

**Yliopistojen elektronisen julkaisemisen
kartoitus**

Tuija Sonkkila

Teknillisen korkeakoulun kirjasto

16.11.1999

Sisällysluettelo

ALUKSI	3
ELEKTRONISET VÄITÖSKIRJAT - PIKAKATSAUS SUOMEN TILANTEESEEN	4
<i>Helsingin yliopisto</i>	4
<i>Jyväskylän yliopisto</i>	4
<i>Kuopion yliopisto</i>	4
<i>Lappeenrannan teknillinen korkeakoulu</i>	4
<i>Oulun yliopisto</i>	4
<i>Tampereen yliopisto</i>	4
<i>Teknillinen korkeakoulu</i>	5
<i>Vaasan yliopisto</i>	5
<i>Valtion teknillinen tutkimuskeskus (VTT)</i>	5
MITÄ ELEKTRONINEN JULKAISEMINEN ON?	5
VÄITÖSKIRJA	6
KÄSIKIRJOITUS	7
KIRJOITTAJA	7
TEKSTINKÄSITTELY.....	8
MALLIPOHJAT	8
JULKAISEMINEN	9
JAKELU	9
LÖYTYMINEN.....	10
PROSESSI	10
<i>RTF ja PDF eli Tampereen malli</i>	11
<i>SGML/XML eli Oulun malli</i>	12
<i>Kuvaformaatti eli Jyväskylän malli</i>	14
<i>Monimuotoisuuden hallinta eli Helsingin malli</i>	15
<i>Vertailua</i>	16
LUKIJA	18
OSAAMINEN	19
YHTEENVETO HAASTATTELUISTA	19
LOPUKSI	21
LÄHTEET	23

Aluksi

Tämä kartoitus on tehty tilaustyönä FinELib-ohjelmalle. Sen tarkoituksena on toimia pohjana kansallisten suositusten laatimisille yliopistojen elektroniseksi julkaisemiseksi, erityisesti ajatellen väitöskirjoja ja muita opinnäytteen omaisia dokumentteja. Kartoituksessa toivottiin käsiteltävän OPM-rahoitusta saaneiden elektronisen julkaisemisen hankkeiden kokemuksia. Lisäksi siihen haluttiin mukaan eri tieteenalojen edustajien mielipiteitä väitöskirjojen elektronisesta julkaisemisesta.

Ottaen huomioon kartoituksen tekoajankohdan - vuosisadan helleheinäkuu, jolloin kontaktien saaminen oli normaalia heinäkuutakin työläämpää - haastattelukierros supistui neljän eri tieteenalan edustajaan. Onneksi haastateltavat olivat sattumalta varsin onnistuneita henkilövalintoja haastateltaviksi. Henkilökohtainen mielipiteeni on, että sekä esitellyt tieteenalat että niissä vallitsevat käytännöt tulivat hyvin esille. Niissä käsitellään mielestäni kattavasti myös erilaiset väitöskirjatyyppit ja niiden tekemisen yleisimmät välineet.

Haastateltavat edustavat vain pääkaupunkiseudun yliopistoja, lähinnä käytännön pakosta. En pidä tätä kuitenkaan ongelmallisena kartoituksen kannalta, koska tutkijoiden sidosryhmistä tärkein on yleensä kansainvälinen, saman alan tutkijoista koostuva tiedeyhteisö. Sille kirjoitetaan ja sen kautta tulevat julkaisemisen käytännöt ja lainalaisuudet.

Haastatteluissa ei ollut käytössä nauhuria, joten muistiinpanot perustuvat kotikutoiseen pikakirjoitukseen. Haastateltavat eivät erityisesti halunneet pysyä nimettöminä, mutta katsoin kuitenkin aiheelliseksi jättää nimet mainitsematta.

Kokemusten analysointi on aina jossain määrin subjektiivista. Mielipiteeni SGML/XML-pohjaisista julkaisemisen hankkeista ovat luonnollisesti värittyneitä siksi, että olen itse osallistunut niihin.

Kiitän lämpimästi kaikkia teitä, jotka korjasitte tekstiluonnoksen virheitä ja kommentoitte sekä asiasisältöä että muotoseikkoja. Erityiskiitos Oulun yliopiston kirjaston suunnittelijalle Ville Varjoselle.

Elektroniset väitöskirjat - pikakatsaus Suomen tilanteeseen

Muiden kuin tässä esiteltävien tahojen suunnitelmista ei ole tarkkoja tietoja. Katsaus perustuu alkukesän tilanteeseen 1999. Huom. esittelyt eivät ole täysin yhteismitallisia.

Helsingin yliopisto

Kaksivuotinen E-thesis -kokeiluhanke on käynnissä vuoden 1999 loppuun asti. Hankkeessa ovat mukana Helsingin yliopiston kirjasto, atk-osasto, tiedekunnat ja Yliopistopaino. Julkaisuja otetaan vastaa tavallisimpina tekstinkäsittely- ja sivuntaitotiedostoina. Jakelumuodot ovat HTML ja PDF. Alkuperäinen tiedosto toimii pitkäaikaistallenteena. Palveluun on tarkoitus liittää myös tarvepainatus, josta vastaa Yliopistopaino.

Jyväskylän yliopisto

Kirjaston Tutkielmapankki Xeroxin välineillä ja ohjelmilla, skannaus paperilta ja verkkojakelu kuvaformaattissa. Korostettu elektronisen aineiston varastoimiseen ja verkkojakeluun liittyviä tekijänoikeudellisia ja asiakirjojen julkisuuskysymyksiä. Professoritason työryhmän raportti aiheesta valmistuu vuoden 1999 aikana.

Kuopion yliopisto

Julkaisuohjeet on uudistettu 1998. Noudattavat OPM:n suositusta vuodelta 1995: painettua väitöskirjaa ei välttämättä tarvita.

Lappeenrannan teknillinen korkeakoulu

Kirjaston ja atk-keskuksen projekti testannut tiivistelmien verkkosyöttöä, -hakua ja -jakelua. Tavoitteena saada kokotekstit verkkoon, projekti alkanut toukokuussa 1999. Projektilla ei ole ulkopuolista rahoitusta.

Oulun yliopisto

Elektronisen julkaisemisen malli ja toteutus tehty yhteistyönä kirjaston, ATK-keskuksen ja YO-painon kanssa. Verkkojakelussa PDF ja HTML, pitkäaikaissäilytyksessä SGML. Rehtorin asettaman työryhmän tekemä suunnitelma valmistui keväällä 1999. Esitetään rinnakkaisen julkaisutuotannon aloittamista vuoden 2000 alusta ja yhden asiantuntijan/suunnittelijan palkkaamista. Resurssointi auki.

Tampereen yliopisto

Työryhmä esittänyt, että sähköisiin väitöskirjoihin siirrytään asteittain 1.8.1999 alkaen, 1.8.2000 alkaen kokonaan. Painettu väitöskirja on silti aina mahdollinen, mutta yliopiston tuki suunnataan vuoden 2000 alusta sähköisiin väitöskirjoihin. Tiedostomuotoina PDF verkossa ja RTF pitkäaikaistallennukseen. Tiivistelmä HTML-muodossa. Muunnokset ja hallinnointi kirjastossa, jonne ehdotetaan palkattavaksi julkaisuamanuenssi. Yliopisto on osoittanut kirjastolle hankkeeseen 90

000 mk vuonna 1999. Vuodesta 2000 alkaen hanketta rahoitetaan väitöskirjojen painatustuesta säästyvin varoin.

Teknillinen korkeakoulu

Kirjaston HUTpubl-projektissa tehty tekninen pilotti TKK:n julkaisusarjojen elektroniseksi julkaisemiseksi. Soveltuu myös opinnäytteisiin. Verkojakelussa toteutettu HTML, pitkäaikaissäilytyksessä SGML. Varsin lähellä Oulun mallia teknisesti. Väitöskirjalautakunta suosittelee 25.5.1999 antamassaan lasunnossa, että kaikki TKK:n väitöskirjat tulisi saattaa elektroniseen muotoon ja että pääkirjasto huolehtisi perustettavasta järjestelmästä.

Vaasan yliopisto

Ei vielä tarkkoja suunnitelmia väitöskirjoille. Panostettu SGML-julkaisemiseen muuten. SGML-tietokantaohjelmisto Astoria käytössä, hyödynnetty Research catalog 1998 -julkaisun teossa.

Valtion teknillinen tutkimuskeskus (VTT)

VTT Tietopalvelu on pohjoismainen elektronisen julkaisemisen osaamiskeskus, ja VTT tarjoaa teknistä asiantuntemusta ja koulutusta elektroniseen julkaisemiseen liittyvissä asioissa. Omien julkaisusarjojen tekemisessä hyödynnetään mallipohjia sekä Adoben Acrobat -tuoteperhettä PDF-muotoisten dokumenttien tekoon. VTT Tietopalvelu on ohjeistanut mallipohjien käytön. Osa PDF-dokumenteista on tulostussuojattu.

Mitä elektroninen julkaiseminen on?

Lainaan Oulun yliopiston elektronisen julkaisutoiminnan suunnitelman tekstiä (Oulun, 1999, 10):

Elektroninen julkaiseminen voidaan määritellä prosessiksi, jossa dokumentti on koko elinkaarensa ajan käytettävissä elektronisessa muodossa.

Kun dokumentin elinkaarta katsotaan lähemmin, elektroninen julkaiseminen nähdään sarjana dokumenttiin kohdistuvia käytännön valintoja ja toimenpiteitä (Wilkinson et al. mukaan):

- määrittely (mm. kuvailukielen valinta)
- tuotanto (käsikirjoitus, korjaukset, tiedostomuunnokset, metadatan lisäys)
- jakelu (jakelukomponentin koon määrittäminen, jakeluhetkellä tehtävä mahdollinen ulkoasutietojen lisäys, erot staattisen ja dynaamisen dokumentin välillä jne.)
- julkaiseminen (löytyvyyden ja paikantamisen turvaaminen, nimeäminen, versionhallinta, tallennus)
- löytyminen (kyselykielet, haku sisällön, metadatan, rakenteen tai navigoinnin avulla, hajautettu haku jne.)

- poisto (ikäntyminen, tallennuspaikan vaihto, arkistointi)

Hyvin harva dokumentti on elinkaarensa alussa muuta kuin elektroninen. Kirjoittaja voi luonnostella työtään ja sen rakennetta levittämällä käsinkirjoitettuja muistiinpanoja eteensä lattialle, mutta kun dokumentin varsinainen kirjoittaminen alkaa, siirrytään tietokoneen ääreen. Kirjoituskoneen käyttö on niin vähäistä, ettei sitä ole otettu huomioon tässä kartoituksessa. On silti huomattava, että tietokoneen tekstinkäsittelyohjelmaa käytetään yleisesti aivan kuin se olisi mekaaninen kirjoituskone. Vain hieman karrikoiden voisi väittää, että tästä nimenomaisesta syystä tämäkin kartoitus on jouduttu tekemään.

Vuokko Palonen on todennut, että perinteiset julkaisuorganisaatiot yliopistossa eivät välttämättä ota hoitaakseen elektronista julkaisemista. On vaikeata päättää, kuka tekee, miten tehdään ja millä rahoilla. Ja kuitenkin (Palonen):

Koska sähköinen julkaiseminen on usein vielä näkyvämpää kuin perinteinen julkaiseminen, eikö sen hallintaan tulisikin panostaa juuri tässä vaiheessa, kun nämä julkaisemisen ja dokumentoinnin menetelmät kehittyvät nopeasti?

Tässä onkin varmasti eräs ongelman ytimistä. Elektroninen julkaisu tuotanto näkyy ja kuuluu. Se kertoo yliopistosta enemmän, eri tavalla ja uudelle, kansainväliselle lukijakunnalle. Se kertoo julkaisu toiminnan volyymin ja tasosta, yliopiston kuvasta lukijan silmissä ja mielissä. Se paljastaa yliopiston halusta satsata asioiden tekemiseen mietitysti ja asiantuntemuksella.

Väitöskirja

Väitöskirjoja julkaistaan Suomessa vuosittain n. 1000 kappaletta¹. Kymmenessä vuodessa määrä on kaksinkertaistunut. Tilastoja hieman yksinkertaistaen voi sanoa, että pienissä yliopistoissa väitöskirjoja ilmestyy tällä hetkellä vuodessa kymmenkunta, keskisuurissa noin 50 ja suurissa yli 100. Määrä on edelleen kasvussa: Opetusministeriö on esittänyt tohtorin tutkintoihin kolmanneksen lisäystä vuosille 2001–2003. Eniten väitöskirjoja vuonna 1998 tuotettiin luonnontieteissä, tekniikassa ja lääketieteissä².

Vallitseva väitöskirjan muoto näillä eniten väitöskirjoja tuottavilla aloilla on ns. nippuväitöskirja, joka koostuu 3-5 artikkelista sekä yhteenveto-osasta. Artikkelit ovat ilmestyneet (tai ilmestyvät) jossakin kansainvälisessä referees-tyyppisessä tieteellisessä aikakauslehdessä, ja lehdelle on siirtynyt niihin tekijänoikeus. Yleistä on, että artikkelit syntyvät tohtorikoulun tai tutkimusryhmän yhteistyönä. Käytännössä on myös usein niin, että tutkimusryhmän vetäjällä on ratkaiseva rooli artikkeleiden syntymiseen johtavassa tutkimusprosessissa sekä artikkeleiden sisällöllisessä - joskus myös kirjoitusteknisessä - laaduntarkkailussa.

Monografiaväitöskirja taas on nimensä mukaisesti kirjan muotoinen, yhtenäinen, yhden kirjoittajan teos. Se on yleinen muoto varsinkin humanistis-

¹ <http://www.minedu.fi/julkaisut/universities/Suomi/1998.htm>

² http://www.minedu.fi/julkaisut/universities/Suomi/postgraduate_education.htm

yhteiskuntatieteellisillä ja niitä sivuavilla aloilla. Mitään varsinaista tilastotietoa ei asiasta kuitenkaan ole saatavilla.

Valmistunut väitöskirja käy läpi esitarkastusvaiheen, jonka aikana se ei useimmiten saa olla julkisesti esillä. Esitarkastajat haluavat käytännössä aina paperimuotoisen version.

Kun esitarkastus on käyty läpi, väitöskirja saa julkaisuluvan. Tämän jälkeen väitöskirjaa pidetään yleensä esillä jokin vähimmäismäärä päiviä ennen julkista väitöstilaisuutta. Julkaisemisen jälkeen tekstiin ei enää saa tehdä muutoksia.

Väitöstilaisuudessa väitöskirjaa käydään läpi kohta kohdalta ja tässä tilaisuudessa paperimuotoinen versio on yleensä välttämätön.

Näyttää siltä, että yliopistot haluavat edelleen jakaa väitöskirjaa tietyn määrän sekä eri puolille yliopistoa että arkistointia (ja mahdollisesti myös edelleen julkaisuvaihtoa) varten kirjastoon. Esimerkkinä Helsingin yliopiston hallintoviraston ohjeet tiedekunnille 24.5.1999³:

Väittelijä voi aiempaan tapaan luovuttaa yliopistolle väitöskirjaansa 160 kappaletta (artikkeliväitöskirjoja 30 täydellistä kappaletta) tai 30 kappaletta ja julkaisemista varten yhden elektronisessa muodossa olevan kappaleen. Jälkimmäisessä tapauksessa tiedekunta voi päättää nostaa luovutettavien kappaleiden kokonaislukumäärän 50:een.

Käsikirjoitus

Kirjoittaja

Elektronisen julkaisemisen tuotantoketjun alkupää, kirjoittaja, on monessa mielessä ketjun tärkeimpiä lenkkejä. Laadittaessa suosituksia pitääkin ottaa kantaa siihen, miten kirjoittajaan suhtaudutaan, miten hän osallistuu elektronisen julkaisemisen talkoisiin. Seuraavassa muutamia mahdollisia vaihtoehtoja, joita sovelletaan pilottiluontoisesti jo nyt:

- kirjoittajan jatkaa kuten tähänkin asti, kaikki dokumentille tehtävä lisätyö (muunnokset, metadatan lisäys jne) tehdään muualla
- kirjoittajalle voidaan antaa suosituksia ja ohjeistusta esim. tyyli- ja tiedostojen käytöstä, mutta hänen ei tarvitse noudattaa niitä. Suositusten mukaan toimittaessa työn tulokselle taataan laadukkaampi ja/tai nopeampi kohtelu ja kirjoittajalle itselleen henkilökohtaista, aineetonta etua
- kirjoittaja luovuttaa työnsä tietyssä tiedostoformaattissa ilman että sen tekotapaan on olemassa mitään erityisiä suosituksia
- kirjoittajan on luovutettava työnsä tietyssä tiedostoformaattissa niin, että se noudattaa ulkoasultaan annettuja suosituksia (mutta erityistä mallipohjaa ei tarvitse käyttää)
- kirjoittaja tekee työnsä jollakin valinnaisella mallipohjalla

³ <http://hul.helsinki.fi/elbanco/ohjeet/dissjakelu.html>

- kirjoittaja tekee työnsä jollakin valinnaisella mallipohjalla ja huolehtii itse myös joistakin tiedostomuunnoksista (esim. HTML tai PDF)

Nyrkkisääntönä voisi sanoa, että kirjoittajalle on syytä tarjota muutamia vaihtoehtoisia tapoja edetä mutta suositella käytettäväksi vain yhtä tai kahta - ja perusteltava hyvin, miksi.

Tekstinkäsittely

Kirjoittajan tekstinkäsittelyohjelman valintaan voidaan tietysti yrittää vaikuttaa. Hänelle voidaan tarjota esim. kampuslisenssien muodossa jotakin tiettyä, suositeltavaa ohjelmistoa ja siihen kytkettynä erilaista ohjeistusta ja kurssitusta. Tämä tie tuskin vie kovin pitkälle. Tekstinkäsittelyohjelmat ovat mikron ja työaseman perustyökaluja ja niihin syntyy pitkäaikainen suhde. Valinta on tullut joskus tehtyä joko vahingossa tai tarkoituksella. Monessa tapauksessa valinnan on tullut tietämättään tehneeksi jo opiskeluainetta valitessaan.

Käytännön realiteetti on, että uuden tekstinkäsittelyohjelman opetteluun on harvalla aikaa ja motivaatiota. Mieluummin opetellaan tutun ohjelmiston lisäpiirteitä, jos siihen löytyy tarpeeksi aihetta.

Yleisimmät tekstinkäsittelyohjelmistot yliopistoympäristössä, niinkuin kaikkialla muuallakin, ovat Microsoft Word, Corel WordPerfect ja Emacs. Viimeksimainittua käytetään yleisesti mm. TeX-tekstinkäsittely- ja ladontakieltä noudattavien dokumenttien tekemiseen, samoin HTML:n. TeX:in käyttäjiä ovat erityisesti matemaatikot, mutta paljon käyttäjiä on myös fysiikassa, tietotekniikassa ja niiden lähitieteissä. Useat alan tieteelliset aikakauslehdet vaativat aineiston sillä koodattuna. Erityisen suosittua on käyttää makropaketteja, joilla voidaan helpottaa huomattavasti dokumentin koodausta⁴.

Niillä aloilla, kuten bio- ja lääketieteissä, joissa matemaattisia kaavoja sisältyy dokumentteihin vähemmän, MS Word ja sen kaavaeditori on yleinen väline.

Mallipohjat

Tekstinkäsittelyohjelman tyyliminaisuuksien ja mallipohjien käyttö helpottaa oleellisesti dokumentin jatkojalostusta, valittiinpa tallennus- ja jakelumuodoksi mikä tahansa ja olipa julkaisuprosessi mikä hyvänsä. Tähän on päädytty kaikissa näkemissäni elektronisen julkaisemisen hankkeissa, ei vain kotimaisissa. Mallipohjilla saadaan dokumenttien rakenteeseen ennustettavuutta, jota voidaan käyttää hyväksi julkaisuprosessin myöhemmissä vaiheissa. Niillä voidaan myös pyrkiä yhtenäistämään jaettavien dokumenttien ulkoasua.

Mallipohjatiedostoja on helppo laittaa jakeluun web-sivuille ja niiden käytön ohjeistuksen laatu ja laajuus on kiinni vain ajasta ja työvoimasta. Mutta mallipohjienkaan käyttö ei ole täysin ongelmaton kolmesta syystä:

- ne ovat ohjelmistoversioihin sidottuja eli niitä pitää olla kohtuullisen iso valikoima
- niissä ei voida ottaa huomioon kaikkia mahdollisia tilanteita, joihin kirjoittaja törmää, jolloin kirjoittajalle tulee kiusaus muuttaa niitä

⁴ TeX:stä lisää esim. <http://hul.helsinki.fi/elbanco/tex/tex.htm>

- kirjoittaja voi muuttaa niitä

IEEE, joka on eräs suurimmista tekniikan alan kustantajista, on arvioinut, että 90% kirjoittajista muuttaa niitä Wordin mallipohjia, joita se tarjoaa kirjoittajilleen (Owens).

Dokumenttien jatkojalostusta on siis lähes mahdotonta automatisoida täydellisesti, sitä voidaan vain yrittää lähestyä.

Realistisin vaihtoehto resurssien kannalta voisi olla mallipohjan tarjoaminen kahdelle yleisimmälle tekstinkäsittelyohjelmalle eli MS Wordille ja (La)TeX:lle. Niiden teko, ohjeistaminen ja käyttökoulutus on sekin jo iso urakka. Kannattaa myös seurata sitä, missä tahdissa WordPerfect valtaa takaisin menetettyä valta-asemaansa. Monet odottavat sitä.

Julkaiseminen

Jos kirjoittaja voi toimittaa tekstinsä eteenpäin tuotantoketjussa kovin monessa tiedostomuodossa, dokumentin jatkokäsittely tallennukseen ja jakeluun sopivaksi voi olla aikamoinen työ⁵. Kun jokainen dokumentti on käsiteltävä omana ainutlaatuisena tapauksenaan, lähestytään arkistointia harjoittavan yksikön toimintaa.

Dokumenttikohtainen kustannus nousee korkeaksi.

Mielessä saattaa käväistä mahdollisuus erityisen käsittely- tai arkistointimaksun periminen kirjoittajalta tai tämän taustayhteisöltä. Tällainen maksu on käytössä mm. Virginia Tech:ssä⁶. Tätä kirjoitettaessa ei ole tiedossa, onko sillä jokin erityinen funktio VT:ssä ja onko se USA:ssa yleinen tapa. Maksua voi olla kuitenkin erittäin vaikea perustella Suomessa. Joka tapauksessa se voisi toimia korkeintaan käytäntöjä suuntaavana keppinä - ei tosin kovin korrektina! - tuskin missään oloissa kustannuksia todella kattavana. On vaikea kuvitella, että kirjastot hypähtäisivät nykytilanteesta yhtäkkiä kustannusvastaaviksi painotaloiksi, kun kustantamisen ammattilaiset, yliopistopainot, ovat nekin tiukoilla. Näin on myös mm. USA:ssa (Fox).

Jakelu

PDF on de facto -standardi verkkojakelussa HTML:n ohella. Perus-PDF -formaattia on helppo tuottaa sekä ilmaisilla välineillä⁷ että Adoben omilla tuotteilla. Muunnos voidaan myös automatisoida erittäin pitkälle.

Oulun projektissa tehtiin kuitenkin se havainto, että lisäpiirteiden tekeminen PDF:ään saatta olla varsin hidasta (Varjonen). Luonnollisesti tämä riippuu myös dokumentista.

Tiedossa ei ole, kuinka laajasti opiskelijoiden käytössä on Adoben Acrobat -ohjelmistopakettia, jolla voidaan lisätä mm. sisällysluettelo ja kirjanmerkkejä, ja miten ohjelmistoa osataan käyttää. Paketti on korkeakoululisenssin piirissä Suomessa, joten jonkin verran levinneisyyttä sillä varmasti on.

⁵ ks. esim. Timo Hellgrenin kokemuksia raportissa <http://hul.helsinki.fi/elbanco/aineisto.html>

⁶ <http://etd.vt.edu/submit/steps.htm>

⁷ Eräs vaihtoehto on tallettaa teksti PostScript-muodossa ja käyttää muunnokseen Ghostview-ohjelman pdfwrite-ajuria

Esimerkki USA:sta: Virginia Tech:ssä Acrobat on asennettuna kaikkiin mikroluokkiin ja opiskelijoista lähes kaikki luovuttavat opinnäytteensä PDF-muodossa. Ohjeet on laadittu sekä itse PDF-muunnoksen että sisällysluettelon, kirjanmerkkien jne. tekemisestä. Pienen summittaisen otannan (=10) tulos oli, että näiden lisäpiirteiden tekemiseen tarvittavaa Exchange-ohjelmaa ei kuitenkaan ole käytetty.

PDF on joka tapauksessa kiistaton de facto -standardi silloin, kun lukijoille halutaan tarjota verkossa tulostettava formaatti. PDF:stä on myös tullut vaivattomin tapa painotalojen kanssa asiointiin. Monet painotalot eivät esimerkiksi suostu enää hevillä ottamaan vastaan MS Word -dokumentteja. PDF-dokumentti voidaan lisäksi varustaa digitaalisella allekirjoituksella niinkuin mikä tahansa tiedosto ja PDF on myös indeksoitavissa hakuun⁸.

Löytyminen

Elektroninen dokumentti on pystyttävä löytämään ja paikantamaan Internet-verkosta. Kansalliskirjasto on antanut suositukset Dublin Core⁹ -metadatatamäärityksen ja kotimaisten, auktorisoitujen URN (Uniform Resource Name)¹⁰ -tunnisteiden käytöstä. Suosituksia on syytä noudattaa, vaikka ne ovatkin vasta varautumista tulevaisuuteen: hakukoneet eivät toistaiseksi osaa indeksoida HTML-dokumentin META-elementin Dublin Core -muotoista dataa - harvoja poikkeuksia on Nordic Web Index¹¹ - ja URN-resoluutiopalvelu on sekin vielä työn alla.

Nordic Metadata -projektissa¹² on tehty julkaisijalle työtä helpottavia apuneuvoja. Dublin Core -tallennusalueella¹³ voi määritellä metadatan ja URN-generaattorilla¹⁴ tunnisteiden. Tavoitteena on lisäksi kehittää Dublin Core -syntaksia siten, että se tukee myös XML-sovellusta RDF (Resource Description Framework)¹⁵.

Hyvä johdatus verkkoaineiston kuvailuun ja identifiointiin sisältyy Helsingin yliopiston kirjaston EVA-projektin loppuraporttiin¹⁶.

Prosessi

Harvalla kirjoittajalla on monipuolinen elektronisen kirjoittamisen ohjelmistorepertuaarin tuntemus ja käytännön sovellustaito. Tietämyksen kasvattamiseen ei ole yleensä ollut tarvetta ja/tai aikaa. Elektronisessa julkaisemisessa tarvitaankin ammattitaitoista toimitussihteeriä kuten kaikessa julkaisemisessa.

Tästä huolimatta olisi suotavaa, että opiskelija oppisi yliopistossa elektronisen julkaisemisen perusvalmiudet.

⁸ Teemu Rautanen on kirjoittanut hyvän katsauksen paitsi PDF:ään myös muihin elektronisen julkaisemisen tiedostomuotoihin: <http://renki.lib.helsinki.fi/elektra/raportti/4.html>

⁹ Dublin Core Metadata Initiative. <http://purl.org/DC/>

¹⁰ URN syntax (RFC 2141). <ftp://ftp.isi.edu/in-notes/rfc2141.txt>

¹¹ <http://nwi.funet.fi/>

¹² <http://linna.helsinki.fi/meta/>

¹³ <http://linna.helsinki.fi/cgi-bin/dc.pl>

¹⁴ <http://hul.helsinki.fi/cgi-bin/urn.pl>

¹⁵ W3C. Resource Description Framework (RDF) Model and Syntax Specification.

<http://www.w3.org/TR/REC-rdf-syntax/>

¹⁶ <http://renki.lib.helsinki.fi/eva/loppurap/html/6.html>

Elektronisen julkaisemisen prosessia suunniteltaessa on tärkeää yrittää löytää tasapaino realismin ja idealismin välillä. Oleellista on mm. käyttää sellaisia dokumentin kuvailukieliä ja tallennusmuotoja, joista voi kohtuullisella vaivalla saada seuraavankin sukupolven de facto -jakelustandardi(e)n mukaisia dokumentteja ja paperimuotoisia versioita. Sukupolvet kun vaihtuvat web-maailmassa nopeasti.

Se, kuinka paljon julkaisujen tuotantoprosessiin satsataan ja mitä prosessissa tehdään, riippuu ainakin seuraavista tekijöistä:

- kuinka tärkeäksi elektroniset dokumentit ja niiden hallinta koetaan
- mitä elektronisilta dokumenteilta odotetaan, mitä niillä halutaan tehdä
- mitä ja kuinka suuria paineita prosessin toteuttamiseen kohdistuu
- kuinka paljon prosessin uskotaan maksavan
- missä työ tehdään
- mitä osataan ja mitä oletetaan osattavan jatkossa
- miten muualla on tehty

Minkälaisia malleja elektronisten dokumenttien käsittelyyn sitten on olemassa? Esittelen seuraavassa kolme mallia, joista kaikista on jo kokemuksia sekä Suomessa että ulkomailla. Keskityn tarkastelemaan malleja lähinnä vain siltä kannalta, mitä valintoja niissä on tehty dokumentin kuvailukielen suhteen. En käsittele työnkulkuun liittyviä asioita muuten kuin sattumanvaraisesti. Elektronisten dokumenttien sopimusoikeudelliset seikat ovat varsin tyhjentävästi tulleet esille Elektra-projektissa, joten en puutu niihin. Lopuksi vertailen malleja keskenään sen perusteella, miten ilmaisuvoimaisia ja joustavia niissä valitut dokumentin kuvailukielet ovat.

RTF ja PDF eli Tampereen malli

Tampereen malli muistuttaa useimpia Ruotsissa käynnistyneitä hankkeita (ks. alla), vastaavia esimerkkejä löytyy runsaasti myös muualta. Perusajatus on se, että kirjasto huolehtii dokumenttien muuntamisesta PDF-verkkojakelumuotoon.

Kirjoittaja käyttää jotakin yleisesti käytössä olevaa tekstinkäsittelyohjelmaa kuten MS Word ja Corel WordPerfect ja toimittaa dokumentin edelleen kirjastolle sähköisessä muodossa. Erityisiä mallipohjia voidaan myös suositella käytettäväksi dokumentteja kirjoitettaessa.

Tampereen mallissa kirjoittaja lähettää lisäksi kirjastolle väitöskirjan tiivistelmän ja viitetiedot sähköisessä muodossa, jotka kirjasto muuntaa HTML-muotoon ja sijoittaa ne tietokantaan.

Ruotsalaisista verrokeista Tampereen malli poikkeaa siinä, että elektroniseksi arkistointiformaatiksi on valittu RTF (Rich Text Format).

Plussat: kevyt ja nopeasti käynnistettävä malli. MS Word on tällä hetkellä yleisin tekstinkäsittelyohjelmisto, joten kirjasto suoriutunee lyhyen harjoittelujakson jälkeen varsin rutiinilla PDF-muunnosten tekemisestä valtaosaan dokumentteja. Mallipohjilla ja ohjeistuksella työtä voidaan edelleen helpottaa.

RTF toimii hyvin vaihtoformaattina tekstinkäsittelyohjelmien välillä ja sillä voi välttää mm. Word-makroviruksen, joka voisi muuten olla pienenä riskinä.

RTF:stä on tunnistettavissa tekstin eri osia ja näin ollen sen muuntamiseen voidaan tehdä ohjelmia, ainakin periaatteessa. Ilmaisia HTML-muunnosohjelmia on olemassa.

Melko varmasti myös XHTML-muuntimia ilmaantuu aikanaan, todennäköisesti käypiä muunnosohjelmia tulee myös joihinkin XML-sovelluksiin. Ilmaiseilla TeX2RTF-ohjelmalla saataneen myös ainakin osa (La)TeX-dokumenteista muunnettua ohjelmallisesti RTF-muotoon.

Miinuksia: RTF on ongelmallinen formaatti neljästä syystä:

1. se on Microsoft-yrityksen omistama ja näin ollen kehitystyö ja dokumentointi on arvoitus
2. sitä pidetään vaikeasti ymmärrettävänä, koska koodit eivät ole selkokieleisiä.
3. se sisältää sekaisin sekä tyyli- ja ulkoasutietoja että dataa¹⁷.
4. siitä on liikkeellä monia versioita, joista mikään ei ole selkeästi dokumentoitu (Gulbrandsen 1999).

Muunnokset (La)TeX:sta RTF:ään voivat osoittautua käytännössä hankaliksi, koska erilaisia (La)TeX-makropaketteja on paljon ja TeX2RTF ei voi millään suoriutua niistä kaikista. Tähän on periaatteellinen syy: (La)TeX ei ole varsinainen merkkäuskieli vaan ohjelmointikieli (makrokieli). Sillä voidaan tuottaa automaattisesti tekstiosioita, joita ei alkuperäisessä käsikirjoituksessa ole vielä lainkaan semmoisenaan nähtävissä. Ainoa tapa päästä käsiksi tuohon automaattisesti lisättävään tekstiin on siis suorittaa samainen ohjelma.

Voidaan myös kysyä, miksi muunnosta kannattaa ylipäänsä tehdä, koska LaTeX on pätevä arkistointiformaattina (ks. esim. vertailutaulukkoja alla).

Viitetieto dokumentista on syötettävä erikseen hakujärjestelmään/iin.

Muuta: RTF:n sisältämä data on sitä käyttökelpoisempaa, mitä selkeämmin kirjoittajalla on ollut hallussa kappaletyyliä käyttä Wordissa ja mitä tarkemmin mallipohjan käytöstä on sovittu. Sama pätee (La)TeX-käyttöön. Jos ongelmia tulee muunnoksissa, on turvaututtava paperimuotoisen arkistoversion käyttöön.

SGML/XML eli Oulun malli

Väitöskirjoille ja muille opinnäytteille ei ole olemassa erillistä nk. teollisuusstandardia dokumenttityypin määrittämisestä (Document Type Definition, DTD). Virginia Tech'in DTD on tunnetuin tähän tarkoitukseen erityisesti tehty, mutta ainoastaan DocBook DTD:llä on vakiintunut asema kansainvälisesti. Seuraavassa lista DTD:eistä, joita ainakin tiedetään käytetyksi opinnäytteisiin:

- **Oulun yliopisto.** DocBook DTD
- **Teknillinen korkeakoulu.** HUTpubl DTD
- **Karlskrona/Ronneby.** Opinnäytteiden XML DTD (ei vielä julkistettu)
- **Virginia Tech.** ETD-ML (XML-muunnosta tästä ei ole vielä julkistettu)
- **Humboldt Universität zu Berlin.** DiML (pohjautuu ETD-ML:ään)

Plussat: dokumenttietokanta on arkistokelpoinen ja monikäyttöinen. Tietokantana voidaan käyttää joko tiedostorakennetta sellaisenaan, jotakin relaatiotietokantaa tai aitoa SGML-tietokantaohjelmistoa. Laaja käyttäjäkunta dokumenttityypin määrittämiselle takaa kehitystyön ja käytännössä ilmaisten muunnosohjelmien, tyylitiedostojen jne. saatavuuden.

¹⁷ Tässä se muistuttaa TeX:iä, mutta TeX:in etuna on sen avoimuus, muunneltavuus ja kohtuullinen ymmärrettävyys.

SGML-dokumentti voidaan automaattisesti muuntaa erilaisiin jakeluformaatteihin.

Viitetieto voidaan siirtää suoraan rakenteisesta dokumentista hakujärjestelmiin tai dokumenttiarkistosta dokumenttiin.

SGML on kansainvälinen standardi¹⁸.

Miinukset: vie aikaa toteuttaa kokonaisuudessaan, joten tulee kalliiksi. Vaatii tietotaitoa, laaduntarkkailua ja suunnitelmallista, pitkäjänteistä henkilöstö- ja julkaisupolitiikkaa. Malli ei myöskään onnistu ilman eri alojen asiantuntijoiden yhteistyötä.¹⁹

Mitä enemmän kirjoittaja haluaa/osaa rakenteistaa dokumenttiaan itse, sitä suurempiin säästöihin päästään. Mutta kirjoittajaa voi olla vaikeaa saada suostumaan tähän. Niinpä lopputulos on se, että rakenteistaminen on tehtävä muualla ja tätä varten rakennettava täysimittainen dokumenttien jatkojalostus- ja hallintaympäristö.

Tekniikan hallitsevia henkilöitä on Suomessa tällä hetkellä erittäin vähän.

SGML:n muista, teknisluontoisista ongelmista, ks. esim. Varjonen.

Muuta: Suomessa tuskin kannattaa ylläpitää kahdenlaista DTD:tä opinnäytteille. Oman DTD:n ylläpitotyö käy raskaaksi, koska kaikki muunnosohjelmat, tyylisivut jne. on tehtävä myös itse. Vaihtoehtoina on siirtyminen kokonaan esim. DocBook:in käyttäjäksi tai HUTpubl DTD:n muokkaaminen DocBook DTD:n alijoukoksi. Jälkimmäisestä on käyty alustavia kahdenkeskisiä keskusteluja RAJU-projektin piirissä.

Toinen vaihtoehto olisi liittyä virallisesti NDLTD:hen (ks. alla) ja testata Saksan Humboldtissa tehdyn DiML:n sopivuutta Suomeen. Etuna olisi kansainvälinen näkyvyys ja NDLTD:n hajautettu hakupalvelu. Mutta varjopuoliakin on. DiML:n ja siihen liittyvien apuvälineiden jatkosta ei ole toistaiseksi takeita. Voi myös olla niin, että NDLTD ei saavuta lukijoitaan, ettei se saa toivottua statusta.

FrameMake+SGML on pätevä taitto-ohjelmistona ja tekstin rakenteistajana, mutta tavalliselle kirjoittajalle se on liian raskas väline. Kaupallisia, aitoja SGML-editoreita on olemassa, mutta niiden lisenssipolitiikka on kehittymätöntä eli ne tulevat kalliiksi. Ja kuten edellä on jo todettu, harva haluaa vaihtaa tekstinkäsittelyohjelmaansa.

Word-mallipohjan tyylivalikoima on todennäköisesti oltava pienempi kuin mitä HUTpubl-projektissa tehdyssä mallipohjassa nyt on. Kun määrää pienennetään, kirjoittajan työ helpottuu, samoin ohjeistaminen. Toisaalta tällöin dokumentin jatkojalostus DTD:n mukaiseen rakenteiseen muotoon vaikeutuu. Mitä lähemmäs kirjoittajaa SGML tuodaan, sen edullisempi tuotantoprosessille. Jokin kompromissi on siis löydettävä.

HTML:n korvaajaksi tituleerattuun W3C-suositukseen XML (eXtended Markup Language) kohdistuu paljon odotuksia, osin sen saaman laajan julkisuuden ja hyvän markkinoinnin ansiosta. XML:n käyttöönotto ei kuitenkaan ole sen yksinkertaisempaa ja helpompaa kuin SGML-standardinkaan, josta se on johdettu. Molemmat ovat metakieliä, eivät sidottuja johonkin elementtivalikomaan kuten HTML. Myös XML vaatii käytettävän elementtivalikoiman ja sen käsittelysääntöjen määrittelyä,

¹⁸ ISO 8879 Information Processing - Text and Office Systems - Standard Generalized Markup Language. 1986

¹⁹ Eri asia sitten on, onko nämä laskettava miinukseksi vai plussaksi!

julkaisuprosessin läpikäyntiä, uusien tekniikoiden oppimista, kehittämistä, soveltamista ja testaamista, kirjoittajien koulutusta ja ohjausta sekä työvoimaa. Ilmaisia lounaita ei todellakaan ole!²⁰

XML:n lupaama täysin vapaa, dokumenttityypitön rakenteinen kirjoittaminen tuskin käytännössä toteutuu yliopistojulkaisemisen kohdalla. Jos dokumenttien rakenteisesta julkaisemisesta halutaan irti kaikki hyöty, silloin on tuskin mitään syytä jättää rakenteistamatta dokumenttia tietyn dokumenttityypin määrityksen pohjalta.

XML:llä on kuitenkin paljon potentiaalia verkkojakelusformaatin, koska se on nimenomaan web-ympäristöön suunniteltu. XML-dokumenttien katsominen web-selaimella eli ulkoasutietojen liittäminen XML-dokumentteihin on tosin vielä kiistelty aihe selainvalmistajien keskuudessa, eivätkä kiistat valitettavasti yleensä tiedä hyvää käyttäjälle. Uudet, monipuolisemmat linkkimääritykset ovat nekin vielä kehittelyn alla.

Kuvaformaatti eli Jyväskylän malli

Jyväskylän yliopiston kirjaston julkaisuyksiköllä on käytössään XEROX COMPANY:n digitaalinen julkaisu- ja tarvepainatusjärjestelmä. Järjestelmä koostuu seuraavista osista: XDOD, DOCUWEB ja DOCUTECH 135²¹. Opinnäytteet ovat omana tietokantanaan nimeltä Tutkielmapankki. DOCUWEB- ja DigiPath-ohjelmistoilla hoidetaan digitoitujen opinnäytteiden haku ja jakelu. Dokumentit ovat web-selainten suoraan näyttäviä, JPEG-pakkausalgoritmia käyttäviä kuvia.

Plussat: Järjestelmä on nopea. Tutkielmapankissa on jo yli tuhat opinnäytettä. Se voidaan ilmeisesti pitää lähes ajantasaisena.

Palvelu näkyy lukijalle yhtenäisenä, ja dokumentit ovat eheitä kokonaisuuksia. Koska dokumentit ovat kuvina, tekstin kopioiminen ja sivujen tulostaminen on käytännössä lähes mahdotonta, joten luvaton käyttö on estetty tehokkaasti²².

XEROX on ilmeisen vakavarainen yritys ja sillä on paljon näyttöä osaamisesta digitaalisen painamisen alueella. Se pyrkii myös vakavissaan yliopistomarkkinoille, joten asiakassuhde on varmasti hyvä.

Miinukset: JPEG on ongelmallinen tekstipainotteisten kuvien pakkausalgoritmina, koska se on ns. hukkaava. Matalaresoluutioiset JPEG-kuvat ovat heikosti luettavissa ruudulta, joten kokotekstin näkymisestä ei ole lukijalle paljon iloa. Voisi ajatella, että pelkkä yhteenveto ja/tai sisällysluettelo olisi riittävä.

Internetin hakukoneet eivät voi indeksoida JPEG-kuvatiedostoja. Siksi palvelua voi olla vaikea ellei mahdoton niveltää osaksi laajempaa, kansallista kokonaisuutta. Tämänkin kartoituksen suppea haastattelukierros antoi tukea sille olettamukselle, että lukijat toivovat "yhden luukun" -periaatetta. Heillä ei ole halua lähteä selailemaan yksittäisten yliopistojen web-sivuja hakiessaan kirjallisuutta.

Sitoutuminen yhteen yritykseen järjestelmätoimittajana voi olla ongelmallista (ei suinkaan välttämättä).

²⁰ Ellei päätetä, että tehtiinpä miten tahansa, sovelletaan dokumentteihin EU:n tuomioistuimen päätöstä koskien tapaus Heidi Hautalaa: "Principen om tillgång till dokumentet gäller bara dokumenten som sådana, inte den information de innehåller" (HBL 20.7.1999)!

²¹ Graafinen kuvaus ks. Dokumentteja...

²² Mutta kolikolla on toinenkin puolensa: luvallinenkin käyttö on hankalaa.

Muuta: Jyväskylän yliopiston kirjasto on profiloitunut elektronisen kirjatalouden suuntaan. Voinee otaksua, että pakettiratkaisuna hankittavan järjestelmän avulla kirjaston on mahdollista ja varmasti järkeväkin suuntautua näin. Kun tekniikka on ikään kuin ulkoistettu, aikaa ja energiaa jää muuhun. Investointi on myös tehtävä kannattavaksi.

Jyväskylä tutki osana EVA-projektia opinnäytteiden varastointia ja verkkojakelua. Kokemuksista ei kuitenkaan ole paljon dokumentoitua tietoa. EVA-loppuraportissa²³ mainitaan lyhyesti ainoastaan projektin määrälliset tulokset ja käytetty perustekniikka.

Digitoitu aineisto on luonnollisesti muunnettavissa, jos niin halutaan, johonkin tekstin rakennetta ilmaisevaan muotoon, jollei mitään originaalidokumentin elektronista versiota ole säästetty. Tarvitaan tosin ohjelmisto, joka pystyy visuaalisten vihjeiden perusteella päättämään, minkälaisesta rakenteen osasta on kyse. Tällä tavalla saadaan kuitenkin vain HTML-tasoinen, ulkoasua ilmentävä rakenne dokumentille. Semanttinen informaatio on menetetty.

Tässäkin mallissa yhtenäinen opinnäytteiden mallipohjavalikoima yleisille tekstinkäsittelyohjelmille ei olisi suinkaan pahitteeksi. Sillä saavutettaisiin mm. tarvittavaa ennustettavutta dokumenttien ulkoasuun siltä varalta, että edellä esitetyn kaltaiseen muunnokseen olisi joskus tarvetta ryhtyä.

Monimuotoisuuden hallinta eli Helsingin malli

Helsingin yliopiston E-thesis (Helsingin yliopiston verkkojulkaisu)²⁴ -palvelussa kirjoittaja voi lähettää valmiin käsikirjoituksen eteenpäin muunnettavaksi HTML- ja PDF-verkkojakelumuotoon, mikäli käsikirjoitus on tehty joillakin seuraavista ohjelmistoista:

- Word (kaikki PC- ja Mac-versiot)
- WordPerfect (kaikki versiot)
- PageMaker (kaikki versiot)
- (La)TeX (kaikki versiot)

E-thesis ottaa lisäksi vastaan aineistoa RTF- ja PostScript-muodossa. Valmiita mallipohjia on tarjolla seuraviin ohjelmistoihin:

- Word 97 (Windows 95/98/NT)
- Word 98 (Macintosh)
- Word 6.0 ja 95 (Windows)
- Word 6.0 (Macintosh)
- WordPerfect 6.0 (Windows)

Julkaisujen tunnisteena toimii väitöskirjoilla perinteinen ISBN-numero ja muilla opinnäytteillä URN (Universal Resource Name). Metadata ilmaistaan DC (Dublin Core) -määrittelyksellä.²⁵

Plussat: Kirjoittajalla on mahdollisuus valita ruhtinaallisen monesta erilaisesta tiedostomuodosta, mikä tekee E-thesis'stä harvinaisen käyttäjäystävällisen. Valinnanvapaus ja matala aloituskynnys on varmasti houkutteleva monelle

²³ <http://hul.helsinki.fi/eva/loppurap/html/1.html#1.3>

²⁴ <http://hul.helsinki.fi/elbanco/index.html>

²⁵ lähemmin ks. Löytyminen

opiskelijalle. E-thesis on malliesimerkki siitä, miten uusi palvelu on syytä markkinoida. Hankkeessa on luultavasti syntynyt myös paljon kokemusperäistä tietoutta siitä, minkälaisia työvaiheita tekstimuunnoksissa tulee vastaan, mitä tietotaitoa ja välineistöä tarvitaan ja paljonko itse työhön kuluu aikaa.

Miinuksset: Hankkeen ensimmäisen vaiheen lopussa on varmasti syytä harkita sitä, mihin resurssit tulevat jatkossa riittämään tai miten palvelu pystytään rahoittamaan. Kaikkien tekstinkäsittelyohjelmien kaikkien versioiden tukeminen on työvoimavaltaista ja siksi kallista. Tähän liittyy myös kysymys pitkäaikaiskäytöstä. Mitä useampi erilainen originaalitekstin tallennusmuoto, sen suurempi kustannus siinä vaiheessa, kun eteen tulee joskus tarve päivittää dokumentit.

Muuta: Helsingin yliopiston hankkeilla ja yliopiston tekemillä ratkaisulla on jo yksistään yliopiston koon takia Suomessa keskeinen asema. Elektronisen julkaisemisen ja jakelun kokeiluissa on tärkeää paitsi teknisten ratkaisujen testaaminen myös yritykset saada tuntumaa mallien skaalautuvuuteen. Julkaisuvolyyminsa ja monialaisuutensa vuoksi Helsingin yliopisto olisi mitä parhain testiympäristö myös muille tässä kartoituksessa esitellyille malleille, ennen muuta dokumentin rakenteistamiseen perustuville. E-thesiksen ja kuvitteellisen "XML-thesiksen" myötä olisi ainutlaatuinen mahdollisuus saada mm. vertailukelpoista aineistoa elektronisen julkaisemisen kustannustekijöistä.

Vertailua

Wilkinson et al. ovat teoksessaan *Document computing* käsitelleet yksityiskohtaisesti dokumenttien kuvailukielten ominaisuuksia ja niiden soveltuvuutta eri tehtäviin. Kuvailukielen valintaan vaikuttavat heidän mukaansa mm. seuraavat seikat (21-24):

- tuki eri kielille
- tuki eri medioille
- tuki tietoturvaratkaisuille
- kielen määrittelyn tarkkuus
- näkyvyys: onko sisältö käytettävissä muille kuin sille sovellukselle, jolla dokumentti on tehty?
- rakenne: kuinka tarkkaan ja syvälle dokumentin looginen rakenne voidaan erottaa ulkoasusta?
- tuki sisäisille ja ulkoisille linkeille
- tuki metadatalle
- tuki eri hakujärjestelmille
- dokumenttien luonnin helppous
- tuki muunnoksille eri kuvailukielten välillä (import ja export)
- dokumentin päivittämisen helppous
- laadukkaan ulkoasun tekemisen helppous
- erilaisten ulkoasujen tekemisen helppous
- säilyvyys
- dokumentin viemä fyysinen tila
- käytännön realiteetit: miten saatavilla olevat ohjelmistot toimivat, mitä ne maksavat ja minkälaista tukea niihin saa?

Sitä, miten hyvin - asteikolla heikko, ok, hyvä, mainio - sekä edellä esitetyt että eräät muut tunnetut kuvailukielet toteuttavat osaa näistä piirteistä, voidaan esittää seuraavien kolmen taulukon muodossa²⁶.

JPEG on jätetty taulukoiden ulkopuolelle, koska se ei ole kuvailukieli vaan (rasteri)kuvaformaatti. Jos sille kuitenkin haluaisi antaa kuvitteellisia arvosanoja, ne olisivat suunnilleet samat kuin PostScript-kielen²⁷, vaikka JPEG:llä ja PostScriptillä ei teknisesti olekaan mitään tekemistä toistensa kanssa. PostScript-dokumentti sisältää yleensä tekstin semmoisenaan; se voidaan vaikka irroittaa omaksi ASCII-tiedostokseen muusta koodista. Rasterikuvasta taas ei löydy hakemallakaan kuin pelkkiä pikseleitä.

	Kielituki	Mediatuki	Näkyvyys	Rakenne	Linkit	Metadata
ASCII	Heikko	Heikko	Mainio	Heikko	Heikko	Heikko
RTF	Ok	Hyvä	Hyvä	Ok	Ok	Ok
LaTeX	Vaihtelee	Hyvä	Hyvä	Ok	Ok	Hyvä
SGML	Hyvä	Ok+	Hyvä	Mainio	Ok+	Hyvä
HTML	Hyvä	Hyvä+	Ok	Ok	Hyvä	Ok+
XML	Hyvä	Ok+	Hyvä	Hyvä	Mainio	Hyvä
PostScript	Vaihtelee	Hyvä	Heikko	Heikko	Heikko	Heikko
PDF	Vaihtelee	Hyvä	Heikko	Heikko	Ok	Ok

Taulukko 1. Kuvailukielten vertailu - ominaisuudet (mukaillen Wilkison et al, 38)

	Luonti	Import	Export	Päivitys
ASCII	Mainio	Hyvä	Heikko	Mainio
RTF	Ok	Ok	Ok	Hyvä
LaTeX	Hyvä	Heikko	Ok	Hyvä
SGML	Ok	Heikko	Hyvä	Hyvä
HTML	Ok	Ok	Heikko	Hyvä
XML	Ok	Heikko	Hyvä	Hyvä
PostScript	Heikko	Hyvä	Heikko	Heikko
PDF	Heikko	Hyvä	Heikko	Heikko

Taulukko 2. Kuvailukielten vertailu - luonti ja muunnokset (mukaillen Wilkison et al, 41)

²⁶ Ks. myös http://renki.helsinki.fi/elektra/rapo2.html#yhteenveto_tiedostomuodot

²⁷ Luonti-piirteen kohdalla JPEG saisi ehkä kuitenkin arvosanan "Mainio" ja Ulkoasun laatu -piirteen kohdalla "Vaihtelee".

	Ulkoasun laatu	Ulkoasun jousto	Ikä	Tila
ASCII	Heikko	Heikko	Mainio	Mainio
RTF	Ok	Ok	Ok	Heikko
LaTeX	Mainio	Hyvä	Hyvä	Hyvä
SGML	Hyvä	Mainio	Hyvä	Ok
HTML	Ok/Mainio	Hyvä	Heikko	Ok
XML	Hyvä	Mainio	Hyvä	Ok
PostScript	Mainio	Heikko	Hyvä	Heikko
PDF	Mainio	Heikko	Ok	Heikko

Taulukko 3. Kuvailukielten vertailu - ulkoasu, ikä ja tila (mukaillen Wilkison et al, 41)

Lukija

OPM-hankkeissa ei tietävästi ole tehty elektronisten dokumenttien lukijatutkimusta. Koska erillinen käyttöliittymähanke on käynnistynyt FinELib:in toimesta, en mene asiaan sen pitemmästi. Luonnostelen vain kuvitteellisen lukijan prototyypin sellaisena miksi se muodostui tässä työssä tehtyjen haastattelujen pohjalta.

Lukija käyttää web-selainta. Hän lähtee etsimään dokumenttia hakupalvelusta, joka on hänelle entuudestaan tuttu tai jonka hän tietää olevan kansainvälisesti kattava. Hän haluaa hakea termeillä tai niiden yhdistelmillä. Joskus hän haluaa vain selailua kokoelmasta. Lukija haluaa tutustua pidempään tai moniosaiseen julkaisuun ensin pintapuolisesti selailemalla ja/tai lukemalla sen tiivistelmän ja sisällysluettelon. Sisällysluettelosta on suorat linkit tekstiin, samoin lähdeluettelosta lähteisiin. Häntä ilahduttaisi se, jos hakusanat tai -fraasi näkyisivät kontekstissaan. Olisi myös hyvä, jos tekstissä esiintyvien lähdelinkkien kohteet (osoitteet) tulisivat näkyviin silloin kun kursori on linkin päällä, ilman klikkausta. Kun lukija vakuuttuu tekstin mielenkiintoisuudesta, hän haluaa tulostaa sen kokonaan tai osittain paperille.

Tutkimuksissa on todettu (ks. esim. Bishop), että varsinkin nuoremmat, tottuneet webin käyttäjät ovat kehittäneet tavan "skannata" nopeasti läpi ruudulla näkyvää HTML-tekstiä. Sen arvellaan selittävän myös osan siitä, miksi dokumenttien metadatan käytetään varsin vähän hyväksi, varsinkin erilaisten testiympäristöjen hakukokeiluissa. Mutta tälle voi olla luontevampikin selitys: lukijoille ei ole voinut muodostua metadatan käyttötapaa, koska metadatan ei WWW-dokumenteilla yleensä ole todellisuudessa, koejärjestelyjen ulkopuolella.

Kun yliopistoissa käydään keskustelua siitä, kumpi valita, paperijulkaisua myötäilevä kuva- tai sivunkuvauskieliformaatti vai dokumentin sisällön ja ulkoasun erotteleva formaatti - niin mitä asiasta ajattelee lukija?

Göteborgin yliopiston kirjastotieteen opiskelija Mats Dahlström kirjoittaa (Dahlström):

[...] Oron för att det vägval högskolorna står inför görs i skymundan, i skydd av det till hälften omedvetna beslutets mörker. Oron för att läsaren blir den som i slutändan förlorar makt i det ställningstagande institutionerna står inför. Oron för en kastrerad läsare.

Vaikka pelko vaikuttaakin hieman dramaattiselta, sitä ei sovi sivuuttaa. Samaan seikkaan kuin Dahlström on mielestäni Vesa Niinikangas viitannut, kun hän on todennut, että elektronisen dokumentin olisi säilytettävä dokumentin alkuperäinen integriteetti ja ne funktiot, jotka dokumentille asetettiin sitä kirjoitettaessa (Niinikangas).

Mutta miksi valita vain yksi vaihtoehto jakeluun? Jos lukijalle tarjotaan dokumentista useampia formaatteja, lukijalla on vapaus valita, mitä niistä haluaa kulloinkin käyttää. Aiheutuva lisätyön määrä riippuu siitä, minkälainen tallennusformaatti on valittu.

Osaaminen

Elektronisen julkaisemisen kaikki osa-alueet - mm. sisällöllinen, tekninen ja sopimusoikeudellinen (Oulun, 11) - vaativat taitoja. Oppiminen vie vuosia silloin kun lähdetään aloittelijatasolta liikkeelle. Suomen yliopistot ovat nyt suunnilleen sillä tasolla. Opetusministeriön rahoittamissa projekteissa on kyllä opittu yhtä ja toista, mutta saatu kokemus elektronisesta julkaisemisesta on ollut monelle ensimmäinen kosketus aihealueeseen.

Julkaisuteknisen osaamisen lisääminen yliopistoissa olisi kuitenkin tärkeää ainakin seuraavista syistä:

- kun tehdään päätöksiä yleisistä suuntaviivoista, vaihtoehtoihin osataan ottaa perustellusti kantaa
- pystytään arvioimaan, kuinka paljon mikin työ vie ajassa ja muissa resursseissa
- ollaan tasavertaisempi neuvotteluosapuoli, kun ryhdytään hankkimaan ostopalveluja tai -tuotteita
- kun osataan jotakin, pystytään tarkemmin määrittelemään se mitä ei osata, jolloin avun pyytäminen muualta helpottuu

OPM-hankkeiden selvimpiä kokemuksia on ollut se, että yhteistyö on a ja o. Atk-keskuksilla on asiantuntemusta ohjelmoinnista, tietojärjestelmien liittämistä toisiinsa, datamuunnoksista ja web-teknologiasta. Kirjastot tuntevat metadatan ja käyttäjäkunnan. Yliopistopainot ovat julkaisemisen eksperttejä. Tietohallinnon yksikköjen yhteispelillä saadaan paljon aikaan. Hyvänä esimerkkinä tästä on mm. Oulu.

Yhteenveto haastatteluista

Haastateltuja oli neljä ja he edustivat seuraavia tieteenaloja: biolääketiede, taidehistoria, tietotekniikka ja tuotantotalous. Seuraavassa lyhyt yhteenveto

haastattelujen tuloksista. Käsitellyt aiheet ja haastattelujen muistiinpanot ovat liitteissä.

Taidehistorioitsijaa lukuunottamatta kaikki ovat rutinoituneita tietokoneen käyttäjiä ja hallitsevat useita eri työpöytäohjelmistoja. Kaikki neljä käyttävät jatkuvasti Internetin kautta saavutettavia palveluja tieteellisessä työssään. Biolääketieteen edustajalla on yksi pää tietolähde, muut käyttävät eri lähteitä vaihtelevasti.

Mitä nopeammin ja suuremmin pääsee oman tieteenalan ja sitä sivuavien alojen globaalin elektronisen dokumenttivarannon äärelle, sen parempi. Tätä mieltä olivat erityisesti tietotekniikan ja biolääketieteen edustajat. Mutta tiedonhaun luonne on kytköksissä tutkimusmetodeihin ja tieteenalan muihin käytäntöihin. Taidehistorian edustaja painotti luovan, sattumanvaraisenkin etenemisen mahdollisuutta tiedonhaussa. Tuotantotaloudessa taas kotimaan lähteet ovat ensisijaisia, ulkomaiset tiedonlähteet löytyvät usein lähipiirin suositusten ja henkilökohtaisten kontaktien kautta.

Elektronisen tekstintekemisen taitojen ja välineiden käyttäminen on lujasti kiinni paitsi tietysti tutkijasta itsestään niin myös tutkijan lähipiiristä: laitoksesta, laboratoriosta, tutkimusryhmän vetäjästä, tutuista väittelijöistä jne. Seuraavaksi tärkein vaikuttaja on tieteenala, sen aikakauslehdet ja muut levityskanavat. Viimeisenä tulee yliopisto.

Yliopiston omien julkaisemisen ohjeiden toivotaan olevan luonteeltaan suosittavia, ei määrääviä. Suositusten tulee olla hyvin perusteltuja ja niiden onnistuneesta käytöstä tulee olla näyttöä. Valmiita mallipohjia on hyvä olla saatavilla, mutta niitä täytyy saada muuttaa. Mallipohjia on oltava ainakin Wordille ja (La)TeX:lle.

Väitöskirjan tekemisessä pyritään laadukkaaseen, näyttävään ja toimivaan lopputulokseen. Niinpä kaikenlaista teknistä kikkailua pyritään välttämään, jottei lopputuloksesta tule arvaamatonta. Kikkailuksi luetaan kaikki still-kuvia monimutkaisempi materiaali kuten digitaalinen video, animaatio, simuloinnit jne. Mutta jo passiivisen kuva-aineistonkin käsittely voi olla hankalaa ja aineiston taitto työlästä.

Elektronisen tekstin plagiointia ei pidetä kovin merkittävänä ongelmana elektroniselle julkaisemiselle. Samansuuntaisia tuloksia saatiin tuoreessa brittiläisessä kyselytutkimuksessa²⁸, samoin Oulun yliopiston kirjaston talvella 1998 tekemissä haastatteluissa²⁹. Tutkimusideoiden ja -tulosten on tarkoituskin levitä, sen parempi mitä laajemmalle. Jos tutkija lähtee kopioimaan tutkimustuloksia, hän on pian entinen tutkija. Toisaalta voi olla, että opinnäytteen tekijän valmistumisen paineet saattavat houkuttaa nuorta tutkijanalkua varkauteen. Tutkijan eettiset säännöt on osattava ja ne täytyy tehdä selväksi viimeistään jatko-opintojen alkuvaiheessa.

Opinnäytteen tekeminen on tieteellisen kirjoittamisen taitojen oppimista: sanaston, ilmaisutyylin, tekstin rakenteen, havainnollistamisen, viittauskäytännön. Kirjoittaminen on työnä henkilökohtainen punnerrus, jossa tietokone ja tekstinkäsittelyohjelma voivat vain yrittää auttaa. Haitaksi ne eivät saa olla.

Väitöskirjan rooli ja status vaihtelevat erittäin paljon tieteenaloittain. Luonnontieteissä kuten biolääketieteessä väitöskirja lähestyy pelkkää akateemista symbolia, jonka

²⁸ <http://libcd2.lib.ed.ac.uk/Theses/implicat.htm>

²⁹ Lähteet: Vesa Kautto, Reino Lipponen.

substanssi on artikkeleissa. Toisessa ääripäässä kuten taidehistoriassa se on elämäntyö, kulttuurihistoriaa tallentava monografia. Puolivälissä ovat sitten ne tieteenalat kuten tietotekniikka ja tuotantotalous, joissa väitöskirjan materiaalilla on kaupallista potentiaalia.

Lopuksi

Document management isn't just about access control or searching, it's about implementing and enforcing business rules for information management. Every group's rules are different, so every management system will be different.

W. Eliot Kimber, postitus uutisryhmään comp.text.sgml 31.3.1998

No single approach can be valid for all countries and all situations. So it is up to you to work out your own policy for electronic information in collaboration with everyone else involved.

Guidelines on best practices for using electronic information, How to deal with machine-readable data and electronic documents. DLM Forum, 1996/7. <http://www.echo.lu/dlm/en/gdlines.pdf>

The recommendations are therefore generic ones and it is intended they should be used with the case studies and bibliography, resources and references, which provide pointers to detailed guidance and application-specific practices.

N. Beagrie and D. Greenstein: A Strategic Policy Framework for Creating and Preserving Digital Collections. Arts and Humanities Data Service Executive, Version 4.0, 14/7/98 Final Draft, <http://ahds.ac.uk/manage/framework.htm>

Elektronisen julkaisemisen kartoitukset ja suositukset ovat yleensä konsulttitoimistojen tekemiä tilaustöitä, hengästyttävän pitkiä dokumentteja, joissa käydään läpi standardeja ja teoretisoidaan niiden käytöllä. Tämä kartoitus ei ollut poikkeus, onhan teillä luettavana vielä liitteetkin! Mutta asialla on vakavampikin puolensa. Yliopistot tarvitsevat akuutteja neuvoja.

Ehdotan, että FinELib kehottaa Suomen yliopistokirjastoja menemään suoraan mukaan joko joihinkin edellä esitellyistä hankkeista tai esimerkiksi Lappeenrannan teknillisen korkeakoulun hankkeeseen, joka on alkanut tiivistelmästä (Knuutila). Se voisi olla hyvinkin realistinen tapa edetä varsinkin niissä yliopistoissa, joissa artikkeliväitöskirja on vallitseva väitöskirjan muoto.

Ehdotan, että se, miten tämä yhteistyön tiivistäminen voisi käytännössä tapahtua, olisi FinELib-suositusten ydin. Käytännössä tämä tarkoittaisi sitä, että FinELib käynnistäisi neuvottelut em. tahojen kanssa muutaman "julkaisuklusterin" perustamisesta Suomeen. Jotta lukijakuntaa palveltaisiin hyvin, klustereiden olisi pystyttävä keskustelemaan keskenään teknisesti vähintään metadatatason tasolla siten, että klustereiden tuottamat julkaisut alkaisivat vähitellen näyttää ulospäin yhtenäiseltä väitöskirjojen elektroniselta kokoelmalta.

Itse asiassa tämä ehdotus ei ole aivan itse keksimäni; Tampereen yliopisto tulee tietämäni mukaan ehdottamaan laajempaa yhteistyötä syksyllä. Ele on kannatettava,

mutta olen kuitenkin itse sitä mieltä, että yksi ja ainut julkaisemisen malli ei ole välttämättä hyväksi. Siihen on kaksi syytä.

Kuten edellä esitin, yliopistot tarvitsevat monipuolista osaamista. Mutta kaikkea osaamista voi olla mahdoton saada jaetuksi tasaisesti joka paikkaan. Yliopistoilla olisi mahdollisuus nostaa omaa osaamisen tasoaan yhden julkaisuklusterin sisällä ja yhteydessä muihin.

Toinen ja tärkeämpi syy on se, että yhtä oikeaa tapaa järjestää elektroninen julkaiseminen ei ole. Emme voi tietää, mihin suuntaan web kehittyy. Meillä ei ole tietoa siitä, miten elektronisten julkaisujen arkistointi ja elvyttäminen onnistuu tulevaisuudessa. Muutamalla valitulla suunnalla eteneminen samanaikaisesti vähentää epävarmuutta. Pysyisimme "omavaraisina kriisin sattuessa", kuten tavataan sanoa.

FinELib on osoittanut voimansa lisenssikonsortioiden muodostamisessa. Se voisi nyt osoittaa taitonsa myös julkaisutoiminnan organisoinnissa.

Lähteet

Bishop, A: Understanding Use in the Real World. IEEE ADL'98, Advances in Digital Libraries, 24.4.1998, Santa Barbara, Kalifornia. (Suullinen esitelmä Illinoisin DLI-projektin käyttäjätutkimuksista)

Dahlström, M.: En nod att stilla bedja om. Human IT, 1998:3. <URL: <http://www.hb.se/bhs/ith/3-98/md.htm>>

Dokumentteja digitaalisesti – tarpeen mukaan, digitaalinen julkaiseminen Jyväskylän yliopiston kirjastossa. <URL: <http://www.jyu.fi/library/julkaisut/Kuvat/digijulk.pdf>>

Fox, E.: Digital dissertations network worldwide. Joint Session of IuK99 Dynamic Documents and Satellite Conference Dissertations Online, 24.3.1999, Jena, Saksa. (Suullinen esitelmä)

Gulbrandsen, Are D (1999): Hovedfagsoppgaver i digitalt bibliotek - En gjennomgang av løsninger og problemstillinger knyttet til innlevering av og elektronisk publisering av hovedfagsoppgaver og "grå litteratur" ved UiO. <URL: <http://www.ub.uio.no/ubit/hopp/vev/Prosjektpresentasjon/>>

Knuuttila, U. ja L. Tissari: Elektroninen julkaiseminen LTKK:ssa. Teoksessa: Rakenteinen...

Niinikangas, V.: Puheenvuoro RAJU-seminaarissa 18.3.1999.

Oulun yliopiston elektroninen julkaisutoiminta. Suunnitelma. Oulu 1999. Elektroninen versio: <URL: <http://herkules.oulu.fi/nbnfi-fe19991068/>>

Owens, E.: Are publishers obsolete? IuK99 Dynamic Documents 23.3.1999, Jena, Saksa. (Suullinen esitelmä, Univ. of Chicago Pressin edustaja)

Palonen, V.: Yliopistot sähköisinä julkaisijoina ja dokumentoijina. Teoksessa: Rakenteinen...

Rakenteinen julkaiseminen yliopistoissa, RAJU-seminaari 18.3.1999. Toim. T. Sonkkila, V. Varjonen. Elektroninen versio: <URL: <http://herkules.oulu.fi/isbn9514252446/>>

Varjonen, V.: Elektronisen julkaisemisen hankkeet Oulun yliopistossa. Teoksessa: Rakenteinen...

Wilkinson, R. et al.: Document computing: technologies for managing electronic document collections. Boston : Kluwer, 1998. (The Kluwer international series on information retrieval; 5). ISBN 0-7923-8357-5

**Yliopistojen elektronisen julkaisemisen
kartoitus - Liitteet**

Tuija Sonkkila

Teknillisen korkeakoulun kirjasto

16.11.1999

Sisällysluettelo

LIITE 1 HAASTATTELUT.....	3
AIHEET	3
BIOLÄÄKETIEDE.....	4
TAIDEHISTORIA.....	6
TIETOTEKNIikka.....	7
TUOTANTOTALOUS	9
LIITE 2 OPM-RAHOITTEISET JULKAISEMISEN HANKKEET.....	13
JYVÄSKYLÄN YLIOPISTO	13
OULUN YLIOPISTO.....	13
TEKNILLINEN KORKEAKOULU	14
VAASAN YLIOPISTO.....	14
LIITE 3 TILANNEKATSAUS RUOTSIIN JA NORJAAN.....	15
RUOTSI	15
<i>Göteborg</i>	15
<i>Karlskrona/Ronneby</i>	15
<i>Karolinska Institutet</i>	17
<i>Linköping</i>	17
<i>Lund</i>	17
<i>Luulaja</i>	18
<i>Uumaja</i>	18
NORJA	18
<i>Oslo</i>	18
LIITE 4 KANSAINVÄLINEN YHTEISTYÖ - ESIMERKKINÄ NDLTD.....	20
VIRGINIA TECH.....	20
HUMBOLDT UNIVERSITÄT ZU BERLIN	21

Liite 1 Haastattelut

Aiheet

- Kuka väitöskirjan tekee, julkaisee ja käyttää? Kuka sen tekemisen maksaa?
- Onko väitöskirja opinnäyte, tieteellinen julkaisu vai jotakin muuta?
- Mikä sen funktio on dokumenttina? Kuka tai mikä sen määrittelee? Onko sillä markkina-arvoa?
- Mikä on väitöskirjan levikki?
- Mikä on väitöskirjojen käytetyin julkaisufoorumi? Entä halutuin?
- Onko väitöskirjan teknisellä tekemisellä opetustehtävää? Voisiko sillä olla? Haluttaisiinko sitä?
- Onko väitöskirjan ulkoasulla jotakin erityistä merkitystä?
- Sisältääkö väitöskirja multimediaalisia elementtejä, joita lukija saattaisi haluta "käyttää" (simuloida, todentaa, kokeilla, leikkiä) jos voisi? Miten kaukana nykytodellisuudesta nämä ajatukset ovat?
- Mikä on väitöskirjan odotettavissa oleva elinikä?
- Onko jollakulla vastuuta väitöskirjan säilyvyydestä?
- Tulevaisuuden väitöskirja: minkälainen se olisi luonteeltaan, fyysisiltä ominaisuuksiltaan ja käyttötavaltaan?
- Tuoko kansainvälistyminen ja tutkimusryhmäpainotteinen väitöskirjan tekeminen mukanaan mitään muutoksia väitöskirjan olomuotoon tai käyttötarkoitukseen?

Biolääketiede

Lääketieteellisen kemian osasto, Helsingin yliopisto. Haastateltava on tutkimusryhmän vetäjä ja kokopäiväinen Suomen Akatemian vanhempi tutkija, väitellyt vuonna 1979. Haastattelupäivä 8.7.1999

Melkein kaikki alan väitöskirjat ovat artikkeliväitöskirjoja; 3-5 artikkelia ja johdanto. Käytännössä johdanto on muodollinen, copy-paste -tekniikalla artikkeleiden pohjalta tehty, välttämättömyyden pakosta rustattava osa väitöskirjaa. Kopioimisen ymmärtää, ei oikein ole järkeä yrittää keksiä saman asian kuvaamiseen uusia muotoiluja.

Väitöskirjalle on vaikea löytää julkaisufoorumia, muutamia mahdollisia sarjoja on Suomessa (mm. Commentations Physica-Mathematicae ja Kansanterveyslaitoksen sarja), mutta haastateltava ei ollut tietoinen, vieläkö näitä ylipäänsä on olemassa. Yleensä väittelijä etsii mahdollisimman halvan painotalon, jossa sitten painattaa työn kansiin. Määrä on kyllä aivan liian suuri, muutama riittäisi sukulaisille jaettavaksi. Ainakin professoreille jaellut kappaleet joutuvat joko paperikoriin tai muutama voi mahdollisesti palauttaa postissa. Kenelläkään ei ole aikaa lukea muuta kuin oman spesifin alansa julkaisuja. Kaupallisella kustantajalla väitöskirjoja ilmestyy tuskin koskaan.

Väitöskirja tehdään yleensä erityisissä tutkijakouluissa ja väitöskirjan luonnekin on muuttunut "koulumaisemmaksi". Väitöskirjan tekeminen on ongelmanratkaisua ryhmässä, tutkimusryhmän johtajan johdolla. Johtaja antaa aiheen ja vastaa hyvin pitkälle myös artikkelien lopullisesta asusta ennen toimitusta lehdille. Tämä on käytännön realiteetti. Koko tutkimusryhmän imago kärsii, jos siltä lähtee huonolaatuista materiaalia.

Haastateltava oli sitä mieltä, että nuoret tutkijat eivät vielä hallitse tieteellisen kirjoittamisen taitoja tarpeeksi hyvin. On osattava kirjoittaa selkeästi, ytimekkäästi ja eksaktisti, lavertelua ei sallita. Teknisten mittalaitteiden käyttökokemus on niin ikään varsin heikkoa, erityisesti tyttöjen keskuudessa.

Tohtorin tutkintoa ollaan tuomassa alas perustutkinnon tasolle ja väittelijöiden ikä on laskussa. Tämä ei välttämättä ole hyvä asia. On vaikea sanoa, tuleeko tällä tavalla alalle hyviä tutkijoita vaiko ei. Ainakaan kokemusta ei ole vielä ehtinyt karttua tieteen tekemisessä. Vakituisia tutkijanpaikkojakaan ei ole tarjolla kuin yksi kymmentä väitellyttä kohti. Toisaalta voi ajatella, että ainahan väitöskirjasta kuitenkin oppii, erityisesti kriittisyyttä. Mutta aikaisemmin tuoret tohtorit olivat tietävämpiä ja näkivät työnsä eteen vaivaa.

Haastateltavan tutkimusryhmässä käytetään datan- ja tekstinkäsittelyyn MS Exceliä ja Wordia. Exceliin tuodaan data ja sen perusteella piirretään plotteja ja graafeja. Word on käyttökelpoinen väline tekstin tekemiseen ja sen kaavaeditori on riittävä mahdollisten kaavojen tuottamiseen. Tekstin tekoa helpottaa autotext-piirre, jolla voi makron tapaan automatisoida usein esiintyvien fraasien kirjoittamista. Se vähentää myös kiusallisia kirjoitusvirheitä.

LaTeX:sta haastateltavalla oli kaukaisia eikä kovin miellyttäviä muistoja: arkaainen käyttöliittymä ja paljon kryptisiä komentoja.

Julkaisut ovat tekstipainotteisia. Liikkuvan kuvan ja muun multimedia-aineiston lisääminen aineistoon on työlästä ja hidasta ja vaivalla saatu hyöty kyseenalaista.

Teksti ajaa useimmiten täysin saman asian. Tekninen kikkailu tuo mukanaan myös epävarmuustekijöitä: toimiiko ja miten kauan? Tosin erilaiset simuloinnit yleistynevät ja niillä voi olla paikkansa opetustehtävissä vaikkei niillä vielä olekaan merkittävää roolia tieteenalalla. Mutta tämä vaatii pätevää ja nopeaa apu työvoimaa. Onko siihen lisäresursseja tulossa?

Tutkimusryhmän kesken vaihdetaan tekstejä joko suoraan tai mailien liitteenä. Usein proseduurilla on se, että tutkimusryhmän johtaja tekee korjauksia työn alla olevaan tekstiin siten, että maalaa käsikirjoituksen korjattavat kohdat eri värillä tai käyttää poikkeavan väristä kirjasinta.

Alalla ei ole preprint-palvelimia eikä haastateltava myöskään tallenna artikkeleja tai muita tekstejään paikallisille palvelimille. Hän ei myöskään näe kovin tärkeiksi paikallisten tutkimus- ja julkaisutietokantojen päivittämistä tai niiltä tietojen etsimistä. Nopein ja käyttöliittymältään selkein palvelu on Public Medline -tietokanta. Jos hän tarvitsee tietoa jonkun tutkijan julkaisuista, hän hakee sen sieltä eikä lähde selaamaan yliopistojen web-sivustoja. Samanlaista "one stop shop":ia hän toivoisi myös fysiikkaan ja kemiaan, koska hänen tutkimusalaansa sivuaa myös niitä.

Medline on haastateltavalle ihanteellinen tietolähde. Hän etsii artikkeleita hakusanoilla tai -fraaseilla ja rajaa haun tulosta niin, että päätyy lopulta n. 50 viitteeseen. Käy sitten ne läpi. Ei näe oleellista käyttöä rakenteiselle haulle. Jonkin termin hakeminen esimerkiksi vain kuvateksteistä on tarpeetonta, koska termin on esiinnyttävä myös ympäröivässä tekstissä. Medlinen artikkeliviitteistä on usein suora pääsy kokotekstiin tai ainakin abstraktiin. Kokoteksti on yleensä PDF-muodossa, joka on helppo tulostaa. Jos kokotekstiä ei ole, hän pyytää paikallista kirjastoa tilaamaan sen. Helpon ja nopean tietolähteen ainut ongelma on tietostressin uhka.

Alan tieteelliset aikakauslehdet ottavat vastaan Word- tai WP-tiedostoja, jos ne on työstetty lehden vaatimalla tavalla (kirjasinkoko, marginaalit jne.) On myös ohjelmia (mm. EndNote), jotka tekevät automaattisesti tietyn kustantajan vaatimat viilaukset Word-tiedostoon. Tekstit lähetetään postissa useampana printtinä ja mukaan laitetaan dokumentti joko korpulla tai nykyään usein myös rompulla, jos kustantaja haluaa kuvat niin korkean resoluution formaatissa, että materiaali ei mahdu korpulle. Referee- ja editointiprosessin kuluessa kirjoittaja sitten korjaa ja työstää tekstiään tarvittaessa edelleen, tarkistaa lopuksi vielä oikovedoksen.

Haastateltava ei käytä Wordin tyylitoimintoja, lähinnä siksi, ettei ole koskaan tullut opetelleeksi niitä ja koska niistä ei vaikuta olevan varsinaista hyötyä. Mikäli väitöskirjojen tekemiseen tehtäisiin mallipohjia, niiden olisi ehkä syytä olla osastokohtaisia, koska käytännöt poikkeavat toisistaan.

Tutkimustyö alalla on hyvin kansainvälistä. Artikkeleiden lukijat ovat todennäköisesti ulkomailta. Niinpä internetin kautta saavutettavat tietolähteet levittävät tietoa parhaiten.

Väitöskirjojen elinikä on täysin kiinni aiheesta ja työn laadusta - yläikärajaa ei ole. Laadukkaasta väitöskirjasta on vastuussa väittelijän ohella erityisesti tutkimusryhmän johtaja; väitöskirjan tarkastajille työ sinänsä kuuluisi mutta käytännössä näin ei tapahdu. Syynä kiire tai tietämättömyys.

Elektronisen tekstin plagiointi on enemmänkin oman työn kopiointia (copy-paste), mutta siihenhän tutkijalla on oikeus. Luonnollisesti olisi myös helppoa esim. kopioida

muutamia valikoituja lauseita jonkun muun tutkimusryhmän tutkimussuunnitelmista ilman että kukaan huomaa mitään. Mutta jos ryhtyy kopioimaan tutkimustuloksia, se on sitten jo toinen juttu, ja siitä jää kiinni jossain vaiheessa. Jos tutkijanurallaan ryhtyy varkaaksi, silloin on aika desperado-meininkiä muutenkin eikä ura voi jatkua kovin pitkään. Originaalitutkimuksen tuottaminen on kunniakysymys. Yleensä tutkija pyrkii tuomaan esiin oman mielipiteensä ja esittämään asian vielä paremmin ja perustellummin kuin muut.

Kaupalliset kustantajat saattavat kyllä olla kiinnostuneita rakentamaan hakurutiineja, jotka vertaavat käsikirjoitusten lauseita verkosta löytyviin lauseisiin.

Elektronisessa julkaisemisessa on tärkeintä lopputulos: tekstin täytyy olla helposti tulostettavissa ja luettavissa ruudulta. Jos linkkien ylläpito on järjestetty, silloin niistä on hyötyä. Jos lähdetään standardoimaan kirjoitusvaihetta tai välineitä, sen on tuettava tutkijan ajatuksenjuoksua. Tekemisessä eivät auta shabluunat ja apuväki, tärkein työväline on tutkijan aivot. Tekstin header-osaan on mahdollista syöttää tarvittavat tunnistetiedot tiedonhakua varten.

Nippuväitöskirjan julkaiseminen sujuisi varsin hyvin niin, että julkaistaan verkossa johdanto ja niiden lisäksi linkit (passiiviset, jos niin halutaan) artikkeleiden verkkoversioihin, jos ne ovat saatavilla.

Taidehistoria

Helsingin yliopisto. Haastateltava on vuonna 1997 väitellyt Suomen Akatemian nuorempi tutkija. Haastattelupäivä 7.7.1999

Taidehistorian laitoksella on toivottu, että väitöskirjojen tuotannossa voitaisiin pitäytyä perinteisessä muodossa. Siellä väitöskirjan muoto on poikkeuksetta monografia ja luonteeltaan väitöskirjat ovat kotimaista, suomenkielistä kulttuurihistoriaa. Työ on monelle "elämäntyö", pitkälinen prosessi. Työtä ei tehdä kompetenssin vuoksi tai tutkimusrahoituksen turvaamiseksi vaan puhtaasti tutkimuksen itsensä takia. Laitos ei tavoittele tehokkuuspisteitä. Tutkimusyhteistyö on varsin vieras käsite, työ on itsenäistä ja yksinäistä(kin). Ulkomaiset kontaktit alalla syntyvät lähinnä museoiden kautta.

Väitöskirjojen ikä on pitkä, mm. Salme Sarajas-Kortteen 60-luvun väitöskirja on edelleen paljon käytetty lähde.

Kaupallisia kustantajia on erittäin vaikea löytää. Kotimainen kustannuspolitiikka on alkeellista. Pienillä kustantajilla ei ole tarpeeksi resursseja, suuret taas haluavat kustantaa vain taatusti myyviä nimekkeitä. Tällöin ei myöskään bibliografioita suvaita. Yliopistopaino pyrkii niin ikään kaupalliseksi kustantamoksi.

Suomessa ei ole yhtään peer review -tasoista alan lehteä ja vain yksi tieteellinen sarja, Taidehistorian seuran Taidehistoriallisia tutkimuksia.

Monografia verkossa on hankala käsitellä. Webiä käytetään tiedon etsimiseen jostain melko spesifistä asiasta ja lukija ei malta syventyä pitkiin ajatusrakennelmiin.

Kuvien käyttö verkossa tulee kalliiksi ja tekijänoikeusasiat ovat vielä keskeneräisiä. Esimerkit eivät kannusta: tekijänoikeusjärjestöt ovat estäneet museoiden ensimmäiset yritykset esittää taidehistorian merkkiteoksia verkossa. Tekijänoikeuteen tulisi suhtautua verkossa eri tavalla kuin painetussa materiaalissa. Tutkija haluaakin

tulostensa ja työnsä leviävän. Taloudelliset intressit luonnollisesti ovat syynä siihen, miksi tekijänoikeudesta taitetaan peistä niin paljon.

Carlöhmanneja on ollut jo ennen nettiäkin! Varkauksia ja petoksia voi tehdä yhtä hyvin painetusta materiaalista. Toisaalta on tuskin tapauksia, joissa painettu kirja väittäisi olevansa esim. Harvard University Pressin kustantama vaikkei todellisuudessa sitä olisikaan. Netissä tämä huijaus onnistuu helposti.

Taidehistoria nojaa vahvasti lähdekritiikkiin; on pystyttävä vakuuttumaan lähteen tekijästä, taustaorganisaatiosta jne. Verkon teksteillä ei tätä vielä ole. Samoin on ongelmallista se, että haun tuloksena putkahtaa keskelle tekstiä, jonka kontekstista voi olla erittäin vaikea sanoa mitään. Kokonaisuus ei hahmotu ja navigointitekniikkakin ontuu: usein ei ole linkkiä koti- tai aloitussivulle.

Verkkoresurssiin saattaisi voida luottaa paremmin, jos sen ylläpitäjä olisi luonut sille rekisteröintipalvelun. Tällä tavoin myös käyttäjä saa tietää jotakin ylläpitäjästä. Maksullisuus ei kuitenkaan ole laadun tae.

Netin käyttöön liittyvien pelkojen, toiveiden ja fantasioiden luonne muistuttaa 700-800 -luvun kuvariitaa. Kirkon piirissä kiisteltiin siitä, edustavatko kuvat itse jumaluutta vai ovatko ne vain keino lähestyä sitä.

Huolen aiheena on verkon tasapäistävä vaikutus. Jos asiat löytyvät verkon hakukoneilla liian helposti (joko näennäisesti tai todella), katoaa jotakin olennaista. Tekijän täytyy nähdä vaivaa asiansa eteen, jotta prosessissa on mahdollista myös löytää jotakin, joka ensi näkemältä saattaa vaikuttaa epäolennaiselta.

Haastateltava haluaisi tukea ilmaisten ja "pienten" järjestelmien ja välineiden käyttöä. Hän pelkää, että standardointi ja tarkat ohjeet tuhoavat tieteen tekemisestä luovuuden.

Haastateltavan oma työ oli ollut raskas kirjoitusprosessi ja hän oli päättänyt koskaan enää olla kirjoittamatta lineaarista tekstiä. Mutta myöhemmin mieli on muuttunut. Nyt hän on erityisen kiinnostunut juuri kirjamuodosta, myös kirjasta fyysisenä esineenä. Kirja on tiivis paketti, jossa on myös runsaasti tekstin sisäisiä linkkejä. Kirjalla on myös fetissiarvoa, ideaaleimmillaan se on kaunis ja laadukas esine.

Haastateltava tunsi syvää huolta humanistisen tieteentekemisen edellytyksistä ylipäättään Suomessa. Amerikasta on plagioitu vain tehokkuuden ja kustannusvastaavuden opit mutta unohdettu ottaa mallia amerikkalaisesta yliopistomaailmasta muuten: nelivuotinen college takaa kaikille perussivistyksen. Suomessa kaikki pisteytetään ja mitataan jo alusta lähtien.

Tietotekniikka

Neuroverkkojen tutkimusyksikkö, Teknillinen korkeakoulu. Haastateltava on vuoden sisällä väittelevä jatko-opiskelija. Haastattelupäivä 5.7.1999

Oma työyhteisö tallentaa väitöskirjat verkkoon jo nyt. Paras väline kirjoittamiseen on LaTeX ja sen latex2html-konvertterilla saatu html-versio paras jakeluun, koska perusselaimet osaavat näyttää html:ää. Latex2html muuntaa kuviksi ne kaavat, jotka osaa. Muulla tavoin tehdyt dokumentit leviävät huonosti, ovat hankalia käyttää eivätkä internetin hakukoneet pysty indeksoimaan niitä hakuun; tämä on mm. PS-tiedostojen hankaluus. LaTeXia käytetään muutenkin paljon, joten se on luonnollinen työväline.

Tärkeintä on se, miten tehokkaasti väitöskirja leviää, että se tavoittaa ne, joille se on tarkoitettukin Tärkeää on myös looginen esillepano, ulkoasulla ei niinkään ole merkitystä. On myös hyvä, jos dokumentin voi tulostaa helposti. Hypertekstimäisyyttä tarvitaan oikeastaan vain sisällysluettelosta sisältöön ja tekstistä viitteisiin. Lähdeviite saisi mieluiten näkyä jo silloin kun kursori on viitteen kohdalla, ettei klikkausta tarvita.

Interaktiivinen materiaali on työlästä ylläpitää, vaikka se periaatteessa olisikin kiva olemassa (mm. kaavat, joihin voi syöttää omia arvoja). Vain tekijä tietää, miten materiaalin pitäisi toimia. Poikkeuksena voisi ajatella videota tms. "puoli-interaktiivista" materiaalia.

Välineiden valinta ja käyttöönotto ovat itseorganisoituvia asioita ja hyväksi havaitut välineet saavat käyttäjiä hyvien mallien avulla. Mm. hyväksi havaittuja LaTeX-tyylisivuja annetaan kollegoille ja tutkimusryhmän muille jäsenille. Usein näkee myös väitöskirjojen kiitoskirjoituksissa erityiskiitokset jonkin tyylisivun tekijälle tai juuri edellä väitelleelle.

Tutkimusryhmät ovat tärkeitä. Ilman niitä tekemisen konventiot eivät leviäisi. Nyt motivaatio tehdä asiat jollakin tietyllä tavalla tulevat alhaaltapäin, eivät ylhäältäpäin ohjausmekanismien tapaan. Suosituksissa on se huono puoli, että niissä ei voida ottaa huomioon kaikkia todellisia tilanteita. Jos suositukset ovat väärille työvälineille (ei Wordia, kiitos!) tai ne eivät jousta käsillä olevassa tilanteessa, niitä ei käytetä. Pitää olla myös näyttöä siitä, että suositusten osoittama tekniikka toimii ja että sitä on menestyksellä käytetty. Tarvitaan siis suositusten verifiointia. Kun tämä puoli on kunnossa, sitten on vain käyttäjistä itsestään kiinni, miten hyvää jälkeä tekee.

LaTeX ja muut unix-ympäristöstä lähtöisin olevat välineet ovat suosittuja ilmaisuvoimaisuutensa ja vapausasteensa vuoksi: jos käyttäjä kohtaa jonkin puutteen esim. mallipohjassa tai ohjelmassa, hän voi, jos osaa, tehdä muutoksen itse lähdekoodiin. Ongelmien yli pääsee omalla työllä.

Monet ovat uskollisia tietyille välineille ja lähes vihaavat joitakin muita. MS-vastaisuus on yleistä. Uusien välineiden opettelu vie myös aikaa ja huomio kiinnittyy väärin asioihin. Tosin nykyisissä välineissä voisi kuvien piirtäminen ja tekstin asettelu olla sujuvampaakin.

Väitöskirja on aloittelevan tieteen tekijän käyntikortti tieteen maailmaan, osoitus siitä, mistä tutkija on kiinnostunut ja mitä on saanut aikaan. Sitä voidaan käyttää monella tavalla: johdannon avulla voi tutustua uuteen aihealueeseen "käyttäjystävällisesti", joku voi katsoa missä kuuluisissa lehdissä tuo onkaan saanut juttunsa julki, yritykset hakevat sovellettavaa, muut tieteilijät ideoita, tutkimustuloksia jne.

Suurin osa alan väitöskirjoista on nippuväitöskirjoja.

Elektronisen tekstin plagiointi ei ole ongelma, huijarit ja varkaat jäävät kiinni. Tutkija on vain iloinen jos työ leviää, kunhan vain hänen nimensä mainitaan. Ideoiden on tarkoituskin levitä. Siksi verkkolevitys sopii oikein hyvin tieteelliselle tutkimukselle. Nippuväitöskirjat ovat kyllä tietynasteinen tekijänoikeusongelma verkossa. Useimmiten artikkelit mainitaankin vain viitteinä johdannon lopussa.

Artikkelien ilmestymisfoorumien ei välttämättä tarvitse olla kaupallinen lehti tai konferenssi. Oleellista on, että se on hyvätasoinen, sopiva, oikealta alalta.

Tutkija on tottunut tekemään artikkeliin tiedonhakua helpottavan abstraktin ja määrittelemään hakusanat.

Väitöskirjojen elinikä on enimillään 10-20 vuotta. Tutkijalla itsellään on tietty vastuu dokumentin saatavuudesta, mutta hankaluutena on henkilökohtaisen osoitteiston puute. Iki.fi on periaatteessa olemassa, mutta se ei tarjoa tallennusta. Yliopistojen omat tallennustavat ovat niin toisistaan poikkeavia ja hajallaan, että se hankaloittaa leviämistä ja hakua. Maantieteellinen hajaannus haittaa myös: kukaan ei lähde selaamaan yksittäisten yliopistojen sivustoja erikseen. Ideaaliarkistoiija voisi olla tieteellinen akatemia, kirjastolaitos tms, joka tarjoaisi yhtenäisen pääsyn maailmanlaajuiseen, saman alan dokumenttikokoelmaan.

Ideaaliväitöskirja ilmestyisi kaupallisen kustantajan monografiana, mm. Elsevierin. Sarjassa (esim. TKK:n Acta-sarja) julkaiseminen edistää myös leviämistä hyvin. Haastateltava ei pidä väittelyä omalta kohdaltaan mitenkään pakollisena, mutta työyhteisön jäsenyys edellyttää sitä. Ja jos väitöskirjan kerran tekee, sen on parasta olla laadukas.

Tuotantotalous

Työpsykologian ja johtamisen laboratorio, Teknillinen korkeakoulu. Haastateltava on vuonna 1998 väitellyt kokopäiväinen tutkija ja projektipäällikkö. Haastattelupäivä 15.7.1999

Suurin osa väitöskirjoista on monografioita, nippuväitöskirjoja on tehty vain muutama. Laadullisia case-tutkimuksia julkaisevia referee-lehtiä on varsin vähän ja myöskään ohjausta artikkelien tekemiseen ei oikein ole saatavilla. Paineita on kyllä nippuväitöskirjoihin, sillä niiden arvostus on lisääntynyt. Usein ratkaiseva tekijä on professorien omat preferenssit. Haastateltavan oma monogafiamallinen työ syntyi sujuvasti lisensointityön jatkeena.

Tuotantotalouden osasto maksaa väitöskirjojen painatuksen. Painettujen väitöskirjojen määrää voisi kuitenkin supistaa rajusti. Elektroninen väitöskirja toimisi FYI-tyyppisenä tiedotteena.

Väitöskirjoilla on myyntiarvoa ja niitä myös myydään. Tätä kautta osasto saa myös takaisin painatukseen menneitä rahoja. Kiinnostunutta ostajakuntaa on paitsi kotimaisten ja ulkomaisten tutkijoiden myös kotimaisten yritysconsulttien piirissä. Konsulttien kiinnostus on ymmärrettävää, sillä osaston tutkimus on lähes kokonaan soveltavaa ja yhteistyö yritysten kanssa on tiivistä. Myös raportteja myydään. Yleistä kiinnostusta on erityisesti sellaisiin väitöskirjoihin, jotka käyttävät kaunokirjallisuutta läheisesti muistuttavaa tyyliä, joiden teema on ajankohtainen ja jotka käsittelevät teemaa yhteiskuntatieteelliseltä tai sosiologiselta kannalta (vrt. Kiiänmaa: Moderni totemismi)

Tutkijoiden piirissä on keskusteltu väitöskirjan tallentamisesta omalle kotisivulle esim. Word-dokumenttina (joitakin esimerkkejä on), mutta tapa ei ole yleistynyt. Haastateltava myönsi, että verkon kautta on ollutkin erittäin helppo saada ja tulostaa käyttöön erityisesti ulkomaisia töitä.

Elektronisessa muodossa olevan väitöskirjan helppo kopiointi voi tehdä tilaisuudesta varkaan; vain paatuneet tiederosvot ryhtyvän tilaamaan vaikeasti saatavia väitöskirjoja ja skannamaan niitä. Paineet nopeaan valmistumiseen voivat niin ikään edistää

copy/paste -tekniikan käyttöä. Tutkijakoulutuksessa on painotettava tutkijan etiikkaa. Toisaalta käytön esteitä ei saisi olla ainakaan tutkijayhteisölle. Voisi ajatella, että väitöskirjojen lukijaksi rekisteröityminen olisi riittävä pelote. Maksullinen käyttö on jo ongelmallisempaa, väitöskirja kun kuitenkin on julkinen asiakirja. Elektronisen materiaalin hinnoittelu on myös hyvin hankalaa; esim. haastateltava itse on tutkinut romppumuotoon pakatun tietotuotteen hinnoittelua.

Väitöskirja on opinnäyte isolla O:lla, mutta periaatteessa siihen suhtaudutaan silti kahdella tavalla. Joko se on ajokortti tieteentekemisen maailmaan, osoitus siitä, että hallitsee metodit ja tieteellisen kirjoittamisen. Tällöin väittelyikä on noin 30. Toisaalta se voi olla elämäntyö, jolloin väittely tapahtuu vasta noin 50-vuotiaana, pienen painostuksen saattelemana.

Väitöskirjasta ei ole opetusmateriaaliksi, ei ainakaan tenttikirjaksi. Tosin tutkimusmenetelmät -kurssilla jatko-opinnoissa analysoidaan väitöskirjoja, erityisesti niiden rakennetta.

3-5 vuoden ajan väitöskirjat ovat käyttökelpoisia, joskus myös 10 vuotta vanhat, jos niissä käytetty tutkimusmetodi on ajankohtainen ja rakenne mielenkiintoinen eli ei ole normaaliformaattia, ennalta ennustettavissa. Keskeinen asia on toimiva "kertomus".

Haastateltava olisi halunnut työtä aloittaessaan, että sen tekemiselle olisi ollut valmis malli, mieluiten Word-mallipohja, johon olisi voinut ryhtyä kirjoittamaan suoraan. Mitään tällaista ei kuitenkaan ollut olemassa, ei myöskään ohjeita ja tukea. Aikaa kului nyt paljon omien tyylien määrittelyyn ja lopuksi vielä kansien tekemiseen.

Laboratorion raporttisarjassa on kolme alasarjaa: kovatasoiset väitöskirjat (myös mahd. lisensiaatintyöt); working papers, joiden tason muutama professori tarkastaa; oppimateriaalit. Väitöskirjat julkaistaan pääsääntöisesti kaikki ensiksi mainitussa. Painopaikka on vapaasti valittavissa.

Haastateltava oli lähtenyt työstämään väitöskirjaansa siten, että oli luonut sisällysluettelon ja kirjoittanut jokaiseen kappaleeseen abstraktin sen sisällöstä. Kun tekstiä alkoi tulla lisää, hän jakoi kappaleet omiksi tiedostoikseen. Hän käyttää Wordia työvälineenä ja on tottunut käyttämään tyylejä. Tällä tavalla lineaarisesti struktuurin mukaan eteneminen ei sovi välttämättä kaikille, tarvittaneen hieman saksalaista mentaliteettiä! Mutta tämä on eräs keino saada henkinen yliote työhön. Hän arveli, että tyylejä kyllä käytetään, erityisesti nuoremman polven kirjoittajien keskuudessa (ongelmia lienee noin 40+ -vuotiailla). Ainakin tuotantotalouden opiskelijat ovat hänen mukaansa valveutuneita tietokoneen käytössä, teekkarit todennäköisesti yleisemminkin. Täällä myös tekninen infrastruktuuri on kunnossa, samoin esim. mikrotuki. Haastateltavalla oli toisenlaisia kokemuksia Ruotsista, jossa hän teki väitöskirjaansa muutaman kuukauden.

Teollisuustalouden laboratorion julkaisut sisältävät jonkun verran matemaattisia kaavoja, mutta Työpsykologian ja johtamisen laboratorion julkaisuissa niitä on hyvin harvoin. Osastolla käytetään yleisesti Wordia ja Exceliä sekä tilastollisia välineitä.

Kirjoittajalle on ollut tärkeää yhteistyö kielikeskuksen englannin kielen lehtorin ja osastokirjaston kanssa kielenhuollon ja tiedonhaun osalta. Samantapaista toimivaa suhdetta hän toivoi myös elektronisten julkaisujen tekemiseen.

Jollakin instanssilla pitäisi olla hanskassaan se, että korkeakoulun julkaisujen ulkoasu olisi yhteneväinen ja että kirjoittaja saisi tarvitsemaansa tukea esim. ulkoasun

viimeistelyssä. Tarvitaan toimittajaa, editoria. Mieluiten niin, että kirjoittajan ei tarvitsi huolehtia ulkoasusta ollenkaan, jos hän ei sitä haluaisi. Korkeakoulun elektronisten julkaisujen tyylin täytyy käydä läpi asiantuntijakäsittelyn, jotta dokumentit ovat luettavuuden kannalta optimaalisia. Väittelijälle pitäisi myös olla esittää todellisia esimerkkejä siitä, minkälainen teoksesta tulee kulloinkin käytettävissä olevilla tyyleillä. Mallipohjan pitäisi mieluiten olla tiettyjä vapauksia antava, ei liian tiukka. Voi myös tulla eteen tilanteita, joissa tyylien käyttö on jostain syystä ollut virheellistä, jolloin editorin on oltava yhteydessä kirjoittajaan, ettei lopulliseen versioon tule virheitä.

Haastateltava ehdotti, että elektroninen versio voisi noudattaa yleisiä ulkoasumäärittäyksiä mutta että paperiversion ulkoasu olisi joiltain osin kirjoittajan määriteltävissä.

Painetulla teoksella on hieman erilainen funktio ja status kuin elektronisella. Ulkoasulla on kiistatta iso merkitys, koska on kyse isosta henkisestä ponnistuksesta, jonka hedelmä väitöskirja on. Samoin, niin kauan kuin väitöstilaisuus on nykyisenkaltainen public defence -tilaisuus, väitöskirja olisi mukavinta olla paperilla jaettavissa, jotta yleisö voi seurata omista kappaleistaan vastaväittäjän huomautuksia. Jos siirrytään amerikkalaistyylliseen komiteaistuntoon, silloin paperiversion merkitys vähenee.

Hän ehdotti, että kirjasto voisi huolehtia esim. Word-mallipohjan esittelystä ja demoamisesta syksyn väitöskirjaseminaarissa, johon uudet jatko-opiskelijat kokoontuvat. Paikallisen opiskelijakunta voisi hyvinkin ajatella olevan tavoitettavissa tällä tavalla, mutta hankaluutena voivat olla ulkopuoliset ohjattavat, jotka tulevat ja menevät.

Kirjoittajan olisi hyvä tietää jo varhaisessa vaiheessa se, miten hänen oletetaan tekevän työnsä. Joudutaan vaikeuksiin, jos kirjoittaja tekee työn valmiiksi jollain omalla metodillaan ja lopuksi häntä vaaditaan siirtämään koko materiaali johonkin mallipohjaan. Vaikka nurinaa aina esiintyy, jos jotain laadittua sääntöä pitää noudattaa, eniten närää on kuitenkin silloin liikkeellä, jos ei ole hyvissä ajoin ohjeita siitä, miten homma pitäisi tehdä. Tieto pitäisi olla myös työn valvojalla. Koordinointia tarvittaneen myös osastotasolla. Periaatteessa vapaa ja rajoittamaton tekeminen tuottaa innovaatioita, mutta tieteellisen tehokkuuden kannalta se ei ole hyvä asia.

Vinkkinä hän sanoi, että näin tärkeä asia kannattaa markkinoida tyyliin "teille ei tule lisää työtä -- me neuvomme -- saavutettavat hyödyt ovat nämäjanämä --". Muuten reaktio voi olla se, että "taas hallinto pompottaa meitä".

Kirjoittajaa voi vaatia määrittelemään ainakin keywordit työlleen, mahdollisesti myös indeksin keskeisestä terminologiasta. Tiedonhaun kannalta olisi ideaalista, jos hakujärjestelmä esittäisi haettavan termin tai fraasin omassa kontekstissaan, jottei tarvitsi kahlata läpi koko työtä.

Ideaalinen elektroninen väitöskirja sisältäisi myös korkealaatuisia still-kuvia, animaatioita ja videokuvaa. Haastateltavan alalla esim. sen ympäristön näyttäminen, jossa tutkittava työ tapahtuu, kertoisi lukijalle paljon. Tosin silloin olisi oltava varovainen, ettei tässä vahingoiteta yritysten imagoa tai vaaranneta heidän kilpailuasemaansa. Multimedia tuo mukanaan myös uuden vaaratekijän: visuaalisen mässäilyn ja liiallisen yrittämisen. Tärkeintä on kuitenkin tieteellinen

asiakirjoittaminen, ja väitöskirjan tärkein kriteeri on kirjoittamisen taso.
Multimediasta ei saa tulla laatukriteeriä väitöskirjalle, ei ainakaan hänen alalleen.

Liite 2 OPM-rahoitteiset julkaisemisen hankkeet

Jyväskylän yliopisto

Osana OPM:n rahoittamaa EVA-projektia Jyväskylän yliopiston kirjasto on käyttänyt vuoden 1997 alusta lukien Xerox Oy:n mm. XDOD- laitteistolla opinnäytteiden skannaukseen ja tallennukseen. Kokoelmaan on vuoden 1998 loppuun mennessä skannattu n. 1050 opinnäytettä eli n. 90 000 sivua tekstiä

Vuoden 1998 alusta lukien opinnäytteet esitetään ja jaellaan verkossa kuvamuodossa, jota www-selaimet osaavat näyttää suoraan.

Tutkielmapankkiin liittyviä tekijänoikeus- ja julkisuuskysymyksiä pohtimaan on perustettu paikallinen professoritason työryhmä, jonka raportti valmistuu vuoden 1999 aikana. Tutkielmapankki on korostanut sitä, että elektronisen aineiston varastointiin ja verkkojakeluun liittyviä tekijänoikeudellisia ja asiakirjojen julkisuuskysymyksiä tutkittaisiin lisää.

Digitaalisen median maisteriohjelma on tutkinut tahollaan SGML/XML-pohjaisia julkaisemisen kysymyksiä erityisesti oppimateriaalin tuotannossa.

Oulun yliopisto

Oulun yliopiston kirjasto aloitti vuonna 1997 kokeiluluontoisesti Acta Universitatis Ouluensis -sarjan elektronisen julkaisemisen yhteistyössä Yliopistopainon, ATK-keskuksen ja julkaisutoimikunnan kanssa. Acta-sarjassa ilmestyy sekä väitöskirjoja että muita julkaisuja. Tavoitteena on laajentaa toimintaa muihinkin julkaisuihin.

Rehtori asetti loppupalvesta 1998 työryhmän, joka teki ehdotuksen keväällä 1999 siitä, miten elektronisten väitöskirjojen julkaiseminen tulvaisiin järjestämään. Työryhmässä olivat mukana mm. kirjasto ja Yliopistopaino.

Tavoitteena on rinnakkaisjulkaiseminen: väitöskirjoja painetaan tarvittava määrä, mutta elektroniset versiot jaellaan myös verkossa, mikäli väittelijä antaa siihen kirjallisen suostumuksen.

Järjestelmän tietokantana on Oracle, mutta muutkin tietokantaratkaisut ovat mahdollisia. ATK-keskus, jonka unix-palvelimella Oracle sijaitsee, hoitaa kannan teknisen ylläpidon ja varmuuskopioinnin. Tietokantaan tallennetaan dokumentin SGML-muotoinen kokoteksti. Tietokannasta saadaan väitöskirjojen viitetietoja mm. myynti- ja jakeluorganisaatioille sekä erilaisia listauksia ja luetteloita web-jakeluun.

Kirjoittajalla on kaksi mahdollisuutta luovuttaa teoksensa verkkojakeluun toimitettavaksi: joko painovalmiina tai muotoilemattomana tekstinä.

Painovalmis materiaali muunnetaan PDF-muotoon, muotoilematon taitetaan FrameMaker+SGML:llä ja tallennetaan SGML-muotoon. SGML:stä tuotetaan DSSSL-määrityksillä ja Jadella HTML-muotoinen versio niin ikään verkkojakeluun. SGML-dokumentti arkistoidaan Oracle-tietokantaan.

Järjestelmän tietokantana on Oracle. ATK-keskus, jonka unix-palvelimella Oracle sijaitsee, hoitaa kannan teknisen ylläpidon ja varmuuskopioinnin. Dokumentit voidaan tallentaa tietokantaan missä tahansa muodossa: SGML, PDF jne. Tietokannasta

saadaan väitöskirjojen viitetietoja mm. myynti- ja jakeluorganisaatioille sekä erilaisia listauksia ja luetteloita web-jakeluun.

Teknillinen korkeakoulu

Kirjaston HUTpubl-projektissa on tehty malli TKK:n julkaisusarjojen elektroniseksi julkaisemiseksi. Se soveltuu myös väitöskirjojen tuottamiseen. Malli on osa tietotekniikan diplomityötä.

TKK:n väitöskirjalautakunta on käsitellyt elektronisten väitöskirjojen tilannetta ja esittänyt kannanottonsa tietohallinnon johtoryhmälle toukokuussa 1999.

HUTpubl on saanut OPM:n hankerahaa kahteen otteeseen; toinen jatko projekti on alkamassa. Se päättyy 1.5.2000. Tarkoituksena on laajentaa projektissa tehtyä pilottia mm. yhteistyössä matematiikan laitoksen kanssa.

HUTpubl on ideoiltaan lähellä Oulun mallia, ja osittain myös käytetyt välineet ovat samat lukuunottamatta Oracle-tietokantaa. Projektien välillä on ollut runsaasti tietojen vaihtoa, muttei toistaiseksi konkreettista yhteistyötä.

HUTpubl on tekninen pilotti, jossa on sen opinnäytteestä johtuen eräitä lisäpiirteitä, joita Oulun mallissa ei ole. Molemmissa malleissa on kuitenkin lähdetty siitä, että elektroninen dokumentti tallennetaan rakenteisessa, pitkäaikaiseen säilytykseen soveltuvassa SGML-muodossa ja tästä tuotetaan tarvittaessa erilaisia jakelumunnoksia, kuten verkkojakeluun HTML tai PDF, paperidokumentti jne.

HUTpublissa on tehty kirjoittajalle valmis mallipohja Word- ja FrameMaker+SGML -tekstinkäsittelyohjelmiin.

Vaasan yliopisto

Vaasassa suunnitellaan opinto-oppaan ja julkaisusarjojen elektronista julkaisemista.

Molemmissa julkaisemisen hankkeissa tullaan käyttämään tiedonhallinnan pohjana joko yliopistolle hankittua Astoria-tietokantaa, jotakin relaatiotietokantaa tai tiedostojärjestelmää. Julkaisusarjojen tekijälle tarjotaan WP- ja Word-mallipohjat. Muunnos SGML:ään tullaan tekemään FMSGML:ssä, ja SGML-muodosta tuotetaan muunnosohjelmien avulla edelleen PDF-muotoa painoon ja web-jakeluun sekä XML-muodon kautta HTML/XHTML-muotoisia web-dokumentteja. Astoriaan tallennettaisiin dokumentit XML-muodossa. Astoria ei sovellu on-the-fly -HTML-muunnoksiin¹.

Vaasan kokemuksen mukaan muutama avainhenkilö tarvitaan, jotta hankkeet saadaan kunnolla liikkeelle. Pienessä yliopistossa on sinänsä helpompi saada idea läpi ja henkistä tukea uusille ajatuksille. Vaasassa on tietotaitoa, mutta toteutuksen ajoitus on kiinni resurssien suuntaamisesta. Kirjastonjohtaja Vuokko Palonen on käynyt läpi yliopistojen julkaisutuotantoa ja selvityksen pohjalta esittänyt sähköisen julkaisemisen ja dokumentoinnin yksikön perustamista (ks. Lähteet: Palonen)

¹ Ari Hovila, postitus keskusteluryhmään raju@hut.fi 4.5.1999

Liite 3 Tilannekatsaus Ruotsiin ja Norjaan

Ruotsi

Svenska Bibliotekariesamfundet on perustanut 1997 yhteistyöfoorumin nimeltään Ämnesgruppen för Elektronisk publicering av fulltextdokument². Tavoitteena on kerätä tietoa ja tarjota eritoten konkreettista, käytännönläheistä apua elektronisen julkaisemisen hankkeille. Erityistä huomiota on kohdennettu metadataan ja copyright-asioihin. Jäseninä ovat kaikki ne yliopistokirjastot, joissa on meneillään kokotekstiprojekteja tai jossa julkaiseminen on jo vakiintunutta.

Foorumi järjesti toukokuussa 1999 kokouksen, jossa käytiin läpi tilannetta jäsenyliopistoissa³. Seuraavat yhteenvedot perustuvat sekä kokousmuistioon että projektien omiin dokumentteihin.

Göteborg

Göteborgin yliopiston Handelshögskolanin elektronisten julkaisujen hanke on edennyt toteuttamisvaiheeseen: Handelshögskolans elektroniska publiceringscenter -niminen yksikkö perustetaan vielä tämän vuoden aikana. Henkilökunta: yksi puolipäiväinen kirjastonhoitaja, yksi puolipäiväinen tekninen henkilö. Pilottiprojektissa kirjasto muunsi PS- tai Word-muodossa tulleet dokumentit PDF-muotoon. Projektissa saatujen kokemusten perusteella päädyttiin de facto -jakelustandardiin PDF mutta todettiin myös XML:n tulevat hyödyt, kunhan selaimet pystyvät esittämään sitä ja editointi helpottuu⁴.

Karlskrona/Ronneby

Karlskrona/Ronneby on vuodesta 1996 lähtien tutkinut elektronisen julkaisemisen formaattivalintoja, mm. SGML:ää ja tehnyt myös käytännön kokeiluja. SGML-sovellusten epäyhteensopivuus on tullut heillekin tutuksi. Dokumentissa Rapport från projekt Elektronisk hantering av forsknings- och examensarbeten vid Högskolan i Karlskrona/Ronneby⁵ projektin vetäjät kirjoittavat:

I teorin ska en SGML-fil vara helt applikations- och systemoberoende, men detta har inte fungerat i praktiken. Ett sätt att lösa detta, är att skaffa sig mer kontroll över hela processen genom att i förväg definiera vilka programvaror som ska användas för informationshanteringen.

SGML merkitsee koko julkaisuketjun uudelleenajattelua, mikä tekee kirjaston aseman kieltämättä problemaattiseksi:

Det största problemet med att använda SGML för att lagra examensarbeten är att biblioteket är en isolerad ö i det system som producerar och konsumerar den informationen arbetena innehåller. I ett fortsatt arbete med att introducera SGML måste även högskolans examensår/forskare vara delaktiga i implementationsarbetet

² <http://www.lub.lu.se/ep/sbsep/sbseptop.html>

³ <http://www.lub.lu.se/ep/sbsep/minnesanteckningar990527.html>

⁴ <http://www.ub.gu.se/Ge/projekt/rapportv1.pdf>

⁵ <http://www.lub.lu.se/ep/sbsep/SGMLrap.html>

Projektin vetäjä Peter Linde tutustui HUTpubl-pilottiprojektiin syksyllä 1997. Linde koki SGML-pohjaisen julkaisemisen mielenkiintoiseksi mutta monella tavalla ongelmalliseksi: rakenteisen tekstin tekemisen oppimiseen kuluva aika, SGML-sovellusten vaikeudet matemaattisten kaavojen kohdalla, järjestelmän kokonaiskustannukset ja implementointivaikeudet olivat hänen päällimmäiset huolenaiheensa.

Erittäin tärkeää uuden julkaisujärjestelmän - tai minkä tahansa järjestelmän - suunnittelussa on kustannusten ja hyötyjen analysointi. Mitä lisäarvoa SGML toisi opinnäytteiden hallintaan? Karlskrona/Ronnebyn projekti päättyi toteamaan seuraavaa:

Svaret på den frågan är att det krävs mycket arbete, både på biblioteket och på högskolan, om man ska skaffa en fungerande SGML-lösning. Vår slutsats är därför att; det man skulle vinna på att lagra examensarbeten i SGML-format, i dagsläget, inte kan anses vara värt allt arbete det skulle innebära.

DELFIN (Direkt Elektronisk Lagring av ForskningsINformation) -projektissa⁶ rakennettiin menestyksellisesti korkeakoulun tutkimusaineistolle tallennus-, tiedonhaku- ja verkkojaketujärjestelmä.

Tutkimusraporttien kokotekstin tallennus tapahtuu useassa eri formaatissa (FrameMaker, Word, LaTeX, HTML), mutta kirjoittajan on toimitettava myös PS-versio PDF-muunnosta varten. Tietokanta- ja käyttöliittymäratkaisusta projektipäällikkönä toiminut kirjastonhoitaja Peter Linde kirjoittaa:

Eftersom vår starka sida inte var själva tekniken tittade vi på programvara från stora och pålitliga leverantörer som kunde erbjuda färdiga och beprövade databaslösningar som fungerade mot WWW och som kunde garantera uppgraderingar. Vårt val föll på Lotus Notes från Lotus Inc., ett dotterbolag till IBM.

Arkistointi tarkoittaa paperiversion säilyttämistä. Tietokannan sisältö poltetaan ajoittain myös rompulle. Muita kokemuksia Delfin-projektista ovat olleet epätietoisuus tekijänoikeusasioista sekä se tosiasia, että internetin hakurobotit eivät pysty indeksoimaan dynaamisia tietokantoja eivätkä ne tue vielä Dublin Core -metadataa NWI:ä lukuunottamatta.

Indeksointiongelma ratkaistiin käyttämällä Internotes Webpublisher -ohjelmistoa, joka luo HTML-dokumentin Notes-tietokannasta. Linde on kuitenkin vakuuttunut siitä, että tulevaisuudessa myös tietokantojen indeksointi onnistuu webistä käsin. Siksi kaikkiin dokumentteihin on lisätty header-kenttä, joka sisältää DC-määrittelyksen mukaisen metadatan.

Tällä hetkellä julkaisujärjestelmää ylläpitää yksi puolipäiväinen henkilö. Arvio on, että yhden dokumentin PDF-muunnokseen ja tietokannan päivittämiseen kuluu vajaa tunti.

Nyt kirjasto valmistelee XML-pohjaisen julkaisujärjestelmän käyttöönottoa opinnäytteilleen projektissa The Xml files⁷, joka on käynnissä vuonna 1999. Rakennekuvaus on tehty ja muutamia dokumentteja on muunnettu XML-muotoon. Word-mallipohja on suunnitteilla ja tekeillä on myös on-the-fly -muunnin XML>HTML, joka on käytössä niin kauan kunnes web-selaimet osaavat renderöidä

⁶ <http://www5.hk-r.se/fou/forskinfor.nsf>, ks. linkki "Projekt Delfin"

⁷ <http://www5.hk-r.se/bib/bibdok.nsf/c1e3e3e46e846aaafc12565e70045d81a>

XML:ää ja siihen liittyvää tyylimäärityä. Lotus Notesin ja XML:n yhteispeliä aiotaan tutkia myös.

Karolinska Institutet

Karolinska Institutetissa on meneillään sekä väitöskirjojen johdanto-osan että kokotekstin elektronisen julkaisemisen hanke⁸. Erikoisuutena Karolinskan projektissa on, että väittelijällä on mahdollisuus julkaista myös korjattu versio työstään väitöstilaisuuden jälkeen. Pilottiprojektiin osallistujia oli niukasti. Tällä hetkellä kokoteksti on luettavissa kahdeksasta väitöskirjasta. Kirjasto tekee muunnostyön PDF-muotoon verkkojakelua varten ja lisää linkit artikkelien kokotekstiin, jos ne ovat tilattavissa PubMed-palvelun kautta. Työn voi jättää kirjastolle seuraavissa tekstinkäsittelyformaateissa:

- Microsoft Word (Win/Mac) 98 tai vanhempi
- PageMaker (Win/Mac) 6.5 tai vanhempi
- Word Perfect (Win/Mac)

Linköping

Linköping University Electronic Press:ssa⁹ (E-Press) on - ilmeisesti kokeiluluontoisesti - kaksi artikkeliväitöskirjaa, joiden julkaisemisessa on noudatettu seuraavaa linjaa: abstrakti on HTML-tiedostona (CSS-tyylisivu on myös luotu), artikkelien bibliografiset tiedot ovat kansisivulla ja Summary on PDF-muodossa. PDF:stä annetaan tarkistussumma PGP-autentikoinnin varmentamiseksi. Niistä artikkeleista, joista on kokoteksti luettavissa kampuslisenssin alaisesta lehdestä, on aktiivi linkki kokotekstiin merkinnällä "restricted access". Myös muutamia kuvia, joihin on ilmeisesti saatu erikseen julkaisulupa, on nähtävissä GIF-muodossa, näistä on myös tarkistussummat.

Mitään ohjeita tai suunnitelmia ei ole vielä julkaistu väitöskirjojen elektroniseksi julkaisemiseksi.

E-Press nojaa vahvasti Eric Sandewallin ja hänen laboratorionsa työpanokseen. Mahdollisesti tästä syystä E-Press ei ole laajentunut siinä määrin kuin mitä aluksi ajateltiin. Se ei myöskään ole lähtenyt vielä esim. väitöskirjojen julkaisemiseen, vasta muutamia kokeiluja on tehty. E-Pressin dokumentoituja periaatteita ja metodeja, kuten PGP:n käyttöä, voitaisiin silti hyvinkin hyödyntää muualla.

Lund

Lundin yliopiston kirjaston LUFT-hanke¹⁰ on tällä hetkellä jäissä. Lund University Press ajetaan mahdollisesti alas ja korvataan Lund University Electronic Press - tyyppisellä yksiköllä. NetLab odottaa myös suosituksia BIBSAM:lta siitä, miten elektroniset väitöskirjat olisi syytä julkaista. BIBSAM on aloittanut "Ruotsin eLib":in eli SVEBIB:n suunnittelutyön, mihin kuuluu osana yliopistojen oma julkaisutuotanto.

⁸ <http://diss.kib.ki.se/info/proj/>

⁹ <http://www.liu.se>

¹⁰ <http://www.lub.lu.se/luft/>

Luulaja

Luulajan yliopistossa on ohjeena lähettää opinnäyte kirjastolle PDF- tai PS-muodossa. Kirjasto vastaa PS>PDF -muunnoksesta. Jos tekijä haluaa työnsä myös HTML:nä, hänen on tehtävä muunnos itse. Kokotekstinä on tällä hetkellä väitöskirjoja viitisen kappaletta, samoin lisensiaatintöitä; tiivistelmiä sen sijaan on runsaasti.

Uumaja

Uumajan yliopiston kirjaston projektissa Elektronisk publicering vid Umeå universitet¹¹ päädyttiin ehdottamaan elektronisten dokumenttien tallentamista siinä muodossa kuin ne oli kirjastoon saatu (odottamaan mahdollista konvertointia myöhemmin) ja konvertointia PDF-muotoon verkkojakelua varten. Päätös vahvistui mm. EP96-konferenssissa¹²:

EP 96, som anordnades för sjätte gången, hade utlovat en del intressanta föredragspunkter som tyvärr visade sig inte riktigt motsvara förväntningarna. I likhet med konferensen i Milton Keynes verkar det som om alla mer eller mindre försöker uppfinna hjulet på nytt. Man hade lagt ner mycken möda på att skapa och skraddarsy egna program inom projekten, som sedan inte fick någon finansiering och således inte kom vidare. Vi blev nu helt övertygade om vikten av att utnyttja de kommersiella program som finns eftersom de hela tiden anpassas efter utvecklingen och blir standard, vare sig man vill det eller inte.

Norja

Oslo

Oslon yliopiston Digital Library -pilottiprojektissa yliopiston USIT (Universitetets senter for informasjonsteknologi) ja sen SGML-työryhmä on luonut opinnäytteille julkaisujärjestelmän, jossa kirjoittaja kirjoittaa tekstin Wordin mallipohjaan ja tallentaa dokumentin RTF-muotoon. Tämä muunnetaan automaattisesti kaupallisella Balise-ohjelmistolla tehtyjen muunnosohjelmien avulla XML- ja SGML-muodon kautta edelleen HTML-dokumentiksi. Tästä muodosta kirjoittaja voi itse päätellä, tuliko hän käyttäneeksi Wordin tyyliä oikein. Jos ei, hän korjaa dokumenttia ja käynnistää kierroksen uudestaan. Dokumenttityyppikuvaus on räätälöity versio Normark DTD:stä, joka on kehitetty ISO 12083:sta. Normarkin omistaa Norjan SGML käyttäjäkerho¹³.

Kirjoittaja tulostaa työnsä Wordilla ja lähettää sen eteenpäin hallinnolle ja kirjastolle. Kirjasto luetteloi sekä elektronisen että paperiversion BIBSYS:iin.

Wordin mallipohjassa käytetään mahdollisimman paljon Wordin standardityylejä, mitä järjestelmän kehittäjät pitivät erityisen tärkeänä. Kirjoittajan oma laadunvalvonta HTML-dokumentin avulla todettiin niin ikään tärkeäksi yliopiston monivuotisessa opetusohjelman rakenteistamisprojektissa. Projektissa uskotaan myös, että tekstinkäsittelyohjelman valintaan ei voida paljontaan vaikuttaa. Jokin aidosti rakenteinen editori, joita markkinoilla on, olisi heidän mielestään lisännyt tuntuvasti "information overload":ia.

¹¹ <http://www.ub.umu.se/ubhtml/proj/elpub/svmain.htm>

¹² <http://www.ub.umu.se/ubhtml/proj/elpub/svrapp.htm>

¹³ http://www.fou.uib.no/fou/grey_lit/sgml-finland9/

Projektissa ollaan myös sitä mieltä, että kirjoittajan luovan vapauden ja dokumentinhallinnan kaipaaman ennustettavuuden ja standardoinnin välillä on periaatteellinen dilemma, jota ei voida tyydyttävästi koskaan ratkaista. Siksi kannattaa lähteä kokeilemalla liikkeelle, kirjoittajan ehdoilla¹⁴.

SGML-ryhmän vetämissä projekteissa on todettu, että kaikkien asianosaisten on hyödyttävä projektin aikana, muuten työ ei etene eikä motivaatiota ole. Jo aikaisessa vaiheessa on myös tehtävä selväksi, minkälaista funktionaalisuutta ja palveluja elektronisilta dokumenteilta odotetaan. Paperikopion mukanaolo prosessissa voi vaikeuttaa tätä, koska siihen on totuttu kiinnittämään suurin huomio.

Oslon projektissa on määritelty selvitettävät ongelmat seuraavaan tärkeysjärjestykseen: 1 - miten kirjoittaja oppii käyttämään työvälinettä oikein; 2 - miten järjestää käyttäjätuki ja kuka sen tekee; 3 - mikä tallennusformaatti valitaan ja missä on riittävä tekninen tietämys. Muita tärkeitä seikkoja ovat:

- kuka vastaa laadusta?
- kuka hyväksyy elektronisen version ja minkälaisilla kriteereillä?
- minkälainen ohjeistus annetaan elektronisen version toimittamista varten?
- miten varmistutaan siitä, että paperi- ja elektronien versio ovat yhtenevät?
- pitäisikö hyväksyä vain elektroninen versio?
- miten taataan elektronisen version löytyvyys (PURL, URN, DOI...)?

¹⁴ Ks. Lähteet: Gulbrandsen 1999

Liite 4 Kansainvälinen yhteistyö - esimerkkinä NDLTD

Virginia Tech

Virginian yliopiston vetämässä Electronic Theses and Dissertations (ETD) -projektissa on tehty opinnäytteille oma DTD, ETD-ML¹⁵ ja käyttäjille ohjeet, miten dokumentti muunnetaan sen määrittelemään muotoon. Tällä tavalla on kuitenkin otettu vastaan vain yksi opinnäyte. Toistaiseksi voimavaroja keskitetäänkin PDF-muotoisen materiaalin saantiin. Adoben Distiller- ja Exchange-tuotteet ovat käytettävissä opiskelijoille kaikissa kampuksen mikroluokissa. Opinnäytteet on yliopistossa pakko jättää elektronisessa muodossa (joko PDF tai SGML), joten hanketta ajaa eteenpäin paitsi kokeilu ja tutkimus, myös käytännön pakko.

ETD on osa laajempaa Networked Digital Library of Theses and Dissertations (NDLTD) -ohjelmaa, jonka tavoitena on lisätä kansainvälistä yhteistyötä yliopistojen julkaisusta vastaavien tahojen kanssa sekä tarjota hajautettu opinnäytteiden hakujärjestelmä.

NDLTD on yhteistyössä sekä UMI:n että kahden väitöskirjoja myyvän yrityksen kanssa. Dissertation.com¹⁶ on osa Amazon.com:ia ja se ottaa myyntiin väitöskirjoja tekstinkäsittelyohjelmien omassa formaatissa tai PDF:nä. Tekijänoikeus säilyy kirjoittajalla. Yrityksellä on yhteinen hakupalvelu NDLTD:n kanssa. Diplomica.com¹⁷ eroaa edellisestä siinä, että se haluaa työn myös paperiformaatissa. Jos tilaaja haluaa paperidokumentin, se kopioidaan tästä originaalista. Organisaatio on saksalais-amerikkalainen yhteistyöyritys.

ETD:n palvelimella on ohjeita sekä aloittelijoille että pidemmälle ehtineille kirjoittajille. Edellinen on suunnattu Word- ja WordPerfect -käyttäjille, jälkimmäinen niille, jotka haluavat kokeilla LaTeX:ia ja/tai ETD-ML:ää.

Word/WP-käyttäjille ei ole olemassa tyylisivua, LaTeXin käyttäjille on. Valmiin textiedoston jatkokäsittely PS:ksi dvips:llä ja edelleen PDF:ksi Adoben Distillerillä on automatisoitu. Käyttäjää ohjataan myös käyttämään tex2pdf-muunnosohjelmaa, jos hän haluaa PDF-tiedostoonsa linkkejä.

Yksityiskohtaisia Word/WP-ohjeita annetaan vain siitä miten asetetaan marginaalit ja sivunumerointi¹⁸. Loput ohjeista koskevat dokumentin asiasisältöä, eräiden dokumenttien osien asettelua valmiissa dokumentissa sekä PDF-dokumentin tuottamista.

XML-kehitystyö on meneillään. ETD-ML DTD on muunnettu XML-standardin mukaiseksi, ja XML:ään pohjautuva workflow-järjestelmä on niin ikään tehty. Tyyli-tiedostot julkistetaan sitten, kun W3C on julkaissut suosituksensa XSL:stä¹⁹:

When XSL is finally refined and becomes an W3C Recommendation, the ETD-ML project team plans to release XSL style sheets "etd2htm.xml" and "etd2pdf.xml" (for conversion to HTML and PDF respectively) and one or more flavors of "etd.xml" for use with formatting

¹⁵ <http://etd.vt.edu/etd-ml/index.htm>

¹⁶ <http://www.dissertation.com/>

¹⁷ <http://www.diplomica.com/>

¹⁸ <http://etd.vt.edu/submit/format.htm>

¹⁹ <http://csgrad.cs.vt.edu/~nkipper/etdsb/>

systems that support the XSL styles directly (expected in future versions of popular Web browsers).

Humboldt Universität zu Berlin

Yliopisto on yksi NDLDT-projektin eurooppalaista partnereista. Se on myös formaattikysymyksiä selvittelevä taho Saksan Dissertationen Online -ohjelmassa²⁰. Humboldtissa väittelijä voi luovuttaa väitöskirjansa elektronisessa muodossa, pakollista se ei ole.

Yliopiston atk-keskuksen vetämässä projektissa on sovellettu ETD-ML:ää ja tehty saksalaisille väitöskirjoille sopiva DTD, DiML. Kirjoittajia varten on Word-mallipohja ja runsaasti ohjeita ja treenausmateriaalia sen käyttämiseksi. Word-kursseja järjestetään kuukausittain (2 x 0,5 päivää). Tekeillä ovat mallipohjat myös LaTeX:lle ja WordPerfectille. XML-pohjaisia työvälineitä tutkitaan, samoin muunnosta TeX>SGML²¹. PDF-muotoista dokumenttia ei oteta vastaan.

Projektissa lähdetään siitä, että osa SGML/XML-pohjaiseen julkaisujärjestelmään siirtymisen lisäkustannuksista vyörytetään kirjoittajalle. Tällöin on kuitenkin otettava huomioon se, että jos halutaan laadukasta jälkeä, kirjoittajaa pitää auttaa²²:

Soll der Qualitätsanspruch an Recherche und Archivierung gehalten werden, ist es jedoch notwendig, die Autoren möglichst frühzeitig zu schulen und mit Grundfragen und -problemen des elektronischen Publizierens vertraut zu machen.

Projektissa korostetaan yliopiston eri yksiköiden yhteistoimintaa elektronisen julkaisutoiminnan järjestämiseksi: kirjaston, atk-keskuksen, mediayksikön jne.

²⁰ http://www.educat.hu-berlin.de/diss_online/

²¹ <http://dochost.rz.hu-berlin.de/epdiss/tp3.html>

²² <http://dochost.rz.hu-berlin.de/epdiss/bibtag/bibtag.pdf>