010. Viitattu 28.9.2011. Saatavilla [http://www.pori.fi/material/attachments/kirjasto/maakuntakirjasto/619CVePon/Mk-toiminta\\_2010.pdf](http://www.pori.fi/material/attachments/kirjasto/maakuntakirjasto/619CVePon/Mk-toiminta_2010.pdf)
- Porvoon kaupunginkirjasto 2011. Maakuntakirjasto. Viitattu 17.9.2011. <http://www.porvoo.fi/fi/palvelut/kirjasto/maakuntakirjasto>
- Rauman kaupunginkirjasto 2011. Rauma-kokoelma. Viitattu 18.9.2011. [http://www.rauma.fi/kirjasto/rauma\\_kokoelma.htm](http://www.rauma.fi/kirjasto/rauma_kokoelma.htm)
- Satakuntaliitto 2011. Satakunnan kunnat. Viitattu 27.9.2011. <http://www.satakuntaliitto.fi/sivu.aspx?taso=1&id=156>
- Search Engine Watch 2007. Major Search Engines and Directories. Viitattu 12.10.2011. <http://searchenginewatch.com/article/2048976/Major-Search-Engines-and-Directories>
- Suomen standardisoimisliitto SFS 2011a. Mikä on standardi? Viitattu 7.11.2011. [http://www.sfs.fi/standardisointi/tietoa\\_standardeista/mika\\_standardi/](http://www.sfs.fi/standardisointi/tietoa_standardeista/mika_standardi/)
- Suomen standardisoimisliitto SFS 2011b. Standardien tarkoitus ja hyödyt. Viitattu 7.11.2011. [http://www.sfs.fi/standardisointi/tietoa\\_standardeista/tarkoitus/](http://www.sfs.fi/standardisointi/tietoa_standardeista/tarkoitus/)
- Stenvall, J. 2002. Dunlin Core -formaatin käyttöopas. Helsinki: Helsingin yliopiston kirjasto. Viitattu 21.10.2011. Saatavilla myös [http://www.kansalliskirjasto.fi/julkaisuuala/dublincore/Files/liitetiedosto2/dc\\_opas.pdf](http://www.kansalliskirjasto.fi/julkaisuuala/dublincore/Files/liitetiedosto2/dc_opas.pdf)
- Suomen yleisten kirjastojen tilastot 2011. Perustilastot. Viitattu 17.9.2011. <http://tilastot.kirjastot.fi/fi-fi/perustilastot.aspx?AreaKey=Y2010T1N1>
- Suomi.fi 2011a. Tietoa portaalista. Viitattu 16.10.2011. [http://www.suomi.fi/suomifi/suomi/yleiset/tietoa\\_portaalista/index.html](http://www.suomi.fi/suomifi/suomi/yleiset/tietoa_portaalista/index.html)
- Suomi.fi 2011b. Kunnat ja kunnallishallinto. Viitattu 16.10.2011. [http://www.suomi.fi/suomifi/suomi/valtio\\_ja\\_kunnat/kunnat\\_ja\\_kunnallishallinto/index.html](http://www.suomi.fi/suomifi/suomi/valtio_ja_kunnat/kunnat_ja_kunnallishallinto/index.html)
- Taideteollinen korkeakoulu 2011. Digitaalinen kuvankäsittely: Bittikartta ja vektorigrafiikka, resoluutio. Viitattu 20.10.2011. <http://www.uiah.fi/mto/digitalimage/002.htm>
- Tiittula, L., Rastas, A. & Ruusuvuori, J. 2005. Kasvokkaisesta vuorovaikutuksesta tietokonevälitteiseen viestintään - Virtuaalihaastattelun näkymiä. Teoksessa Ruusuvuori, J. & Tiittula, L. (toim.) Haastattelu - Tutkimus, tilanteet ja vuorovaikutus. Tampere: Vastapaino. 264–271.
- Tiittula, L. & Ruusuvuori, J. 2005. Johdanto. Teoksessa Ruusuvuori, J. & Tiittula, L. (toim.) Haastattelu - Tutkimus, tilanteet ja vuorovaikutus. Tampere: Vastapaino. 9–21.
- Tuominen, K. 2008. Tiedon partaalla – Kuinka hallita informaatiotulvaa. 2. uud. ja täyd. laitos. Helsinki: BTJ Kustannus.



Tähtinen, M. 2008. Sivut järjestykseen. Tietokone 5/2008. Viitattu 12.10.2011. Saatavilla [http://www.tietokone.fi/lehti/tietokone\\_5\\_2008/sivut\\_jarjestykseen\\_872](http://www.tietokone.fi/lehti/tietokone_5_2008/sivut_jarjestykseen_872)

W3C 2002. XHTML™ 1.0 The Extensible HyperText Markup Language (Second Edition). Viitattu 20.10.2011. <http://www.w3.org/TR/xhtml1/>

W3C 2003. JPEG JFIF. Viitattu 20.10.2011. <http://www.w3.org/Graphics/JPEG/>

Wendler, R. 2006. The status of preservation metadata in the digital library community. Teoksessa Deegan, M. & Tanner, S. (ed.) Digital preservation. London: Facet Publishing. 60–77.

Yahoo 2011a. The Difference Between Yahoo! Directory and Yahoo! Search. Viitattu 12.10.2011. [http://help.yahoo.com/l/us/yahoo/search/basics/basics-06.html;\\_ylt=Aj5e6S6nxeijALGZP8ktLBEuqCN4](http://help.yahoo.com/l/us/yahoo/search/basics/basics-06.html;_ylt=Aj5e6S6nxeijALGZP8ktLBEuqCN4)

Yahoo 2011b. Pipes: Kirjastovirta. Viitattu 19.10.2011. [http://pipes.yahoo.com/pipes/pipe.info?\\_id=cqrwQKJn3RGaMPyppgt1Yg](http://pipes.yahoo.com/pipes/pipe.info?_id=cqrwQKJn3RGaMPyppgt1Yg)

Yippy 2011. About Yippy Search. Viitattu 13.10.2011. <http://search.yippy.com/about-yippy-search>

## Haastattelukysymykset maakuntakirjastojen edustajille

Verkossa olevan aineiston löytymiseen liittyvät kysymykset

Miten aineisto löytyy verkosta? Sattumalta vai onko tietty kaava/prosessi/suunnitelma sen löytymiseen?

Minkätyyppistä materiaalia verkosta yleisesti ottaen etsit?

Haetko erityyppistä aineistoa erilaisesti, ts. eroavatko niiden haut toisistaan? Esim. kaunokirjallisuus, tietokirjallisuus, kartat yms.

Käytkö tietyillä verkkosivuilla säännöllisesti / lähes säännöllisesti seuraamassa niiden tarjontaa? Millä sivustoilla?

Etsitkö tietoa tietokannoista, esim. artikkelitietokannoista?

Käytätkö Googlea/muuta hakukonetta hakemiseen? Millä hakusanoilla esimerkiksi haet? Kuinka usein? Minkälaisia tuloksia hakukoneella tulee?

Löytyykö aineistoa?

Käytätkö metahakukoneita? Esim. Dogpile, Clusty?

Käytätkö Wikipediaa tai sosiaalista mediaa hakujesi tueksi?

Onko joidenkin julkaisijoiden/yritysten/yhteisöjen kanssa sovittu uusien julkaisujen ilmestymisen ilmoittamisesta (esim. sähköpostitse)?

Saatko tietoa verkossa olevasta aineistosta informaaleista lähteistä eli esimerkiksi työtovereilta tai asiakkailta?

Lähtevätkö haut yleensä asiakkaiden tiedontarpeista vai henkilökunnan tai aineiston kokoamisen tarpeista (esim. huomataan puutteita kokoelmassa)?

Onko muita tapoja, joita en ole osannut kysyä?

Tuleeko mieleen mitään kehittämis- tai muutosideoita? Tai toisaalta joitakin erityisiä ongelmakohtia?

Verkossa olevan aineiston luettelointiin liittyvät kysymykset

Miten verkossa oleva aineisto luetteloidaan?

Kuvaillaanko sitä jotenkin?

Tallennetaanko jotakin metadataa?

Verkossa olevan aineiston säilytys lyhytaikaisesti ja pitkäaikaisesti

Onko aineiston säilytysaikaa mietitty? Säilytetäänkö kaikkea aineistoa yhtä pitkään?

Onko aineiston pitkäaikaissäilytystä mietitty? Onko sille löydetty jo jokin ratkaisu? Millainen, jos on? Miten pitkään aineistoa on tarkoitus säilyttää?

Kuka vastaa säilytyksestä? Onko säilytys järjestetty omassa organisaatiossa vai onko se ulkoistettu?

Onko aineiston säilytykselle mietitty tietty tiedostomuoto?

Mitä tapahtuu, jos aineiston verkko-osoite muuttuu tai tulee muita muutoksia?

Onko lyhytaikaisesti säilytettävälle materiaalille eri menetelmä kuin pitkäaikaiselle?

Minkälaisia tulevaisuuden suunnitelmia säilytykseen on olemassa?

## **Haastattelukysymykset Porin kaupungin ympäristöviraston johtajalle**

Millä perusteella ympäristöviraston verkkosivuille laitettavat julkaisut ja raportit valitaan? Laitetaanko sinne kaikki julkaisut ja raportit vai vain osa niistä?

Miten kauan julkaisuja ja raportteja pidetään esillä verkkosivuilla? Toisin sanoen: vaihtuvatko julkaisut silloin tällöin, vai pysyvätkö sinne kerran jo laitettut julkaisut sivuilla aina?

Miten uusien julkaisujen ja raporttien ilmestymisestä tiedotetaan?

Kenellä on julkaisujen ja raporttien tekijänoikeudet, ympäristövirastolla vai kyseisen julkaisun tehneellä henkilöllä/henkilöillä?

## **Haastattelukysymykset Kansallisen digitaalisen kirjaston edustajalle**

Miten yksittäinen maakuntakirjasto voi hoitaa pitkäaikaissäilytysasiaa, vai onko tarkoitus, että tulevaisuudessa kaikki säilytys hoidetaan keskitetysti?

Onko jo nyt olemassa joitain konkreettisia ratkaisuja asian hoitamiseksi?

Mitä on siis jo tehty yksittäisen maakuntakirjaston näkökulmasta?