

PDF-tiedostomuodon hyödyntäminen eduskunnassa

Jari Petersen-Jessen



Tietotekniikan koulutusohjelma

<p>Tekijät Jari Petersen-Jessen</p>	<p>Ryhmä</p>
<p>Opinnäytetyön nimi PDF-tiedostomuodon hyödyntäminen eduskunnassa</p>	<p>Sivu- ja liitesivumäärä 46 + 11</p>
<p>Ohjaajat Ari Apilo, John-Erik Wigren</p>	
<p>Tämä selvitysraportti kertoo, miten eduskunnassa käytetään pdf-tiedostomuotoa, minkälaisia ongelmia eduskunnan käyttäjillä on asiakirjojen tallennusmuotojen suhteen ja minkälaisia pdf-tiedostoja on tallennettu eduskunnan valtiopäiväasiakirjojen asiakirja-arkistoon sekä olisiko pdf-tiedostojen tehokkaammasta käytöstä hyötyä asiakirjojen tallentamisessa ja jakelussa. Eduskunnan sähköiseen asiakirja-arkistoon on tallennettu lähes 75 000 asiakirjaa, joista lähes 40 000 on pdf-muodossa.</p> <p>Pdf-tiedostojen käyttöä selvitettiin keskustelun muodossa tehdyn kyselyn avulla. Kysely kohdistettiin 23 eduskunnan virkamieheen, joilta kysyttiin heidän osallistumisestaan valtiopäiväasiakirjojen ja julkaisujen tekoon. Lisäksi selvitettiin, kuinka paljon he tietävät pdf-tiedostojen käyttömahdollisuuksista ja kuinka paljon he käyttävät pdf-tiedostoja omassa jokapäiväisessä työssään. Eduskunnan asiakirja-arkiston pdf-tiedostojen määrää ja ominaisuuksia tutkittiin käymällä läpi asiakirja-arkiston sisältöä. Tämä tehtiin käyttämällä eduskunnan Internet-sivustossa olevaa valtiopäiväasiakirjojen hakutoimintoa.</p> <p>Työn tavoitteena oli selvittää pdf-tiedostojen käyttötarvetta ja tunnettuutta. Toisena tavoitteena oli saada selville, miksi jotkut asiakirja-arkistossa olevat pdf-tiedostot tulostuvat väärin suhteessa niiden alkuperäiseen tulostusasuun, ja kuinka paljon tällaisia asiakirjoja arkistossa on. Lisäksi selvitettiin pdf-tiedoston ominaisuuksia, käyttömahdollisuuksia ja tulevaisuuden näkymiä. Tulevaisuuden näkymissä kiinnostuksen kohteena oli ensisijaisesti pdf-tiedostojen standardoinnin tila ja näistä ennen kaikkea PDF/A-standardi.</p> <p>Selvitys, joka toteutettiin kesäkuun 2008 ja tammikuu 2009 välisenä aikana, rajattiin koskemaan vain eduskunnassa nykyisin käytössä olevia ohjelmistoja ja tiedostomuotoja. Pdf-formaatista selvitettiin lisäksi PDF/A-formaatin keskeisimmät ominaisuudet ja käyttömahdollisuudet.</p> <p>Selvitystyön tuloksena löydettiin syyt pdf-tiedostojen virheelliseen tulostumiseen ja syyt kyseisten tiedostojen syntyyn. Selvitystyön myötä selvisi myös, että eduskunnan käyttäjien pdf-tiedostojen ominaisuuksien ja käytön tuntemisessa on selvästi parantamisen varaa. Tämän perusteella esitetään, että pdf-tiedostojen tuottamistapoja tehostetaan kehittämällä palvelinperustainen pdf-tiedostojen tuottamisympäristö, johon voitaisiin liittää myös kuvamuotoisten pdf-tiedostojen tekstin tunnistamispalvelu.</p>	
<p>Asiasanat PDF, tiedostomuodot, sähköinen arkistointi, pitkäaikaissäilytys, tietoturva</p>	

Education program for Information technology

<p>Authors Jari Petersen-Jessen</p>	<p>Group</p>
<p>The title of thesis Using the PDF format in Parliament of Finland</p>	<p>Number of pages and appendices 46 + 11</p>
<p>Supervisors Ari Apilo, John-Erik Wigren</p>	
<p>This report describes how the Pdf file format is used by the Finnish Parliament, what kind of problems the user in the Parliament have with the document formats, and what kind of Pdf files there are saved in the archive of the parliament documents. The report will also describe whether the more efficiency use of Pdf files would offer some benefits to save and distribute documents. The electronic document archive for the parliament documents consists of 75 000 documents and almost 40 000 of them are in Pdf format.</p> <p>The usage of the Pdf files was investigated by interviewing the personnel of the Finnish Parliament. There were 23 civil servants who were asked about their participation in producing parliamentary documents or publications. In addition to that they were asked what they know about Pdf files and how much they are using Pdf files in their everyday working. The number of the Pdf files in the electronic document archive was found out by using the search function in the the Web Site of the Finnish Parliament.</p> <p>The main task was to find out the need of the Pdf files and how well known they are. Another task was to find out why some Pdf files in the archive did not have the same printout format than the original document does and how many documents like that there are. In addition to that the features, the usability, and the future visions of Pdf files were investigated. The main target of interest was the state of the standardisation of Pdf files and first of all the standard of PDF/A.</p> <p>The study that took place between June 2008 and January 2009 was restricted to cover only the software and file formats currently used by the Parliament. PDF/A among the other Pdf formats was investigated by its main features and usefulness.</p> <p>The cause of misprinting and the reasons for causing this kind of Pdf files were one of the results of this study. In this study it came obvious that the knowledge of the Pdf files and their usability could be much better. Because of these results it is suggested to establish a server environment to improve methods to produce pdf files. This environment should consist also a OCR system in order to recognize characters in bitmap-type Pdf file and reproduce a searchable Pdf file.</p>	
<p>Key words PDF, file-formats, electronic archiving, long-term preservation, data security</p>	

Sisällys

1	Johdanto	3
1.1	Raportin rakenne	3
1.2	Pdf-formaatin käyttö eduskunnassa: tutkimuksen kohde.....	3
1.3	Tiedostona lähetetyn asiakirjan ulkoasun säilyminen — kaksi esimerkkiä sähköisestä asiakirjasta	4
1.4	Tehtävän rajaus	5
1.5	Projektissa käytetyt menetelmät	6
1.6	Tutkimusongelmat (asiakirja-arkiston tiedostojen käsittely)	6
1.7	Keskeiset käsitteet	7
1.7.1	Asiakirja-arkiston valtiopäiväasiakirjat.....	7
1.7.2	Sanasto	9
2	Pdf-selvityksen lähtökohdat	11
2.1	Eduskunnan tietohallinto	11
2.2	Eduskunnan työasemaympäristö	13
2.2.1	Pdf-tiedostojen käsittely eduskunnan työasemilla	13
2.2.2	Graafiset työasemat ja niiden pdf-tuotanto	13
2.3	Eduskunnan asiakirja-arkisto	14
2.3.1	SGML:n käyttöönotto	15
2.3.2	Adobe FrameMakerin käyttö.....	16
2.3.3	Rakenteiset asiakirjat	16
2.3.4	Ei-rakenteiset asiakirjat	17
2.4	PDF (Portable Document Format) yleisesti.....	17
2.4.1	Pdf:n alkuvaiheet ja kehitys.....	17
2.4.2	Pdf-yhteensopivuustasot	19
2.4.3	Pdf:n keskeisiä ominaisuuksista.....	21
2.4.4	Pdf-standardoinnin tila	23
2.5	PDF/A — arkistoinnin standardi.....	24
2.5.1	Digitaalisen arkistoinnin ongelmat.....	24
2.5.2	Miksi PDF/A eikä pelkkä pdf?.....	25
2.5.3	PDF/A-standardi	26
2.5.4	Kuinka luoda PDF/A-tiedostoja?.....	26
2.5.5	PDF/A:n välittömät hyödyt	28

2.5.6	PDF/A-1a ja PDF/A-1b	29
3	Kyselyitten ja pdf-selvitysten toteutuminen.....	31
3.1	Menetelmien esittely.....	31
3.1.1	Käyttäjäkyselyiden toteutus.....	31
3.1.2	Asiakirja-arkiston tutkiminen.....	33
3.1.3	Pdf ja PDF/A selvitysmenetelmät.....	36
3.2	Käyttäjäkyselyn kautta selvinnyttä.....	36
3.2.1	Julkaisujen tuottaminen	37
3.2.2	Valtiopäiväasiakirjojen tuottaminen.....	37
3.3	Asiakirja-arkiston aineistosta tehdyt havainnot.....	38
3.3.1	Asiakirja-arkiston asiakirjojen kokonaismäärä.....	38
3.3.2	Pdf-tiedostojen määrät.....	39
3.4	PDF/A:n mahdollisuudet ja tulevaisuus.....	40
4	Mitä pdf-asioista saatiin selville ja ehdotuksia jatkosta	41
4.1	Selvityksen keskeiset tulokset ja päätelmät	41
4.1.1	Tietokannan pdf-tiedostojen hakemisen ongelma.....	41
4.1.2	Tutkimusmenetelmien kehittäminen jatkotutkimuksia varten.....	42
4.1.3	Tulosten yleistettävyyys	43
4.2	Toimenpidesuositukset ja jatkotutkimusaiheet	43
4.2.1	Tallenna tekstiarkistoon -makron muuttaminen.....	43
4.2.2	Pdf-tietouden lisääminen.....	43
4.2.3	PDF/A-version jatkoselvitys ja kokeilun käynnistäminen	45
4.2.4	Pdf-tiedostojen tuottamisen automatisointi	45
	Lähteet	46
	Liitteet	
	Liite 1. Kyselylomake	47
	Liite 2. Kyselyaikataulu	49
	Liite 3. Kyselylomakkeiden yhteenveto.....	51
	Liite 4. Valtiopäiväasiakirjojen lukumäärät	57

1 Johdanto

1.1 Raportin rakenne

Raportin rakenne on tiivistettynä ja vapaasti ilmaistuna seuraavanlainen.

- Johdanto

Näin aluksi kuvataan tiivistettynä miten eduskunnassa käytetään pdf-formaattia sekä kerrotaan parin esimerkin avulla tiedostoihin liittyvistä ongelmista. Johdannossa todetaan myös raportin rajaukset, sen tekoon käytetyt menetelmät ja siihen liittyneet tutkimusongelmat. Lisäksi johdannon loppuun on sijoitettu kohta keskeiset käsitteet, josta selviää valtiopäiväasiakirjoihin liittyvät termit sekä raportin sanasto.

- Pdf-selvityksen lähtökohdat

Tässä kerrotaan eduskunnan tietohallinnon organisoinnista, eduskunnan työasemaympäristöistä ja ohjelmistoista sekä eduskunnan sähköisestä asiakirja-arkistosta ja sen tiedostojen tuottamisesta. Tähän on myös kerätty pdf-formaattiin ja PDF/A-standardiin liittyvät kartoitukset.

- Kyselyitten ja pdf-selvitysten toteutuminen

Toteutumisesta kertovassa kohdassa kerrotaan mitä menetelmiä eri asioiden selvityksissä käytettiin ja mistä tietoa haettiin. Tässä kuvataan myös käyttäjäkyselyiden kulkua sekä minkälaisia asioita saatiin selville. Asiakirja-arkiston sisältöön liittyvät asiat löytyvät myös tästä kohdasta.

- Mitä pdf-asioista saatiin selville ja ehdotuksia jatkosta

Tähän koottiin selvityksen keskeiset tulokset ja päätelmät, miten tutkimusmenetelmiä pitäisi mahdollisia jatkotutkimuksia varten kehittää. Lopuksi esitetään jatkotoimiksi joitakin ehdotuksia, joiden avulla voitaisiin parantaa tekstiarkistoon tallennettujen pdf-asiakirjojen tarkkuutta, yleistä pdf-tietoutta ja sen myötä tietoturvaa.

1.2 Pdf-formaatin käyttö eduskunnassa: tutkimuksen kohde

Tässä opinnäytetyössä selvitetään, a) miten eduskunnassa käytetään ja hyödynnetään pdf-formaattia, b) minkälaisia ongelmia eduskunnan käyttäjillä on asiakirjojen tallennusmuotojen suhteen ja c) olisiko pdf-formaatin tehokkaammasta käytöstä hyötyä asiakirjojen tallentamisessa ja jakelussa.

Eduskunnassa tuotetaan runsaasti erilaisia asiakirjoja, kuten muistioita, pöytäkirjoja, esityslistoja, kirjeitä ja niin edelleen. Valtaosa näistä toimitetaan vastaanottajille tiedonantotarkoituksessa, hyvin usein jopa paperitulosteena¹, eikä heillä ei ole yleensä enää tarvetta tai edes lupaa asiakirjan muokkaamiseen. Näin ollen ei ole mitään tarvetta lähettää asiakirjaa alkuperäistiedostona. Riittäisi, että asiakirja lähetetään alkuperäistä paperitulostetta vastaavassa muodossa pdf-tiedostona.

Osa eduskunnassa tuotettavista asiakirjoista syntyy puolestaan tarkoin ohjeistetun tuottamisprosessin mukaan. Näin tuotetaan suurin osa valtiopäiväasiakirjoista, jotka tallennetaan asiakirja-arkistoon ja jotka kootaan yhteen valtiopäiväasiakirjojen vuosittaiseksi arkistokirjasarjaksi.

Eduskunnan sähköiseen asiakirja-arkistoon on tallennettu lähes 75 000 asiakirjaa. Vuosittain valtiopäiväasiakirjojen määrä lisääntyy noin 4 000–5 000 asiakirjalla, joiden yhteinen laajuus on noin 55 000 sivua. Varhaisimmat asiakirja-arkistoon tallennetut asiakirjat ovat vuodelta 1991. Tätä aikaisemmista asiakirjoista ei ole sähköisessä muodossa olevia versioita viitetietoja lukuun ottamatta. Joitakin poikkeustapauksia on digitoitu myös varhaisemmasta aineistosta.

Alun perin asiakirja-arkistoon tallennettiin vain asiakirjojen tekstisisältö. Seuraavassa vaiheessa asiakirjasta tallennettiin myös asiakirjan tekstinkäsittelyohjelman mukainen tiedosto. Vuodesta 2001 lähtien suurimmasta osasta asiakirjoista on tallennettu myös painoasun mukainen pdf-tiedosto.

1.3 Tiedostona lähetetyn asiakirjan ulkoasun säilyminen — kaksi esimerkkiä sähköisestä asiakirjasta

Sain keväällä 2008 kutsun erään yksikön tilaisuuteen. Sähköpostilla toimitettu kutsukortti oli erittäin viehättävä, ja sen ulkoasun eteen oli selvästi nähty paljon vaivaa. Tekstin merkkilaji olisi voinut tosin olla näyttävämpi ottaen huomioon tilaisuuden luonteen. Tekstiä varten oli kyllä valittu erikoismerkkilaji, mutta tämä kävi ilmi vasta itse tilaisuudessa, kun näin kutsun alkuperäisen version painettuna.

Kesällä 2008 tarvittiin eduskunnan asiakirjatoimistossa erästä *hallituksen esitystä*² siinä olleen ulkoasuun liittyvän yksityiskohdan tarkistamista varten. Kyseinen hallituksen esitys löytyi asiakirja-arkistosta helposti, mutta sen ulkoasu ei vastannutkaan alkuperäistä tulostusasiaa. Rivit ja

¹ Eduskunnassa arvioidaan vuoden 2009 tulostus- ja monistuspaperin kulutukseksi 25500 riisiä.

² Hallituksen esitys eduskunnalle -tyyppinen valtiopäiväasiakirja

sanat olivat pääsääntöisesti kohdallaan, mutta etenkin tavutettujen rivien loppuissa kirjaimet menivät usein päällekkäin, joissakin paikoissa niin pahoin, ettei sanasta saanut heti selvää.

Edellä olevat esimerkit tuntuvat olevan melko kaukana toisistaan, ja sitähän ne tietyssä mielessä ovatkin, tilaisuuden kutsukortti ja hallituksen esitys. Yhteistä niille on kuitenkin laatijansa pyrkimys säilyttää alkuperäinen ulkoasu ja uskomus, että tehdyillä toimenpiteillä tässä myös onnistutaan – onhan kaikilla Word tai kyllähän pdf säilyttää ulkoasun.

Kutsukortti oli tehty Microsoft Word -tekstinkäsittelyohjelmalla ja lähetetty suoraan .doc-muotoisena liitetiedostona sähköpostin liitteenä. Kortin tekstiin oli käytetty uuden kuvankäsittelyohjelman mukana tullutta merkkilajia, joka sopi kirjasinleikkaukseltaan erinomaisesti kutsukorttiin. Minulla ei kuitenkaan ollut työasemassani kyseistä merkkilajia, joten Wordini korvasi merkkilajin ”lähinnä sopivalla”.

Wordissa on toiminto, joka automaattisesti korvaa avattavan asiakirjan ne merkkilajit, joita työasemaan ei ole asennettu. Korvaaminen tapahtuu ohjelmiston sisäisen säännösten avulla ja periaatteena on korvata puuttuva merkkilaji mahdollisimman samannäköisellä merkkilajilla. Tässä tapauksessa kortin alkuperäinen merkkilaji oli muuttunut Times New Romaniksi, mikä oli vain kalpea varjo tarkoitetusta, eikä Word edes varoita tai muuten ilmaise merkkilajin korvaamisesta.

Hallituksen esitys eduskunnalle oli tallennettu eduskunnan asiakirja-arkistoon sekä pdf-, html-että word-muodoissa. Näin siksi, että asiakirja olisi tallessa alkuperäisen tekstinkäsittelyohjelman tiedostomuodon (doc) lisäksi myös alkuperäisessä tulostusasussaan (pdf) sekä Internetin hakukoneille helposti tavoitettavassa tekstimuodossa (html). Pdf-muodossa oli kuitenkin ongelmana sen muodostustapa. Wordista PDFWriter-ajurin avulla tulostettuun pdf-tiedostoon ei ollutkaan tallentunut kaikkea tarvittavaa tietoa asiakirjan merkkilajeista, joten jotkut merkit näkyvät ja tulostuvat väärin.

Näiden esimerkkien mukaisiin ongelmiin on kuitenkin löydettävissä hyvin yksinkertaisia ratkaisuja. Tässä selvitysraportissa kuvataan myös tämänkaltaisten ongelmien perusratkaisut.

1.4 Tehtävän rajaus

Tässä selvitysraportissa keskitytään eduskunnassa nykyisin käytössä oleviin ohjelmistoihin ja tiedostomuotoihin. Pdf-formaatista selvitetään lisäksi uuden, erityisesti arkistointiin tarkoitetun PDF/A-version ominaisuudet.

Raportissa ei ole tarkoitus antaa yksityiskohtaisia käyttöohjeita virheettömien pdf-tiedostojen luomiseksi kaikissa eri tilanteissa, vaan tuoda esille pdf-tiedostojen ominaisuuksia ja mitä hyötyä niistä voisi olla normaalissa toimistotyöskentelyssä. Raportissa ei ole myöskään tarkoitus antaa ohjeita olemassa olevien työprosessien muuttamiseksi, mutta luonnollisesti, jos prosessin joissakin vaiheissa havaitaan ongelmia, tuodaan ongelmakohtat esille.

1.5 Projektissa käytetyt menetelmät

Tässä selvitysprojektissa keskityttiin kahteen eduskunnassa tärkeään asiakirjojen tuottamisalueeseen. Projektissa selvitettiin päivittäisten toimistoasiakirjojen tekemistä ja käyttöä sekä valtiopäiväasiakirjojen arkistointia, johon liittyen tutkittiin myös pdf-tiedostojen ominaisuuksia.

1) Käyttäjäkyselyiden avulla selvitettiin, miten eduskunnan virkamiehet hyödyntävät pdf-tiedostoja jokapäiväisessä työssään ja millä ohjelmilla he tekevät niitä. Lisäksi kysyttiin, miten hyvin he tuntevat pdf-tiedostojen eri käyttömahdollisuuksia ja minkälaisia tarpeita tai ongelmia niihin liittyy.

2) Eduskunnan sähköistä asiakirja-arkistoa tutkittiin eduskunnan Internet-sivuilla olevan *Valtiopäiväasiat ja -asiakirjat* -sivuston avulla. Sivustolla olevan asiakirjahaun kautta selvitettiin, minkälaisia pdf-asiakirjoja eduskunnan asiakirja-arkistoon on tallennettu, mitä pdf-versioita niissä on käytetty ja millä välineillä niitä on tuotettu.

3) Pdf-tiedostojen eri versioiden ominaisuuksia ja mahdollisuuksia selvitettiin tutkimalla pdf-tiedostoihin liittyvää kirjallisuutta ja Internet-sivustoja. Lisäksi perehdyttiin keväällä 2008 Amsterdamissa järjestetyn kansainvälisen PDF/A-konferenssin aineistoon.

1.6 Tutkimusongelmat (asiakirja-arkiston tiedostojen käsittely)

Kuten todettu, asiakirja-arkistossa olevien pdf-tiedostojen olemusta selvitettiin eduskunnan Internet-sivuston avulla. Selvittämistä kuitenkin hankaloitti, että arkistossa olevien pdf-tiedostojen hakuun ei ollut tarjolla mitään muuta välinettä kuin arkiston yleinen html-pohjainen käyttöliittymä.

Jos tiedostojen saantiin olisi ollut käytettävissä normaalin tiedostojärjestelmän resurssienhallinta-ohjelmaan verrattavissa oleva menetelmä, olisi tiedostojen käsittely ollut huomattavasti helpompaa ja nopeampaa, mikä olisi mahdollistanut myös tiedoston tarkemman tutkimisen. Näin olisi ehkä saatu täsmällisempää tietoa pdf-tiedostojen versioista.

1.7 Keskeiset käsitteet

1.7.1 Asiakirja-arkiston valtiopäiväasiakirjat

Eduskunnan asiakirja-arkistoissa on valtiopäiväasiakirjoja vuodesta 1991 lähtien. Verkkopalvelu sisältää sekä eduskunnan että sen sidosryhmien tuottamia asiakirjoja. Eduskunta huolehtii kaikkien asiakirjojen arkistoinnista. Arkistot sisältävät teknisesti kahdenlaisia asiakirjoja:

Ei-rakenteiset asiakirjat

Vanhimmat ei-rakenteiset asiakirja-arkistot sisältävät asiakirjat vain tekstimuodossa (ascii-muodossa). Vuodesta 2001 lähtien asiakirjat on tallennettu myös pdf-muodossa ja vuodesta 2003 lähtien Word-muodossa. Ei-rakenteisia asiakirjoja voi hakea perustunnisteiden tai tekstihaun keinoin.

Rakenteiset asiakirjat

Eduskunta laatii valtiopäiväasiakirjat nykyisin rakenteisessa muodossa. Rakenteinen asiakirjastandardi otettiin käyttöön vuosina 1998-2000. Standardin käyttöönotto on parantanut verkkopalvelujen hakuominaisuuksia ja asiakirjan esitystapaa. Rakenteiset asiakirjat on saatavissa HTML- ja PDF-muodoissa.

Seuraavassa taulukossa on yksityiskohtaisia tietoja asiakirja-arkistoista ja asiakirjojen tuottajista. Lisäksi taulukosta selviää, mistä vuodesta lähtien tiettyjen tallennusmuotojen käyttö on aloitettu.

Taulukko 1. Eduskunnan asiakirja-arkiston valtiopäiväasiakirjojen kuvaukset, tuottajat ja aloitusvuosi (Valtiopäiväasiakirjat eduskunnan asiakirja-arkistossa).

Valtiopäiväasiakirja	Kuvaus	Tuottaja	Vuodesta
Hallituksen esitykset - HE	ei-rakenteisia asiakirjoja, tallennusmuodot: ascii, pdf (2001-), Word (2003-)	valtioneuvosto	1991-
Valtioneuvoston selonteot ja tiedonannot - VNS, VNT	ei-rakenteisia asiakirjoja, tallennusmuodot: ascii, pdf (2001-), Word (2003-), osasta asiakirjoja vain pdf-muoto	valtioneuvosto	VNS: 1995- VNT: 1996-
Valtioneuvoston kirjelmät ja selvitykset (EU) - E, U, UTP	ei-rakenteisia asiakirjoja, tallennusmuodot: ascii, pdf (2001-), Word (2003-)	valtioneuvosto, ministeriöt	U ja E: 1995- UTP: 2003-
Valtioneuvoston kirjelmien ja selvitysten jatkokirjelmät (EU) - E, U ja UTP	ei-rakenteisia asiakirjoja, tallennusmuodot: ascii, pdf (2001-), Word (2003-)	valtioneuvosto, ministeriöt	2001-

Valtiopäiväasiakirja	Kuvaus	Tuottaja	Vuodesta
pääministerin ilmoitukset	rakenteisia asiakirjoja, tallennusmuodot: sgml ja pdf	valtioneuvosto; eduskunta liittää osaksi pöytäkirjaa	1999-
Valtioneuvoston kirjelmät (ministerin sidonnaisuudet) - VN	ei-rakenteisia asiakirjoja, tallennusmuodot: ascii, pdf (2001-), Word (2003-)	valtioneuvosto	1996-
Valiokuntien mietinnöt ja lausunnot	sisältää valiokuntien mietintöjä ja lausuntoja koostuu kahdesta erillisestä palvelusta: ei-rakenteiset asiakirjat 1991- 31.8.1998, tallennusmuoto: ascii rakenteiset asiakirjat 1.9.1998-, tallennusmuodot: sgml ja pdf	eduskunta	1991-
Eduskunnan vastaukset ja kirjelmät - EV, EK	sisältää eduskunnan vastauksia ja kirjelmiä koostuu kahdesta erillisestä palvelusta: ei-rakenteiset asiakirjat 1991- 1999/2000, tallennusmuoto: ascii rakenteiset asiakirjat 1999-/2000-, tallennusmuodot: sgml ja pdf)	eduskunta	1991-
Aloitteet - KA, LA, TAA, LTA, RA, TA, TPA	sisältää keskustelu-, laki-, talousarvio-, lisätalousarvio-, raha-asia-, toivomus- ja toimenpideoitteita koostuu kahdesta erillisestä palvelusta: ei-rakenteiset asiakirjat 1991–1998/1999, tallennusmuoto: ascii rakenteiset asiakirjat 1999-/2000-, tallennusmuodot: sgml ja pdf lakialoitteet rakenteisina asiakirjoina 1.2.2000-, toimenpideoitteet 1.3.2000- ja muut aloitteet 25.3.1999 - raha-asia-aloitteen tilalle on tullut talousarvioaloite vuonna 1992, toivomusaloitteen tilalle toimenpideoite vuonna 2000	eduskunta	1991-
Kysymykset - KK, KVN, VK	sisältää kirjallisia kysymyksiä, kysymyksiä valtioneuvostolle ja välikysymyksiä koostuu kahdesta erillisestä palvelusta: ei-rakenteiset asiakirjat 1991–1998, tallennusmuoto: ascii rakenteiset asiakirjat 1999-, tallennusmuodot: sgml ja pdf kysymyksiä valtioneuvostolle on laadittu vuoteen 1998 asti	eduskunta	1991-
Puhemiesneuvoston ehdotukset ja luettelot - PNE, PNL	sisältää puhemiesneuvoston ehdotuksia ja luetteloita koostuu kahdesta erillisestä palvelusta: ei-rakenteiset asiakirjat 1994–1999, tallennusmuoto: ascii rakenteiset asiakirjat 2000-, tallennusmuodot: sgml ja pdf	eduskunta	1999-
Kertomukset - K	ei-rakenteisia asiakirjoja, tallennusmuodot: ascii, pdf (2001-), Word (2003-), osasta asiakirjoja vain pdf-muoto	valtioneuvosto, eräät toimielimet	1996-

Valtiopäiväasiakirja	Kuvaus	Tuottaja	Vuodesta
Muut asiakirjat - MINS, M	ei-rakenteisia asiakirjoja, tallennusmuodot: ascii, pdf (2001-), Word (2003-), osasta asiakirjoja vain pdf-muoto	eduskunta, valtioneuvosto	1992-
Talousarviomietinnöt ja -kirjelmät pääluokittain ja osastoittain	rakenteisia asiakirjoja, tallennusmuodot: sgml ja pdf	eduskunta	1998-

1.7.2 Sanasto

Adobe Acrobat	Adoben 1993 julkistama pdf-tiedostojen tuottamisohjelma.
Adobe FrameMaker	Adoben kehittämä kirjoitus- ja julkaisuohjelmisto. Eduskunnassa on käytössä versio FrameMaker+SGML v5.5.
ASCII	American Standard Code for Information Interchange. Tietokoneiden nykyisin käyttämä merkistö, joka sisältää englannin kielen kirjaimet, numerot, joukon väli- ja ohjausmerkkejä. Ns. perus-Ascii sisältää 128 merkkiä. Eri merkistöt ovat yleensä keskenään yhteensopivia näiden 128 ensimmäisen merkin osalta.
CMYK	Lyhenne neliväripainatuksen osaväreistä cyan, magenta, keltainen ja musta. Värijärjestelmä, jota käytetään painoteollisuudessa. vrt. RGB
DTP	Desktop Publishing, sivunvalmistus tietokoneella
Eirak	Ei-rakenteinen tekstiarkisto, johon on tallennettuhallinnon ja sidosorganisaatioiden tuottamia lainsäädäntötyön ja hallinnollisten päätösten asiakirjoja sekä vanhaa lainsäädäntötyön aineistoa.
Fakta	Eduskunnan sisäinen tiedonjakelujärjestelmä, intranet.
HTML	Hypertext Markup Language, kuvauskieli, jota käytetään muun muassa web-sivujen teossa.
ISO	International Organization for Standardization, 1947 perustettu kansainvälinen standardointiorganisaatio.
JPEG	Join Photographic Experts Group, kuvien pakkausmuoto. Käytetään myös lyhennettä JPG, Windowsin tiedostopäätteen mukaisesti.
PDF	Portable Document Format, Adoben Systemsin kehittämä laite- ja ohjelmistoriippumaton tiedostomuoto.
PostScript	Vektorigrafiikkaan perustuva sivunkuvauskieli.
PTJ	Valtioneuvoston päätöksenteon tukijärjestelmä, jonka avulla valmistellaan valtioneuvoston istuntojen esittelylistat ja niiden liitteet.
RGB	Lyhenne kolmivärijärjestelmän väreistä punainen (red), vihreä (green) ja sininen (blue). vrt. CMYK

SGML	Standard Generalized Markup Language, 1986 standardoitu (ISO 8879:1986) metakieli, jolla voidaan määritellä muita merkintäkieliä, kuten HTML ja XML
Sivunkuvauskieli	Tietokonekieli, jonka avulla voidaan sijoitella monentyyppistä sivulle tulevaa aineistoa.
Sivuntaitto	Valmiin tekstin, kuvien ja muiden elementtien sijoittelu tulostusta tai painamista varten.
Skannaaminen	Kuvan tai asiakirjan sivun lukeminen kuvanlukijalla kuvatiedostoksi. Joillakin kuvanlukijoilla voidaan kuvatiedosto tallentaa edelleen pdf-tiedostoksi.
TRIP	Eduskunnan asiakirja-arkiston toteutukseen käytetty tekstitietokanta.
Tulostus tiedostoon	Sivunkuvauskielillä valmistettujen sivujen muuttaminen tulostuskelpoiseen muotoon. Nykyiset pdf-ohjelmat osaavat muuntaa lähdetiedoston tulostusmuotoon generoimalla automaattisesti tarvittavan sivunkuvauskielisen välitiedoston.
U-asiat, U-asiakirja	U-asiat ovat ehdotuksia sellaisiksi EU:n säädöksiksi, jotka koskevat eduskunnan toimivallan piiriin kuuluvia kysymyksiä. U-asiat tulevat eduskunnassa vireille valtioneuvoston eduskunnan puhe- miehelle lähettämällä kirjelmällä. U-asiakirjat ovat tähän asiaan liittyviä asiakirjoja.
Unicode	merkistöstandardi, joka kattaa suurimman osan maailman kirjoitet- tujen kielten käyttämistä merkeistä.
WP, WP5	WordPerfect Corporationin käyttämiä WordPerfect-ohjelmien tiedostomuotoja.
WPSPLUS	Digital Equipment Corporationin Allin1-järjestelmän tiedosto- muoto.
XHTML	eXtensible Hypertext Markup Language, XML-kielen muotovaa- timukset täyttävä HTML-kieli, jonka tarkoituksena oli alun perin korvata HTML-kieli.
XML	eXtensible Markup Language, SGML-kieleen pohjautuva metakie- li, jota käytetään tiedon kuvaamiseen.

2 Pdf-selvityksen lähtökohdat

Tämän selvitysprojektin viitekehyksen muodostavat eduskunnan tietohallinto sekä erityisesti eduskunnan työasemaympäristö ja asiakirja-arkisto. Näiden lisäksi selvitettäviin asioihin kuuluvat pdf-tiedostoformaatti ja sen eri versiot. Seuraavissa kappaleissa kuvataan tätä kokonaisuutta eri näkökulmista.

2.1 Eduskunnan tietohallinto

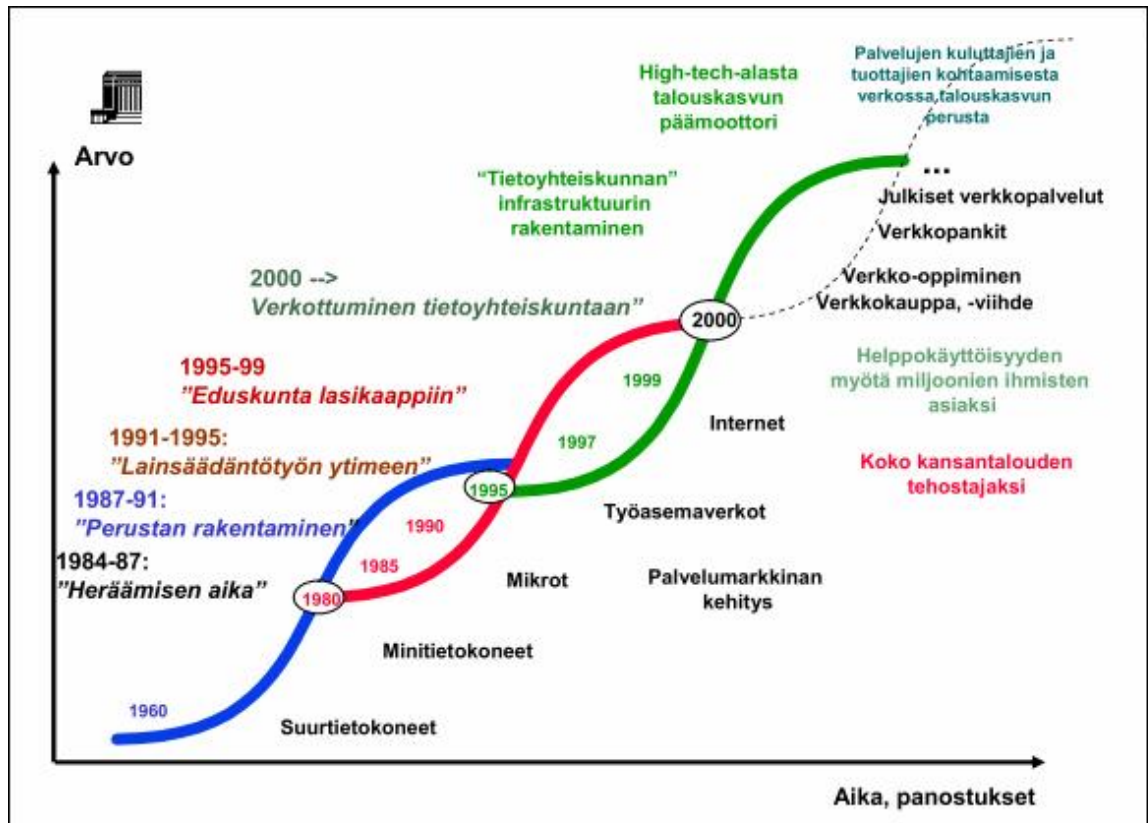
Tietotekniikan hyödyntäminen eduskunnassa juontaa juurensa vuoteen 1984, jolloin eduskunnan kehittämistyöryhmä asetti automaatiotyöryhmän, jonka tehtäväksi annettiin kartoittaa ja arvioida eduskunnan asiakirjatuotannon ja tietojenkäsittelyn nykytilaa ja kehittämistarpeita. Automaatiotyöryhmälle annettiin myös tehtäväksi laatia eduskunnalle tietojenkäsittelyn kehityssuunnitelma. Eduskunnan kansliatoimikunta hyväksyi automaatiotyöryhmän asettamisen 29.3.1984. (Eduskunnan automaatiotyöryhmä 1985, esipuhe.)

Yksi automaatiotyöryhmän selvityksen tuloksista oli esitys tietojenkäsittelyn tukipalvelujen organisoinnista. Työryhmä esitti, että tietojenkäsittelyn kehittämistä ja eduskunnan sisäisen koulutuksen organisointia varten perustetaan vähintään kolme uutta tointia. Nämä tukipalvelut, tietotekniikan kehittäminen ja suunnittelu, koulutuksen suunnittelu ja kehittäminen sekä tietojärjestelmien ylläpito- ja tukitoiminta, ehdotettiin liitettäväksi eduskunnan hallinto-osastoon osaksi eduskunnan sisäistä hallintoa. (Eduskunnan automaatiotyöryhmä 1985, 80.)

Näin ollen vuonna 1987 palkatun atk-päällikön ensimmäisiä tehtäviä olivat tietohallintoyksikön perustaminen sekä tietohallinnon eri toimintojen suunnittelu ja organisointi. Aloituskvaiheen tehtävinä oli muun muassa:

- tekstinkäsittelyn tarpeiden yksilöinti ja tarpeiden mukainen toteutus
- elektronisen tekstiarkistoinnin määrittely ja toteutus myöhemmässä kehitysvaiheessa
- valtiopäiväasioiden seurantajärjestelmän kokeiluprojektin käynnistäminen

Näin sai alkunsa eduskunnan kuten monen muunkin organisaation syklinen kehitysmalli, jossa tiettyä järjestelmää käytettäessä kehitetään jo seuraavaa, jotta se olisi valmiina, kun nykyinen järjestelmä on käyttöikänsä päässä (Mustajärvi O, 14.1.2009).



Kuvio 1. Eduskunnan IT-kehityksen vaiheet (Mustajärvi, O 14.1.2009).

Noista ajoista eduskunnan tietohallinto on kasvanut 26 hengen organisaatioksi, joka huolehtii 1000 hengen tietotekniikkaympäristöstä. Vuonna 2008 tietotekniikkalaitteiston muodosti 1200 työasemaa, 600 kannettavaa tietokonetta, yli 60 palvelinta ja lisäksi joukko tulostimia sekä runsaasti erilaisia tietoliikennelaitteita ja kaapeleita.

Eduskunnan tärkeimpiä sovelluksia ovat:

- täysistunnon tietojärjestelmät
- valiokuntien tietojärjestelmä (VKFakta)
- valtiopäiväasioiden käsittelyjärjestelmä (VEPS)
- tekstiarkistot
- intranet-palvelut (Fakta)
- toimistojärjestelmä (tekstinkäsittely, sähköposti, kalenteri ym)
- puhelinjärjestelmät

Kaikki yllä mainitut sovellukset tuottavat tai käyttävät pdf-tiedostoja tai vähintäänkin materiaalia pdf-tiedostoja varten. Puhelinjärjestelmistä uusissa matkapuhelimissa on Adobe Reader LE 1.5 -lukuohjelma, joten myös niillä voidaan lukea pdf-tiedostoja.

2.2 Eduskunnan työasemaympäristö

Eduskunnan työasemien ohjelmistorungon muodostavat Microsoft Windows XP -käyttöjärjestelmä, Internet Explorer -selain, Microsoft Office 2002 tai 2003 -toimisto-ohjelmisto ja TeamWare Office -sähköposti- ja kalenteriohjelmisto. Kannettavissa tietokoneissa on Cisco VPN ja Citrix Metaframe -ohjelmistot, joiden avulla saadaan suojattu etäyhteys eduskunnan tietojärjestelmiin.

Valtiopäiväasiakirjojen tuottamisprosessissa mukana olevien henkilöiden työasemissa on lisäksi Adobe FrameMaker+SGML v5.5 -ohjelmisto, joiden avulla tuotetaan rakenteisia asiakirjoja kokoamalla tekstejä eri lähteistä ja käyttämällä vakiofraasistoja.

2.2.1 Pdf-tiedostojen käsittely eduskunnan työasemilla

Kaikissa eduskunnan työasemissa on Tracker Softwaren PDF-XChange 3.0 -pdf-ohjelmisto, jonka pdf-tulostusohjainta käyttämällä voi luoda pdf-tiedostoja kaikista ohjelmista, joissa on ylipäättään tulostusmahdollisuus. FrameMaker työasemissa on myös Adobe Acrobat 5 -ohjelmisto, jonka Distiller-ohjelmaa käytetään valtiopäiväasiakirjojen pdf-tiedostojen luontiin.

Eduskunnan kopiokoneita on viime vuosina vaihdettu yhä enemmän monitoimilaitteiksi. Monitoimilaitteen kuvanlukijalla voidaan skannata asiakirja tiedostoksi, joka edelleen voidaan tallentaa tiedostoksi kovalevylle tai lähettää sähköpostilla edelleen.

PDF-XChange-ohjelmiston PDF-Tools-ohjelmalla voidaan käsitellä pdf-tiedostoja jakamalla tai yhdistämällä niitä sekä erottamalla niistä sivuja tai muuttamalla niiden ominaisuuksia. Valtiopäiväasiakirjojen pdf-versioiden käsittelyyn käytetään Adobe Acrobat -ohjelmaa, jolloin voidaan käyttää FrameMaker-ohjelman kanssa samoja *.jobsettings*-asetustiedostoja.

Pdf-tiedostojen lukua varten kaikissa työasemissa on Adobe Reader -ohjelma, josta käytössä on pääsääntöisesti versiot 7 ja 8. Adobe Reader, aiemmin Acrobat Reader, on Adoben kehittämä pdf-tiedostojen lukuohjelma, joka on ilmaiseksi ladattavissa Adoben Internet-sivuilta.

2.2.2 Graafiset työasemat ja niiden pdf-tuotanto

Eduskunnassa tuotetaan runsaasti erilaisia kirjapainoon painettaviksi lähetettäviä julkaisuja. Nykyaikainen digitaalinen painaminen mahdollistaa, että julkaisu tuotetaan eduskunnassa val-

miiksi ja toimitetaan painoon pdf-muodossa. Tällöin painamisesta jää pois filmin valmistus, mikä nopeuttaa julkaisun painamista ja laskee kustannuksia. (Helkama 2002, 85).

Eduskunnassa julkaisuja tekevien henkilöiden käytössä on niin sanotut graafiset työasemat, joihin on asennettu Adobe Creative Suite -ohjelmisto. Ohjelmistoon kuuluu taitto-ohjelma InDesign, jolla julkaisut suunnitellaan ja jolla lopuksi tuotetaan painoa varten tarvittava pdf-tiedosto. Pdf-tiedostojen oikeiden asetusten varmistamiseksi kirjapaino on toimittanut tulostimiaan vastaavan asetustiedoston, jolla varmistetaan paras mahdollinen painolaatu.

Julkaisujen ulkoasu tilataan usein ulkopuoliselta graafiselta suunnittelijalta, joka lopulta lähettää taittoa varten tehdyn taittopohjan suoraan InDesign-ohjelman mallitiedostona. Tämän käyttö edellyttää kuitenkin, että eduskunnassa julkaisua tekevällä henkilöllä on sama ohjelmaversio kuin suunnittelijalla.

InDesign-ohjelman taittotöitä voidaan siirtää henkilöltä toiselle edellyttäen, että heillä on sama ohjelmaversio, nimittäin jokainen kerta, kun InDesign-ohjelmasta on tullut uusi versio, on myös sen oma tiedostomuoto muuttunut. Tästä seuraa, että käytännössä kaikilla eduskunnan InDesign-käyttäjillä on oltava sama ohjelmaversio.

InDesignissa on kyllä mahdollista tallentaa taittotyö niin sanotussa yhteensopivassa muodossa, jolloin sen saa auki myös edellisellä InDesign-versiolla. Tämä yhteensopiva muoto toimii kuitenkin vain edeltävään ohjelmaversioon eikä sitä edeltäviin, eikä tuolloinkaan saada välttämättä kaikkia taittotyön ominaisuuksia siirrettyä.

Tällä hetkellä eduskunnassa on käytössä Creative Suiten CS2-versio, mutta keväällä 2009 Creative Suite -ohjelmisto päivitetään versioon Creative Suite Design CS4, joka sisältää uuden version InDesign-ohjelmasta ja Acrobat-ohjelman version 9. InDesign-ohjelmilla tehdyt pdf-tiedostot ovat kyllä yleisten pdf-määrittysten ja standardien mukaisia.

2.3 Eduskunnan asiakirja-arkisto

Vuonna 1984 asetetun automaatiotyöryhmän selvityksen tuloksena esitettiin myös, että eduskuntaan luodaan järjestelmä, joka mahdollistaa sähköiset arkistopalvelut eduskunnan omille yhteiskäyttöisille teksteille. Arkistopalveluiden tulee mahdollistaa myös valtion painatuskeskuksessa säilytettävien tekstien arkistointi. Automaatioryhmän tutkimuksessa esitetään saavutettaviksi hyödyksi seuraavaa:

Tekstiarkistojen muodostaminen tehostaa valtiopäiväasiakirjoihin kohdistuvaa tiedonhakuja sekä helpottaa asiakirjojen laadintaa ja tekstinkäsittelyä. Tekstiarkistot korvaavat ainakin osittain paperiarkistopalveluja ja helpottavat mm. arkistointiin liittyviä tilaongelmia. (Eduskunnan automaatio-työryhmä 1985, 92.)

Näin on heti tietohallinnon käynnistymisen alkuvaiheessa määritelty myös asiakirja-arkiston tarve ja toiminnot. Lisäksi jo tuolloin oli selvää mitä hyötyjä sähköisellä asiakirja-arkistolla on saatavissa.

Eduskunnan valtiopäiväasiakirjojen tekstiarkisto on jaettu kahteen osaan, rakenteisiin ja ei-rakenteisiin asiakirjoihin. Rakenteisessa arkistossa on SGML-muotoisia asiakirjoja, joita voidaan hakea HTML-lomakkeen avulla asiakirjoissa olevien metatietojen perusteella. Lisäksi asiakirjoihin voidaan kohdistaa myös vapaa tekstihaku.

Nyt tutkittavien valtiopäiväasiakirjojen lisäksi eduskunnan arkistoon on tallennettu myös pöytäkirjoja. Täysistuntojen pöytäkirjoja on tallennettu vuodesta 1999 alkaen, ja niitä on kertynyt vuoden 2008 loppuun mennessä 1411 kappaletta. Pöytäkirjat on haettavissa sekä täysistunnoitain että puhujittain. Vuodesta 2002 alkaen on arkistoon tallennettu myös valiokuntien pöytäkirjoja, joita on yhteensä 6907 kappaletta. Valiokuntien pöytäkirjoja voidaan hakea normaalien hakukenttärvojen lisäksi myös valiokunnittain.

2.3.1 SGML:n käyttöönotto

SGML liittyy olennaisesti Internetin käytön alkutaipaleeseen eduskunnassa. Vuonna 1994 eduskunnan tietohallinto selvitti samaan aikaan sekä asiakirjatuotannon ongelmia että eduskunnan intranetin FAKTA-järjestelmän perusteita. Tietohallinnossa päädyttiin lopulta ratkaisuun, jossa asiakirjojen tuottaminen ja tallentaminen toteutettiin SGML-muodossa. (Mustajärvi, O. 14.1.2009)

SGML:n käyttöönotto tuolloin ei ollut mitenkään itsestään selvä asia. Samaan aikaan, kun eduskunta selvitti SGML:n mahdollisuuksia yhdessä Jyväskylän yliopiston Airi Salmisen johtaman työryhmän kanssa, tutkittiin valtioneuvoston kansliassa WordPerfect 5.1:n käyttöä tekstiarkistojen tallennusmuotona. (Kuronen 1998, 64.)

Ensimmäinen valtioneuvoston PTJ (päätöksenteon tukijärjestelmä) pohjautui WordPerfect 5.1 Windows-versioon. WP 5.1 -ohjelmaa tai -asiakirjamuotoa ei enää käytetä arkistointimielessä, eikä juuri muutenkaan. Nykyinen PTJ pohjautuu puolestaan MS Wordin tiedostomuotoon.

2.3.2 Adobe FrameMakerin käyttö

Eduskunnassa on käytetty Adobe FrameMaker+SGML v5.5 + SGML -ohjelmaa syksystä 1998 alkaen valiokuntien lausuntojen ja mietintöjen laatimiseen. SGML-muotoiset asiakirjat on tallennettu asiakirja-arkistoon, josta automaattisesti muodostetaan hakuehtojen mukaiset tuloluettelot HTML-muodossa. (Kuronen 1998, 64).

Myöhemmin FrameMaker-ohjelman käyttöä on lisätty niin, että nykyään lähes kaikki valtiopäiväasiakirjat tuotetaan kyseisellä ohjelmalla. SGML-tiedostomuoto ja FrameMaker-ohjelmaan tehdyt räätälöinnit ovat kuitenkin osaltaan aiheuttanut sen, että FrameMaker-ohjelman uudempiä versioita ei ole voitu ottaa käyttöön, mikä on edelleen johtanut siihen, että pdf-tiedostojen taso on jäänyt Acrobat 5:n ja pdf-versio 1.4 tasolle.

2.3.3 Rakenteiset asiakirjat

Eduskunnan tuottamat valtiopäiväasiakirjat laaditaan Adobe FrameMaker+SGML v5.5 -ohjelmalla. Lainsäädäntötyössä käytetään asiakirjatyypikohtaisia sovelluksia seuraaville asiakirjoille:

- aloite
- kysymys
- valiokunnan mietintö ja lausunto
- talousarviomietintö/talousarviokirjelmä
- valiokunnan esityslista ja kokouspöytäkirja
- eduskunnan vastaus ja kirjelmä.
- lakialoite/puhemiesneuvoston ehdotus tai luettelo
- päiväjärjestys
- täysistunnon pöytäkirja

Kullekin asiakirjatyypille on määritetty tyyppikohtainen rakennemäärittely ja toteutettu oma laadintasovelluksensa. Sovellukset sisältävät yhden tai useamman ulkoasumäärittelyn, asiakirjan laadintaan liittyviä tukitoimintoja, fraasistojen käyttömahdollisuuden ja Tekstiarkisto-toiminnon, jolla käynnistetään asiakirjan arkistointiprosessi. Asiakirjasta tallennetaan verkkolevylle sgml-, ascii- ja pdf-muodot, jotka tallennetaan Trip-tietokantoihin. Toiseen kantaan siirretään ascii- ja sgml-muodot sekä kiinteitä kenttiä ja toiseen kantaan siirretään asiakirjan pdf-muoto.

Rakenteisia asiakirjoja on tuotettu vuodesta 1998 lähtien, ja viimeisimpänä rakenteisiin asiakirjoihin siirrettiin eduskunnan vastaukset ja kirjelmät vuonna 2000. Vanhemmat asiakirjat löytyvät ei-rakenteisista arkistoista.

Rakenteisia asiakirjoja voidaan hakea verkkopalvelun avulla, jolloin käyttäjä valitsee joko peruslomakkeen tai laajennetun hakulomakkeen. Hakulomakkeen avulla käyttäjä voi hakea asiakirjoja erilaisten metatietojen avulla. (Nurmeksela, R. Virtanen, M. Lehtinen, A. Järvenpää, M. & Salminen, A. 2006. 8–9).

2.3.4 Ei-rakenteiset asiakirjat

Eduskunnan ei-rakenteisten asiakirjojen tekstiarkisto sisältää sidosorganisaatioiden tuottamia lainsäädäntöön liittyviä asiakirjoja. Hallituksen esitykset, valtioneuvoston selonteot ja tiedonannot sekä U-asiakirjat haetaan PTJ-järjestelmästä. Ministeriöt lähettävät EU-asiakirjoja sähköpostilla, ja erilaiset valtiopäiväasiakirjoihin luettavat kertomukset tulevat laatijaorganisaatioilta.

Ei-rakenteiset asiakirjat tallennetaan Word-, ascii- ja pdf-muodoissa. Ennen vuotta 2003 laaditut asiakirjat ovat Wordin asemesta WP-muodossa. Verkkopalvelut päivittyvät automaattisesti uusia asiakirjoja lisättäessä tai vanhoja muutettaessa.

Ennen sgml-tiedostojen käyttöönottoa laaditut asiakirjat on myös tallennettu ei-rakenteisiin tekstiarkistoihin. Näissä tallennusmuotona on ascii- ja WPSPPLUS- tai WP5-muoto. Sähköinen arkistointi aloitettiin vuonna 1991 ja vanhemmista asiakirjoista on vain viitetietoja. (Nurmeksela ym. 2006. 10).

2.4 PDF (Portable Document Format) yleisesti

2.4.1 Pdf:n alkuvaiheet ja kehitys

PDF-tiedostoformaatin tekniikka tuli ensimmäisen kerran esille Seyboldin kokouksessa San Josessa 1991. Silloin asiasta puhuttiin termillä IPS.³ PDF versio 1.0 esiteltiin Comdex-messuilla 1992, jolloin pdf-teknologia palkittiin "Best of Comdex" -palkinnolla. Koodinimenä oli ensin Camelot ja sitten Carousel, mutta väline PDF-tiedostojen tekoon ja katseluun, Acrobat 1.0, julkistettiin 15. kesäkuuta 1993. (Prepressure.com.)

³ Interchange PostScript

Tätä ensimmäistä versiota ei voinut käyttää pre-press-toimintaan. Pre-press on painotekniikassa käytetty termi, joka sisältää mm. painoaineiston vastaanottamisen ja arkkiaseinnin (Katajauuri & Loikkanen 1999, 33). Acrobat 1.0:n ominaisuuksiin kuuluivat kuitenkin jo sisäiset linkit ja kirjanmerkit sekä upotetut kirjasimet, mutta ainoa käytettävissä oleva väriavaruus oli RGB, kun painotekniikka taas edellytti CMYK-väriavaruuden hallintaa.

Jotta Adobe ei olisi päässyt yksinään valtaamaan sähköisten asiakirjojen aluetta, ilmestyi nopeasti joukko kilpailevia tekniikoita. Eräs näistä oli Tumbleweed Communications Corporationin kehittämä Envoy, joka tuli markkinoille 1993 ja liitettiin osaksi WordPerfect Office -pakettia 1994. Envoy:n etuna oli mahdollisuus liittää lukuohjelma osaksi jaettavaa asiakirjaa. Tai oikeammin asiakirja liitettiin osaksi lukuohjelmaa, jolle annettiin asiakirjan nimi ja tiedostotunniste .exe – siihen aikaan käyttäjät vielä uskalsivat ja saivat ajaa "tuntemattomia" ohjelmia!

Adoben ajatuksena oli aluksi ansaita myös pdf-tiedostojen lukuohjelmalla. Kun Acrobat Distiller, jolla pdf-tiedostoja tuotettiin PostScript-muotoisista tulostustiedostoista, maksoi 695\$, maksoi Acrobat Reader -lukuohjelma 50\$.

Alkuvaiheessa kilpailu vaikutti hetken aikaa kohtuullisen tasaväkiseltä, vaikka Adobella oli avoin, julkistettu tiedostoformaatti ja Envoy:n tiedostoformaatti oli yksin WordPerfectin ja sittemmin Novell Inc:in käytettävissä. Siinä vaiheessa kun Adobe lisäsi pdf-versioon 1.2 CMYK-väriavaruuden ja painotalojen yleisesti käyttämien spot-värien hallinnan sekä lopetti Reader-ohjelmalla rahastamisen, kilpailu ratkesi Adoben eduksi.

Adoben voittoon saattoi toki vaikuttaa WordPerfect Officen seikkailut Novellin ja Corelin omistuksessa, mutta kun pdf-tiedostoformaatti loi mahtavat markkinat ohjelmistotaloille, oli selvää, millä asiakirjaformaattilla mentiin eteenpäin. Markkinoille ilmestyi useita ohjelmistotaloja, jotka halusivat tuoda esille erikoisosaamisensa painoteollisuuden sovelluksissa tarjoamalla pdf-tiedostojen tulkintasovelluksia samaan aikaan, kun pienet freeware-ohjelmien⁴ tuottajat olivat innoissaan, sillä heidän pdf-ohjelmalla tehtyjä asiakirjoja pystyi kuka tahansa lukemaan laite- ja ohjelmistoympäristöstä riippumatta Adoben tarjotessa ilmaisen lukuohjelman lähes kaikkiin käyttöjärjestelmiin. Nykyään pdf-ohjelmia valmistavia yrityksiä on useita satoja, sillä pelkästään TUCOWS-ohjelmienjakelusivustolta löytyi tammikuussa 2009 pdf-aiheisia ohjelmanimikkeitä 820 kappaletta.

Pdf-ohjelmaa on kehitetty jatkuvasti lisäämällä siihen uusia ominaisuuksia, kuten eri väriprofiilien tai jopa 3D-sisällön hallinta. Jotta uusia ominaisuuksia voidaan hyödyntää, on myös pdf-

⁴ ilmaisohjelmat

tiedostoon määritettävä näiden ominaisuuksien kuvausmenetelmät, ja näin pdf-tiedostoistakin tulee aika-ajoin uusia versioita. Seuraavassa taulukossa kuvataan eri ohjelmaversioiden tärkeimmät ominaisuudet ja niitä vastaavat pdf-tiedoston versionumerot.

Taulukko 2. Pdf-versioiden ja -tiedostojen kehitys.

Versio	Vuosi	PDF	Ominaisuudet
Acrobat 1	1993	1.0	PDF 1.0 sisältää suurimman osan niistä ominaisuuksista, joita on sivunkuvauskielessä PostScript Level 2
Acrobat 2	1994	1.1	Tämän version myötä tuli mukaan La-väriavaruus ja kalibroitu RGB. Myös TrueType merkkilajien tuki oli mukana.
Acrobat 3	1996	1.2	Tämä versio mahdollisti värierottelun käytön. Lisäksi mukaan tulivat Unicode-merkistö ja CID-merkkilajit.
Acrobat 4	1999	1.3	PDF 1.3 sisältää koko PostScript Level 3 -grafiikkamallin. Se mahdollistaa monikanavaisen väriavaruuden käytön sekä tukee ICC-profiileja.
Acrobat 5	2001	1.4	Tästä versiosta lähtien PDF-tiedostossa voi olla mukana läpinäkyviä objekteja. Tämän myötä tuli rakenteellinen PDF, mikä mahdollistaa sisällön käsittelyn. Turvaominaisuuksia on lisätty ja JPG-pakkaus on otettu mukaan.
Acrobat 6	2003	1.5	PDF-dokumentteihin tuli tasot, mikä mahdollistaa vaihtoehtoiset sisällöt.
Acrobat 7	2004	1.6	Tämä versio tukee OpenType-merkkilajeja. Tiedostoon voi lisätä 3D sisältöä ja käyttäjä voi luoda virtuaalisivuja
Acrobat 8 Acrobat 9	2006 2008	1.7	Unicode-polkumäärittely yksinkertaistaa linkkien oikeaa määrittystä. PDF-pakkaukset mahdollistavat useiden eri tiedostomuotojen liittämisen samaan pdf-tiedostoon.

PDF-version 1.5 myötä tullut taso-ominaisuus tarkoittaa, että asiakirjalla voi olla vaihtoehtoisia sisältöjä riippuen siitä, mikä taso valitaan näkyväksi. Tätä voidaan hyödyntää esimerkiksi tehtäessä samasta esitteestä erikielisiä versioita sijoittamalla kunkin kielen tekstit omalle tasolle, jolloin selvittää yhdellä taitto työllä. Tasojen käyttö edellyttää, että ne ovat käytettävissä jo lähdeohjelmassa. Tällaisia ohjelmia ovat esimerkiksi Adobe Illustrator ja Adobe InDesign -ohjelmien uudemmat versiot.

2.4.2 Pdf-yhteensopivuustasot

Kun haluaa luoda hyviä pdf-tiedostoja, on päätettävä, mitä pdf-versiota käytetään. Pdf-version voi vaihtaa vaihtamalla toiseen asetusmäärittelyyn⁵ tai valitsemalla halutun yhteensopivuusasetuksen.

Seuraavassa taulukossa verrataan joitakin pdf-tiedostojen eri ominaisuuksia. Mitkä ominaisuudet ovat käytettävissä missäkin pdf-versiossa, ja miten kyseisen version tiedostot ovat avattavissa aiemmilla Acrobat- tai Reader-versioilla. Vaikka taulukko on tehty Adoben Acrobat-ohjelman versioiden mukaan, on se sovellettavissa myös muiden valmistajien ohjelmiin, niiden

⁵ Jobsettings. Asetusmäärittely, johon on tallennettu kaikki pdf-tiedoston luontiin tarvittavat määrittelyt

pdf-versiomäärittysten mukaisesti. Taulukossa mainittu RC4-salaus on eräs SSL-salauksen⁶ muoto.

Taulukko 3. Acrobat-versioiden yhteensopivuusvertailu. Eri pdf-versioiden vertailu tiedostojen avautumisen sekä läpinäkyvyyden, tasojen, väritilojen, kirjasimien ja suojausten tukemisen suhteen.

Seikka tai ominaisuus	Acrobat 4.0 (PDF 1.3)	Acrobat 5.0 (PDF 1.4)	Acrobat 6.0 (PDF 1.5)	Acrobat 7.0 (PDF 1.6), Acrobat 8 ja Acrobat 9 (PDF 1.7)
Pdf-tiedoston avaaminen	PDF-tiedostoja voi avata Acrobat 3.0:lla ja Acrobat Reader 3.0:lla sekä uudemmilla.	PDF-tiedostoja voi avata Acrobat 3.0:lla ja Acrobat Reader 3.0:lla sekä uudemmilla. Uudempien versioiden ominaisuudet voivat kuitenkin kadota tai eivät ehkä näy.	Useimmat PDF-tiedostot voi avata Acrobat 4.0:lla ja Acrobat Reader 4.0:lla sekä uudemmilla. Uudempien versioiden ominaisuudet voivat kuitenkin kadota tai eivät ehkä näy.	Useimmat PDF-tiedostot voi avata Acrobat 4.0:lla ja Acrobat Reader 4.0:lla sekä uudemmilla. Uudempien versioiden ominaisuudet voivat kuitenkin kadota tai eivät ehkä näy.
Kuvien läpinäkyvyys	Ei voi sisältää kuvaa, jossa käytetään läpinäkyvyyshosteita. Läpinäkyvyys on yhdistettävä muihin tasoihin ennen muuntamista PDF 1.3 -muotoon.	Tukee läpinäkyvyyden käyttämistä kuvassa. (Acrobat Distiller -toiminto yhdistää läpinäkyvyyden.)	Tukee läpinäkyvyyden käyttämistä kuvassa. (Acrobat Distiller -toiminto yhdistää läpinäkyvyyden.)	Tukee läpinäkyvyyden käyttämistä kuvassa. (Acrobat Distiller -toiminto yhdistää läpinäkyvyyden.)
Tasojen käsittely	Tasoja ei tueta.	Tasoja ei tueta.	Säilyttää tasot luotaessa PDF-tiedostoja sellaisista sovelluksista, jotka tukevat monitasoisten PDF-dokumenttien luomista, kuten Illustrator CS tai InDesign CS.	Säilyttää tasot luotaessa PDF-tiedostoja sellaisista sovelluksista, jotka tukevat monitasoisten PDF-dokumenttien luomista, kuten Illustrator CS tai InDesign CS.
Väritilojen tukeminen	DeviceN-väritila ja kahdeksan väriä ovat tuettuja.	DeviceN-väritila ja kahdeksan väriä ovat tuettuja.	DeviceN-väritilaa ja jopa 31 väriä tuetaan.	DeviceN-väritilaa ja jopa 31 väriä tuetaan.
Kirjasimien sisällyttäminen	Monitavuisia kirjasimia voidaan sisällyttää. (Distiller muuntaa kirjasimet sisällytyksen yhteydessä.)	Monitavuisia kirjasimia voidaan sisällyttää.	Monitavuisia kirjasimia voidaan sisällyttää.	Monitavuisia kirjasimia voidaan sisällyttää.
RC4-suojausten taso	40-bittistä RC4-suojausta tuetaan.	128-bittistä RC4-suojausta tuetaan.	128-bittistä RC4-suojausta tuetaan.	128-bittistä RC4-suojausta ja 128-bittistä AES-suojausta (Advanced Encryption Standard) tuetaan.

⁶ Secure Sockets Layer, mahdollistaa yhteyden vahvan salaamisen yksittäisen käyttäjän Internet-selaimen ja palveluntarjoajan Internet-palvelimen välillä. Yhteyden salaamisen lisäksi SSL-protokolla mahdollistaa molempien osapuolien vahvan todentamisen sertifikaattien eli varmenteiden avulla.

Yleisesti ottaen on suositeltavaa käyttää alinta pdf:n versiotasoa, jonka tietää riittävän halutunlaisen pdf-tiedoston tekemiseen. Näin varmistetaan, että mahdollisimman moni kykenee avaamaan tiedoston ja saa sen näkyviin laatijan tarkoittamalla tavalla. Käyttämällä pdf-versiota 1.4 varmistetaan, että tiedoston saa auki Acrobat 3:sta lähtien, ja kuitenkin pdf-versio 1.4 tukee asiakirjassa mahdollisesti olevia läpinäkyvyyksiä

2.4.3 Pdf:n keskeisiä ominaisuuksista

- Pdf-asiakirjan suojaus

Yleensä kun asiakirja lähetetään vastaanottajalle, sen pystyy lukemaan hänen lisäkseen kuka tahansa, joka saa tiedoston käsiinsä. Joissakin ohjelmissa tai joitakin salausohjelmia käyttäen voi tiedoston suojata salasanalla, mutta sen jälkeen kun vastaanottaja on avannut tiedoston, hän voi joko tallentaa asiakirjan uutena suojaamattomana asiakirjana tai tulostaa sen paperille, joka sitten jääkin vahingossa muiden nähtäville.

Nykyisissä pdf-versioissa voidaan asiakirja suojata usealla eri tavalla. Ensinnäkin asiakirjalle voidaan määrittää yleissuojaus, jolloin asiakirjan suojausasetukset ovat samat kaikille henkilöille ja vain asiakirjaan liitetyillä salasanoilla on merkitystä. Tämän lisäksi on mahdollista luoda asiakirjalle käyttäjäluettelo, jossa eri käyttäjille on määritetty yksilölliset suojausasetukset ja -oikeudet.

Asiakirjakohtaisessa suojauksessa voidaan määrittää kaksi eri suojaustasoa, joista kumpikin voidaan suojata omalla salasanallaan. Yksinkertaisimmillaan salanasuojaus tarkoittaa, että pdf-asiakirjalle määritetään salasana, jota kysytään tiedostoa avattaessa.

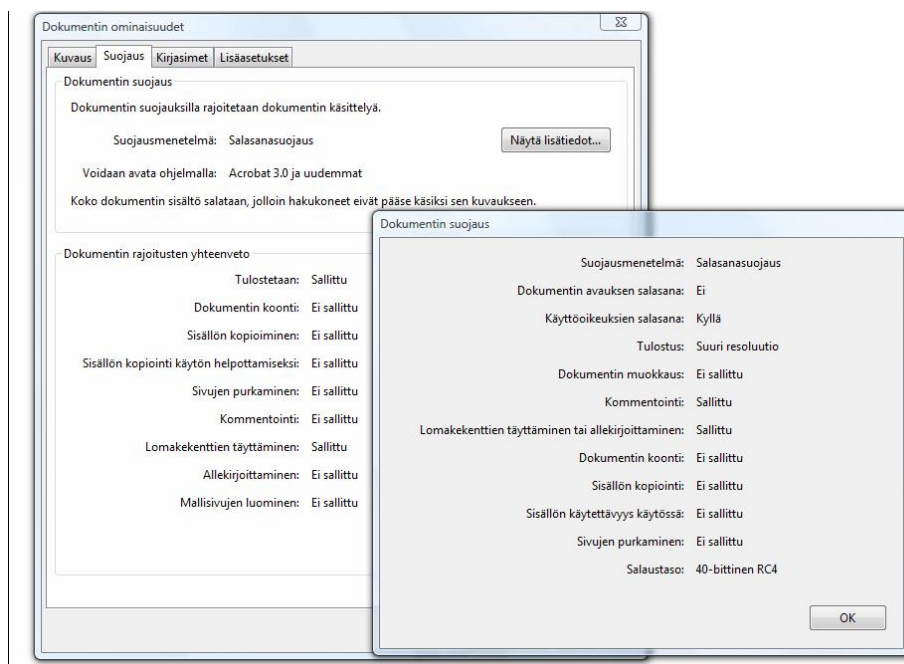
Pdf-asiakirjan käsittelylle voidaan määrittää myös erilaisia rajoituksia ja asettaa rajoituksiin liittyvä salasana, jonka avulla nämä rajoitukset voidaan ohittaa. Taskisen (2006, 80) mukaan Pdf-tiedoston rajoituksiin liittyvä salasana kannattaa määrittää, vaikka avauksen salasanaa ei käytätkään.

Pdf-tiedoston käyttöä voidaan rajoittaa esimerkiksi seuraavilla asetuksilla:

- asiakirjan tulostaminen (normaali/matala tarkkuus/ei lainkaan)
- asiakirjan sisällön kopioiminen leikepöydälle
- asiakirjan muuttaminen (sivujen poisto/lisäys)
- asiakirjan kommentointi/allekirjoittaminen
- asiakirjan tallentaminen jne.

Käyttäjakohtaisissa suojuksissa luodaan käyttäjille omat käyttäjäprofiilinsa. Tässä voidaan käyttää hyödyksi myös yleisiä, hyväksytyjä henkilötunnistusjärjestelmiä, jolloin henkilöprofiilit voidaan rakentaa näiden tietojen pohjalta. Suojattavaan asiakirjaan voidaan sen jälkeen määrittää käyttäjakohtaiset oikeudet huomioiden edellä mainitut tulostus-, kopiointi- ja muokkausmääritykset.

Asiakirjan suojuksen voi tarkistaa Adobe Readerillä *Tiedosto / Dokumentin ominaisuudet...* (*File / Document properties...*). Alla olevasta näyttöleikkeestä näkee, että kyseinen asiakirja on suojattu kopiointilta ja muokkaukselta. Asiakirjan avaaminen, tulostaminen ja lomaketäyttö on sallittu.



Kuvio 2. Pdf-asiakirjan suojausasetusten tarkistaminen Adobe Reader 9 -ohjelmalla sekä perustasolla että yksityiskohtaisesti. Esimerkkinä Elektroninen painoviestintä -pdf-julkaisun suojaus- ja rajoitustiedot.

Uusimmassa Acrobat-ohjelman versioissa 9 on mahdollista määrittää pdf-asiakirjaan liittyvä asiakirjan suojaus myös ajallisesti. Tämä tarkoittaa, että asiakirjan avaaminen voidaan sallia vain tietyssä ajankohtana, mikä perustuu siihen, että asiakirjaa avattaessa otetaan yhteyttä suojuksesta vastaavaan palvelimeen, ja asiakirja saadaan auki ainoastaan, jos suojauspalvelimen määrittäykset sen sallivat.

- Asiakirjan aitouden varmistaminen

Asiakirja voidaan myös allekirjoittaa sähköisesti, joko sijoittamalla asiakirjaan pelkkä allekirjoitus tai vahvistamalla asiakirjan sisältö oikeaksi erillisellä varmennusmenetelmällä. Jos asiakirjaa on muutettu, lukija saa joko suoraan asiakirjan avauksen yhteydessä varoituksen tai hän voi

allekirjoituksen tarkistuksessa saada tiedon, että asiakirjan sisältö ei vastaakaan allekirjoitetun version sisältöä. Kummassakin tapauksessa allekirjoitukseen käytetään digitaalista allekirjoitusta, joka voidaan luoda joko Adobe Acrobatissa itsessään tai käyttää jotakin yleistä, kolmannen osapuolen vahvistamaa henkilötunnistusmenetelmää, josta varmenne (certificate) saadaan.

Käytettäessä kolmannen osapuolen varmennemenetelmää voi pdf-asiakirjan vastaanottaja hakea lähettäjän vastatunnisteen varmenteen myöntäjän edustajalta. Jos taas käytetään Acrobatissa luotua käyttäjätunnistetta, pitää se jollakin tavalla siirtää asiakirjan vastaanottajalle, jotta hän voi varmistua allekirjoittajan henkilöllisyydestä. Varmenteen sähköpostilla lähettäminen sisältyy suoraan Acrobatin varmennekäsittelyyn, joten sen käyttö on helppoa ja turvallista.

Turvallisuutta voidaan vielä lisätä niin sanotun sormenjälkitiedon avulla. Siinä varmenteen vastaanottaja tarkistaa varmenteessa olevan sormenjälkitiedon esimerkiksi soittamalla varmenteen lähettäjälle, jolloin molemmat voivat varmistaa varmenteen muuttumattomuuden postisiirron aikana.

- Asiakirjan kommentointi

Asiakirjojen kommentointi kuuluu työryhmyöskentelyn perusasioihin. Pdf-asiakirjojen kommentointiin on useita eri tapoja. Siihen voidaan vaikka kirjoittaa huomautuksia muistilappuina, liittää erilaisia tiedostoja, piirtää kuvioita ja kaavioita tai korostaa tekstiä. Pdf-tiedoston kommentointi voidaan tehdä sekä Acrobatin omassa ohjelmaikkunassa että selainliittymän avulla. Selainliittymää käytettäessä useat käyttäjät voivat kommentoida samaa asiakirjaa lähes reaaliaikaisesti. Tätä toimintoa kutsutaan Online-kommentoinniksi.

Kommentointivälineinä Acrobat tarjoaa *Huomautustyökalun*, jolla tehdään kellovia muistilappuja, jotka voidaan avata myöhemmin kaksoisnapsautuksella. Muistilapun otsikkoriviltä näkee kommentintekijän nimen. *Leimasintyökalulla* voidaan asiakirjaan "lyödä" vaikka Hyväksyty-leima ja toimittaa asiakirja edelleen seuraavalle käsittelijälle.

Kommentoidun pdf-asiakirjan vastaanottaja voi helposti tutkia kaikki asiakirjaan tehdyt kommentit erillisessä kommentti-ikkunassa, jossa kommentin napsautus vie automaattisesti siihen asiakirjan kohtaan, johon kommentissa viitataan.

2.4.4 Pdf-standardoinnin tila

Pdf-tiedostoihin liittyviä ISO-standardointeja on vuoden 2009 alussa kaikkiaan 13 kappaletta. Osa niistä liittyy pdf-tiedostoissa käytettyihin toimintoihin tai muihin pdf-tiedostojen ja mui-

den järjestelmien välisiin liittyisiin. Varsinaisia eri pdf-versioihin liittyviä standardeja on neljä kappaletta:

- ISO 15930-1:2001 Graphic technology -- Prepress digital data exchange -- Use of PDF -- Part 1: Complete exchange using CMYK data (PDF/X-1 and PDF/X-1a)
- ISO 19005-1:2005 Document management -- Electronic document file format for long-term preservation -- Part 1: Use of PDF 1.4 (PDF/A-1)
- ISO 24517-1:2008 Document management -- Engineering document format using PDF -- Part 1: Use of PDF 1.6 (PDF/E-1)
- ISO 32000-1:2008 Document management -- Portable document format -- Part 1: PDF 1.7

Tässä selvitysraportissa perehdytään ennen kaikkea ISO 19005-1:2005 standardin mukaiseen PDF/A-1 -formaattiin.

2.5 PDF/A — arkistoinnin standardi

Asiakirjoja joudutaan usein arkistoimaan lainsäädännöllisistä syistä johtuen. Tällaisia asiakirjoja voivat olla esimerkiksi veroilmoitukseen liitteiksi tarvittavat kuitit, vakuutustodistukset tai sopimukset.

Aikana, jolloin kaikki oli vielä paperilla, ongelmana oli muistaa mihin kaappiin, kansioon tai laatikkoon sopimukset oli arkistoitu. Nykyään voidaan sähköisiä asiakirjoja hakea niiden nimien, sisällön tai jopa asiakirjan rakenteeseen liittyvien metatietojen perusteella.

2.5.1 Digitaalisen arkistoinnin ongelmat

Tämän päivän digitaalisessa maailmassa arkistoinnin ongelmat liittyvät yleisesti tiedostojen avaamiseen. Hakutoimintojen ja tietokantapalveluiden ansiosta asiakirjat ovat kyllä löydettävissä, mutta niiden avaaminen tai tulostaminen alkuperäisessä muodossa saattaa olla vaikeaa tai jopa mahdotonta. Ongelman merkitys korostuu, jos arkistointiaika on pitkä esimerkiksi yli 15 vuotta.

Ohjelmaa, jolla asiakirja oli aikanaan tehty, ei ole enää käytössä. Toisaalta merkkilaji, jota asiakirjassa oli käytetty ja jolla asiakirja oli tulostettu, olikin silloisen kirjoittimen muistikasetilla oleva merkkilaji, eikä kyseistä muistikasettia enää valmisteta. Monissa 80-luvun tekstinkäsittelyohjelmissa saatettiin merkkilaji ilmaista ainoastaan merkkilajin numerolla, eikä kukaan tiedä nykyään edes minkä kirjoittimen tai kasetin merkkilajeista oli kysymys.

Tällaisia ongelmia on esiintynyt muun muassa Eduskunnan veteraanikansanedustajien haastatteluaineistoissa. Äänikaseteille tallennetuista haastatteluista litteroitujen paperiaineistojen pohjalta on aikoinaan tehty sisällysluetteloita, joissa sivunumerot viittaavat otsikkotason asemesta suurpiirteisesti vain käsiteltävään aiheeseen.

Kun silloinen tekstitiedosto avataan nykyisin käytössä olevalla tekstinkäsittelyohjelmalla, ei tiedostomuunnoksen jälkeen olekaan enää tiedossa, millä merkkilajilla tai sivuasetuksella alkuperäinen teksti on tehty. Itse tekstisisältö on saatu muunnettua, mutta ulkoasuun liittyvää tietoa ei.

Asiakirjan sivutuksen muuttumisesta puolestaan seuraa, että sisällysluettelo ei enää pidäkään paikkaansa. Sivutusta voi kyllä yrittää muuttaa manuaalisesti vaikka marginaaleja muuttamalla, mutta tämä edellyttää yleensä, että käytettävissä on alkuperäinen paperiversio, josta selviää tavoiteltava sivumäärä ja marginaalit. Marginaalien säätäminen vaatii jo perusteellisempaa ohjelmiston ominaisuuksien hallintaa ja ennen kaikkea kärsivällisyyttä.

2.5.2 Miksi PDF/A eikä pelkkä pdf?

Eivätkö perinteiset pdf-asiakirjat ole riittävän hyviä pitkäaikaissäilytykseen? Pdf:llä on muutamia loistavia ominaisuuksia, joiden ansiosta se soveltuu arkistoasiakirjojen luontiin. Säilytyslaatikon tavoin myös pdf voi sisältää kovinkin erilaista aineistoa, kuten tekstiä, kuvia ja merkkilajeja. Lisäksi pdf pystyy tuottamaan uudelleen alkuperäisen asiakirjan ulkoasun tietojärjestelmäalustasta riippumatta. Tiettyjen ehtojen on kuitenkin täyttyvä, jotta asiakirja voitaisiin tuottaa täsmälleen samanlaisena.

Yhtenä ehtona on, että käyttäjällä on oltava käytettävissään kaikki asiakirjan elementit. Esimerkiksi asiakirjan merkkilajin täytyy olla upotettu, pelkkä linkki kyseiseen merkkilajiin ei riitä. Tämä tarkoittaa, että jos käyttäjä avaa kymmenen vuoden kuluttua asiakirjan eikä hänellä ole tietokoneellaan tarvittavia merkkilajeja, eivät tietyt erikoismerkit tai symbolit näykään oikein.

Lisäksi pitää välttää tiettyjä pdf:n ominaisuuksia. Ne saattavat heikentää asiakirjalta vaadittavaa luotettavuutta. Vältettäviä ominaisuuksia ovat esimerkiksi vuorovaikutteiset elementit ja pdf-

tasot. Nämä kaksi asiaa estävät asiakirjan yksiselitteisyyden, joka on edellytys PDF/A-asiakirjalle. Asiakirjassa saattaa esimerkiksi olla eri tasoilla vaihtoehtoisia tekstejä, joiden voimassaoloa ei voida enää viidenkymmenen vuoden kuluttua varmistaa. Asiakirjan sisältö pitää määrittää sen luontivaiheessa, eikä enää luettaessa.

PDF/A-asiakirja on periaatteessa perinteinen pdf-asiakirja, joka täyttää tietyt tarkat määrittelyt. Jotta käyttäjien ei olisi tarvinnut toistuvasti kokeilla eri versioita ja kiistellä parhaiten arkistointiin sopivasta pdf-tiedoston versiosta, teollisuuden asiantuntijat päättivät vuonna 2002 yhdessä kehittää PDF/A-standardin.

2.5.3 PDF/A-standardi

ISO hyväksyi PDF/A-standardin pitkäaikaisarkistointiin syksyllä 2005. PDF/A-standardi julkaistiin numerolla ISO 19005-1:2005 ja se pohjautuu pdf-määrittelyyn 1.4. Standardiin valmistellaan parhaillaan⁷ lisämäärittelyä PDF/A-2, joka perustuu pdf-versioon 1.7.

PDF/A-standardin tavoitteena on mahdollistaa sellaisten pdf-asiakirjojen luonti, joiden ulkoasu säilyy järjestelmien muuttumisesta huolimatta. Näiden tiedostojen tulee olla ohjelmistoriippumattomia, eikä niiden käyttöä saa rajoittaa niiden luontiin, varastointiin tai uudelleen tuottamiseen käytetyt käyttöjärjestelmät. Niin kauan kun toimitaan PDF/A-standardin mukaisesti, käytäntö ja teoria kulkevat käsi kädessä. Adoben ohjelmista Acrobat Professional 7 sisältää ainoastaan luonnoksen PDF/A:n toiminnoista, mutta Acrobat 8 ja 9 tarjoavat täysin PDF/A-standardin mukaiset luonti- ja tarkistusominaisuudet.

PDF/A-standardin julkaisemisen jälkeen markkinoille on tullut monia uusia tiedostojen luontiin ja tarkistukseen tarkoitettuja PDF/A-työkaluja ja -sovelluksia, pienistä yksittäiskäyttäjälle tarkoitetuista ohjelmien lisäosista laajoihin palvelinratkaisuihin. Näiden avulla voidaan luoda satojatuhansia arkistoasiakirjoja muutamassa tunnissa.

2.5.4 Kuinka luoda PDF/A-tiedostoja?

Kun luodaan PDF/A-tiedostoja, on monta ehtoa täytettävänä. Kulloinkin käytettävä prosessi riippuu siitä, onko asiakirja jo valmiiksi pdf-tiedostona vai pitääkö se luoda jostakin työtiedostosta.

⁷ 12.1.2009

- PDF/A-tiedoston teko tiedostosta tai datasta

Yleisin tapa pdf-tiedostojen tekemiseen tekstinkäsittelyohjelmasta, kuvankäsittelystä tai taitto-ohjelmasta on tulostaminen pdf-tulostinohjaimella. Pdf tuotetaan lähdeohjelmasta Acrobat Professionalilla, Distillerillä⁸ tai jollakin muulla pdf-muuntimella. Massakonversiota varten on ohjelmia, jotka tuottavat PDF/A-asiakirjoja tietokannoista tai tulostusjonoista.

- Skannattujen paperiasiakirjojen muunto PDF/A-muotoon

Usein asiakirjasta on olemassa ainoastaan paperiversio, kuten sopimuksista, laskuista tai kirjoista. Viime vuosina skannaustulokset on tallennettu usein bittikartta-TIFFinä.⁹ Nykyään yhä skannataan jo enenevässä määrin pdf-muotoon ja ennen pitkää jo suoraan PDF/A-asiakirjaksi. Käyttäjät voivat skannata asiakirjan käyttäen esimerkiksi Acrobat Professional 8:aa ja tallentaa tiedoston PDF/A-muodossa. Usein voidaan hyödyntää tekstintunnistusta ja tehdä asiakirjasta pdf-tiedosto, johon myöhemmin voidaan kohdistaa hakutoimintoja. Kuvia ja vanhoja asiakirjoja voidaan myös skannata PDF/A-muotoon. Myös massakäsittelyä varten löytyy ratkaisuja, joilla suuri joukko paperisia asiakirjoja voidaan skannata ja muuntaa PDF/A-muotoon.

- PDF-tiedoston muuntaminen PDF/A-asiakirjaksi

Monilla on pdf-asiakirjoja, jotka eivät kuitenkaan ole PDF/A-yhteensopivia. Joskus PDF/A-asiakirjan tekeminen alkuperäisestä asiakirjasta ei onnistu, koska se sijaitsee jossakin muualla, ja asiakirjasta on lähetetty sähköpostilla pelkästään pdf-versio. Pdf-asiakirjan muuntamiseen PDF/A-muotoon on olemassa useita menetelmiä. Adobe Acrobat 9 on eräs näistä ohjelmista. Adobe ei ole kuitenkaan ainoa markkinoilla oleva, tälle alueelle ohjelmia tuottava ohjelmistotalo. On olemassa useita eri tuotteita yhden käyttäjän ratkaisuista suurtuotantoon saakka.

- PDF/A-tiedoston varmistaminen

Kun PDF/A-asiakirjojen kanssa työskennellään päivittäin, on tiedostojen tarkistaminen tärkeää. Onko järkevää uskoa asiakirjan lähettäjää, kun hän väittää asiakirjansa olevan PDF/A-muodossa? Ennen kuin vastaanotetut asiakirjat tallennetaan arkistoon, on varmistuttava, että ne todella ovat PDF/A-yhteensopivassa muodossa. Asiakirjan PDF/A-kelpoisuuden tarkistamiseen on useita eri välineitä. Acrobat 8 Professionalin lisäksi on muitakin ohjelmia, kuten Callas Softwaren pdfaPilot-ohjelma, joka monien muiden toimintojensa lisäksi kykenee tarkistamaan ja luomaan PDF/A-asiakirjoja.

⁸ Adoben ohjelma pdf-tiedostojen tuottamiseen eräajona.

⁹ TIFF Tagged Image File Format, häviötön bittikarttamuoto

- Kuka hyötyy PDF/A-muodosta?

Monilla toimialoilla ja ammateissa on odotettu arkistointiin sopivaa pdf-standardia. PDF/A-standardista hyötyvät myös tutkimus ja opetus, koska pdf-asiakirjaan voidaan sisällyttää useita eri asiakirjatyyppisiä. Alla on lueteltu muutamia satunnaisesti poimittuja esimerkkejä erilaisista käyttökohteista.

- Sähköpostiviestien tallentaminen PDF/A-asiakirjana

Nykyään on yhä enemmän sähköistä kirjeenvaihtoa, joilla on sopimuksellinen tai muu juridinen merkitys. Jokainen, joka on joskus joutunut vaihtamaan sähköpostijärjestelmänsä toiseen, tietää mitä vaikeuksia siihen liittyy. Vanhan postijärjestelmän viestit eivät olekaan enää käytettävissä. Koska PDF/A on turvallinen asiakirjamuoto, on järkevää tallentaa tärkeät sähköpostit aika-ajoin tallennusmedialle PDF/A-muotoon.

- Esitteet, käsikirjat ja info-sivut PDF/A-muotoon

Useat yritykset ja viranomaiset tarjoavat jo runsaasti informaatiota ladattavaksi pdf-muodossa. Miksi ei näitä asiakirjoja tehtäisi saman tien tulevaisuuteen säilyvässä PDF/A-muodossa?

- Pdf-tiedostojen käytettävyys

Tiedon saannin mahdollistaminen myös näkövammaisille on tärkeää. Koska PDF/A tukee pdf-asiakirjan rakenteista sisältöä, on sen käsittely tekstinlukulaitteilla helppoa.

2.5.5 PDF/A:n välittömät hyödyt

Vaikka PDF/A-standardi suunniteltiinkin pitkäaikaisarkistointia silmällä pitäen, on sen käytössä etuja, joista on hyötyä välittömästi. Kaikki, jotka välittävät tietoa ja asiakirjoja asiakkaille, lukijoille tai yhteistyökumppaneille, voivat hyötyä PDF/A:sta.

PDF/A takaa tiedon eheyden. Tämä poistaa erilaiset ongelmat, kuten puuttuvat merkkilajit, mikä voi pahimmillaan tuottaa sotkua, kun merkit eivät näykään oikein¹⁰. Värienhallintatoiminnot korjaavat liian tummat tai vaaleat kuvat. Lisäksi PDF/A estää ongelmia, joita voi seurata pdf-asiakirjojen salauksen avaamisen tai tulostuksen yhteydessä.

¹⁰ Näin on käynyt esimerkiksi EK:ssa Wordistä makron avulla asiakirja-arkistoon tallennetuille PDF-tiedostoille. Writerilla PDF 1.2-tiedostoja.

Näin ollen henkilöt, jotka käyttävät PDF/A-tiedostomuotoa, helpottavat oman työskentelynsä lisäksi myös tiedostojen vastaanottajia. PDF/A ehkäisee monet niistä ongelmista, joihin muuten voitaisiin joutua jo heti tiedoston avaamisen yhteydessä.

2.5.6 PDF/A-1a ja PDF/A-1b

ISO on jakanut PDF/A-standardin kahteen yhteensopivuustasoon, PDF/A-1a ja PDF/A-1b. PDF/A-1a (taso A) edellyttää kaikkien PDF/A-standardin osien noudattamista. PDF/A-1b (taso B) edellyttää ainoastaan ulkoasuun liittyvää yksiselitteisyyttä. Seuraavassa taulukossa kuvataan PDF/A-standardiin liittyvät määrittelyt ja aliversioiden väliset erot.

Taulukko 4. Standardien PDF/A-1a ja -1b vertailu. (Drümmer, Oettler & von Seggern 2007, 15)

	ISO 19005-1:2005: PDF/A-1a (taso A)	ISO 19005-1:2005: PDF/A-1b (taso B)
Säännönmukaisuus	Täydellinen PDF/A yhdenmukaisuus	Rajoitettu PDF/A yhdenmukaisuus
Tarkoitus	Tuottaa arkistokelpoisia PDF-asiakirjoja, joissa on täysi saatavuus kaikkeen sisältöön	Tuottaa arkistokelpoisia PDF-asiakirjoja, jotka ovat ainoastaan ulkoasultaan täsmällisiä
PDF-versio	PDF 1.4	
PDF/A tunniste	Käyttäjät on veloitettu ilmaisemaan käytetty PDF/A-taso ja sen noudattaminen	
Metadata	Määritysten, kuten kirjoittaja, asiakirjan otsikko, luontipäivä ja lähdeohjelma, pitää esittää XMP-yhteensopivassa muodossa.	
Looginen rakenne	Rakenne ja saatavuus on toteutettava tagien avulla ja vaihtoehtoisilla kuvien kuvailumääreillä. Myös käytetty kieli on ilmaistava.	Ei mitään loogiseen rakenteeseen liittyviä vaatimuksia.
Salaus	Turva-asetukset on kielletty. Kyseisen PDF-tiedoston avaamisen ja käsittelemisen pitää olla mahdollista ilman salasanan tarvetta.	
Värit	Kaikki värit täytyy identifioida. Laiteriippuvat väriavaruudet pitää identifioida tulostusmääritysten mukaisesti.	
Läpinäkyvyys	Ei sallittu	
PDF-tasot (layers)	Ei sallittu	
Pakkaus	LZW-pakkaus ei sallittu	
	JPEG2000-pakkaus ei sallittu	
Merkkilajit	Kaikkien PDF-tiedostossa käytettyjen merkkilajien pitää olla käytettävissä (vähintäänkin merkkilajin alijoukkona) kyseisessä tiedostossa (upotettuna).	
	Merkkien koodit pitää kyetä liittämään glyyfeihin ¹¹ ilman tulkintavaihtoehtoja	
	Jokaisella merkillä on oltava Unicode-vastaavuus	-
Merkinnät	Kommentit, jotka ovat ääni- tai video-muodossa on kielletty. Perinteiset merkki- ja otsikkotyypiset kommentit on sallittu.	

¹¹ Kirjoitusmerkin esitysmuoto

	ISO 19005-1:2005: PDF/A-1a (taso A)	ISO 19005-1:2005: PDF/A-1b (taso B)
Viitattu sisältö	Viitattut kuvat tai sivusisällöt (ei-upotetut) on kielletty	
Vaihtoehtoiset kuvat	Vaihtoehtoiset kuvat (matalan tarkkuuden näyttökuvat) on kielletty	
Ohjelmointikiel	Upotetut Java-scriptit on kielletty	
Toiminnot	Tietyt toiminnot, kuten elokuva- tai äänitiedoston avaaminen tai lomakkeen lähettäminen tai alustaminen on kielletty	
Lomakkeet	Sallittu tietyin rajoituksin	

Aliversioiden 1a ja 1b kirjainten merkityksen muistamista helpottaa, kun ajattelee, että a (= all) sisältää koko standardin, kaikki sille määritetyt ominaisuudet, kun taas standardin perusmäärittäisiin riittävä osastandardi on merkitty versiokirjaimella b (= basic). Näitä 1b-version helpoituksia koko standardiin nähden ovat:

- saatavuutta sisältöön ei tarvita, ulkoasullinen täsmällisyys riittää
- loogiseen rakenteeseen liittyviä vaatimuksia ei ole, pelkkä ulkoasu riittää
- sisällön merkeillä tarvitse olla vastaavuutta Unicode-merkistössä.

Suuri osa erilaisista ohjelmista, jotka kertovat olevansa PDF/A-yhteensopivia, ovat yhteensopivia alistandardiin PDF/A-1b nähden.

3 Kyselyitten ja pdf-selvitysten toteutuminen

3.1 Menetelmien esittely

Tässä selvitysprojektissa keskityttiin kahteen asiakirjojen tuottamisalueeseen. Projektissa selvitettiin päivittäisten toimistoasiakirjojen tekemiseen ja käyttöön sekä valtiopäiväasiakirjojen arkistointiin liittyviä asioita kysymällä valituilta käyttäjiltä heidän pdf-tiedostojen käytöstään ja tutkimalla sähköisessä asiakirja-arkistossa olevia tiedostoja. Lisäksi projektin edetessä tuli esille tarve perehtyä PDF/A-arkistointimuotoon, joten siihen perehdyttiin tutustumalla käytettävissä olevaan materiaaliin.

3.1.1 Käyttäjäkyselyiden toteutus

Käyttäjäkyselyiden avulla selvitettiin miten eduskunnan käyttäjät hyödyntävät pdf-tiedostoja jokapäiväisessä työssään, miten hyvin he tuntevat pdf-tiedostojen eri käyttömahdollisuuksia ja minkälaisia tarpeita tai ongelmia heillä on näiden suhteen.

Kyselyt toteutettiin kahdenkeskisten keskustelujen muodossa, jotka yhtä puhelimitse toteutettua keskustelua lukuun ottamatta käytiin henkilökohtaisen tapaamisen muodossa. Keskustelun ohjaamista varten oli laadittu kyselylomake, johon myös vastaukset kirjattiin. (Liite 1).

Kyselylomakkeen kysymykset olivat lähinnä keskustelun kulkua ohjaavia, ja kysymyksiä tarkennettiin tarvittaessa, niin että kaikkien vastaukset olisivat keskenään yhteismitallisia. Kaikki kysymykset esitti sama henkilö, joten mainitunlainen toiminta oli mahdollista. Näin ei testilomakkeen käyttöäkään pidetty tarpeellisena.

- Kysymykset

Kaksiosaisen kyselylomakkeen ensimmäisessä osassa selvitettiin, minkälaisen asiakirjojen tuottamisessa kyseinen henkilö on mukana. Tähän käytettiin seuraavia kysymyksiä:

- Oletko mukana tuottamassa valtiopäiväasiakirjoja?
- Tuotatko julkaisuja?
- Tuotatko tiedotteita?
- Tuotatko asiakirjoja muiden tarkastettaviksi?

Jos henkilön vastaus johonkin näistä kysymyksistä oli myönteinen, selvitettiin kyseistä kohtaa tarkentavilla kysymyksillä, joilla saatiin lisätietoa käytetyistä asiakirjoista ja niiden tuottamisesta. Tarkentavat kysymykset olivat seuraavat:

- Minkälaisia?

- Käytetäänkö pdf-tiedostoja?
- Millä ohjelmilla?
- Miten?

Lomakkeen toisessa osassa käytiin läpi kyseisten henkilöiden pdf-tietämystä. Heiltä kysyttiin, missä muodoissa he saavat asiakirjoja ja onko niiden kanssa mahdollisesti esiintynyt ongelmia.

- Miten pdf-tiedostoja voi tehdä?
- Miten pdf-tiedostoja voi lukea?
- Mitä ominaisuuksia tai asetuksia pdf-tiedostoille voi tehdä?
- Oletko tehnyt pdf-tiedostoja omaan käyttöösi?
- Missä muodoissa saat tiedostoja?
- Mitä ongelmia?
- Muuta?

Kyselylomakkeessa ei alun perin ollut viimeistä kohtaa *Muuta?*, mutta kun heti ensimmäisessä keskustelussa tuli esille myös muita asiakirjojen käsittelyyn liittyviä asioita, lisättiin lomakkeeseen tuo kysymys. Näin varmistuttiin, ettei se jää huomiotta seuraavissa keskusteluissa.

- Henkilöiden valinta

Kyselyt kohdistettiin subjektiivisesti valittuihin henkilöihin, joiden työtehtävistä ja tietotekniikan käytöstä oli kokemukseen perustuvaa tietoa ja joilla uskottiin olevan tietoa pdf-tiedostojen käytöstä. Kyselyyn valitut 27 henkilöä olivat tekemisissä joko valtiopäiväasiakirjojen tai julkaisujen kanssa, tai sitten heidät tiedettiin henkilöiksi, jotka käyttivät pdf-tiedostoja muista syistä.

Selvityksen tarkoituksena ei ollut tutkia, kuinka moni henkilö eduskunnassa käyttää pdf-tiedostoja, vaan lähinnä se, miten he käyttävät pdf-tiedostoja. Tästä johtuen kysely kohdistettiin henkilöihin, joiden tiedettiin tai oletettiin käsittelevän pdf-tiedostoja.

Alun perin kyselyyn oli valittu 20 henkilöä, mutta muilta haastatelluilta saatujen vihjeiden perusteella kyselyyn lisättiin vielä 7 henkilöä. Keskustelujen edetessä ilmeni, että muutamat ennakkoon valitut eivät olekaan niin laajasti pdf-tiedostojen kanssa tekemisissä, kuin alun perin oli ajateltu ja koska kummallekin osapuolelle sopivan keskusteluajankohdan löytäminen osoittautui vaikeaksi, päätettiin heidän osalta luopua kyselystä.

Kyselyyn osallistuneiden määrä oli lopulta 23 henkeä. Kyselyt suoritettiin 26.6.–6.10.2008 välisenä aikana. Liitteessä 2 on esitetty yksityiskohtainen kyselyaikataulu, josta selviää myös kyselyyn osallistuneet henkilöt ja kyselyajankohdat.

Asiakirjatoimiston ja valiokuntien henkilöiden työstä muodostaa pääosan valtiopäiväasiakirjojen käsittely. Kirjaston henkilöitä valittiin kyselyyn sen vuoksi, että heidän pääasiallinen toimintansa tähtää asiakirjojen luettavaksi tarjoamiseen sekä julkaisujen tekoon. Kirjaston julkaisut ovat usein erilaisia raportteja ja selvityksiä, eikä niitä yleensä lueta valtiopäiväasiakirjoihin kuuluviksi. Alla olevasta taulukosta ilmenee, mihin eduskunnan organisaatio-osiin kyselyyn osallistuneet kuuluivat.

Taulukko 5. Kyselyyn osallistuneiden henkilöiden jakautuminen eri käyttäjäryhmiin.

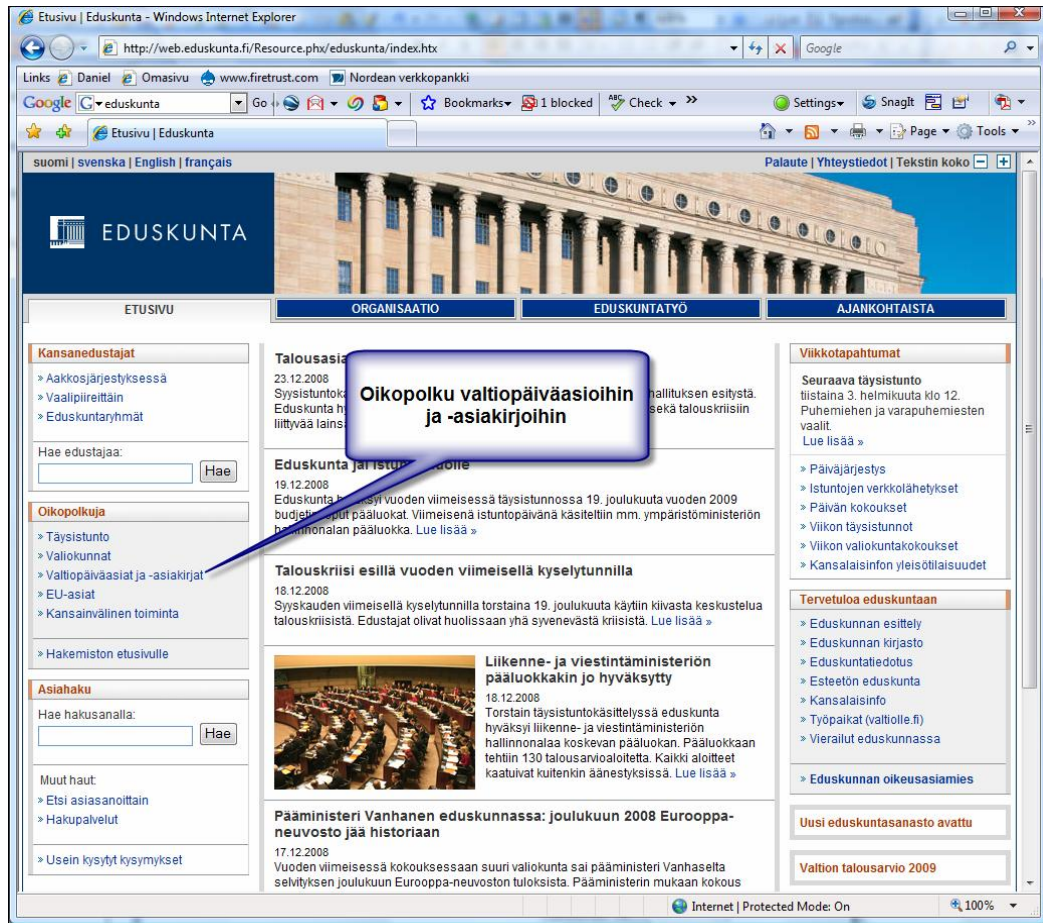
Ryhmä	lkm
Asiakirjatoimisto	7
Kirjasto	6
Pöytäkirjatoimisto	1
Tietohallintotoimisto	3
Valiokunnat	3
Avustajat	2
Eduskuntaryhmät	1
yhteensä	23

Haastateltavien henkilöiden kanssa käydyt keskustelut kestivät keskimäärin 55 minuuttia. Pissimmillään keskustelut kestivät lähes puolitonta tuntia ja vähimmillään vain 30 minuuttia.

3.1.2 Asiakirja-arkiston tutkiminen

Projektin toisena tehtävänä oli selvittää, minkälaisia pdf-asiakirjoja eduskunnan asiakirja-arkistoon on tallennettu, mitä pdf-tiedoston versioita niissä on käytetty ja millä ohjelmilla tai välineillä niitä on tuotettu. Tämän lisäksi tutkittiin, minkälaisia ongelmia pdf-asiakirjojen eri versioihin saattaa liittyä.

Eduskunnan asiakirja-arkisto on toteutettu usean eri tietokannan yhteissovelluksena, ja niille on kehitetty yhteinen html-selaimeen pohjautuva käyttöliittymä. Asiakirja-arkisto on kaikkien Internet-käyttäjien käytettävissä eduskunnan Internet-sivujen kautta.



Kuvio 3. Eduskunnan Internet-sivujen aloitusivu. (www.eduskunta.fi).

Asiakirja-arkiston hakupalvelu on liitetty eduskunnan www-sivustoon www.eduskunta.fi. Pääsivulla olevan *Valtiopäiväasiat ja -asiakirjat* -linkin kautta käyttäjä pääsee valtiopäiväasioiden ja -asiakirjojen hakusivulle.

Asiakirjojen haku on toteutettu monipuolisen dynaamisen hakulomakkeen avulla. Käyttäjä voi valita haun kohteeksi joko kaikki valtiopäiväasiakirjat tai sitten halutun asiakirjatyyppin mukaiset asiakirjat.

Asiakirjatyyppikohtaiset hakulomakkeet löytyvät seuraaville asiakirjatyypeille:

- Hallituksen esitykset - HE
- Valtioneuvoston selonteot ja tiedonannot - VNS, VNT
- Valtioneuvoston kirjelmät ja selvitykset (EU) - U, E, UTP
- Valtioneuvoston kirjelmien ja selvitysten (EU) jatkokirjelmät - Y, U, E, UTP
- Pääministerin ilmoitukset - PI
- Valtioneuvoston kirjelmät (ministerien sidonnaisuudet) - VN
- Valiokuntien mietinnöt ja lausunnot - VM, VL
- Eduskunnan vastaukset ja kirjelmät - EV, EK
- Aloitteet - LA, TPA, TA, KA, TAA, LTA, RA
- Kysymykset - VK, KVN, KK
- Puhemiesneuvoston ehdotukset ja luettelot - PNE, PNL
- Kertomukset - K
- Muut asiakirjat - MINS, M

– Talousarvioasiakirjat pääluokittain ja osastoittain

Asiakirjatyyppikohtaisissa hakulomakkeissa hakukenttien määrä vaihtelee haettavasta valtiopäiväasiakirjasta riippuen kolmesta yhteentoista. Periaatteena on, että lomakkeella on näkyvissä ainoastaan kyseisen haun kannalta oleelliset kentät.

The screenshot shows the search interface for parliamentary documents on the Eduskunta website. The page is in Swedish. The main heading is "Kaikki valtiopäiväasiakirjat alkaen 1991". Below this is a search form with the following fields:

- Hakusana: (empty text box)
- Asiakirjan tyyppi: Kaikki (dropdown menu)
- Asiakirjan tunnus: HE 25/2005 (text box)
- Valtiopäivät/Vaalikausi: Valtiopäivät 1991 -> (dropdown menu)

Buttons for "Hae" and "Tyhjennä" are located below the form. To the right of the form is a section titled "Ohjeet" (Instructions) with the following text:

Valtiopäiväasiakirjat — yhteishaku asiakirja-arkistoista
Katkaisumerkki on *.

Käytä tätä yksinkertaista haku, jos et tiedä, milloin asiakirja on laadittu tai minkä tyyppinen se on. Haulla löydät asiakirjoja vuodesta 1991 eteenpäin. Kaikki asiakirjasarjat eivät ole täydellisiä. Vuotta 1991 vanhempia asiakirjoja ei ole saatavissa sähköisessä muodossa.

Valitse asiakirjatyypittain räätälöity ja rajattu haku, kun tiedät asiakirjan tyyppin, esim. etsit hallituksen esitystä, valiokunnan mietintöä tai kansanedustajan kysymystä. Haku asiakirjatyypin mukaan valitaan sivun vasemmassa laidassa olevasta luettelosta.

Lisätietoa tietosisällöstä >

Haku kohdistuu asiakirjan koko tekstiin. Hakusana-kentän täyttämisen riittää haun suorittamiseen. Voit rajata hakuja valitsemalla asiakirjatyypin jaltai valtiopäivävuoden.

Ohjeita tiedonhausta >

Jos haet asian käsittely- ja päätöstietoja, valitse Valtiopäiväasiat >

At the bottom of the page, there is a footer with contact information: "Eduskunnan vaihde: (09) 4321, 00102 Eduskunta | Yhteystiedot | Aakkosellinen hakemisto | Ohjeet" and "Käyttöehdot | Palaute".

Kuvio 4. Valtiopäiväasiakirjojen hakulomake, jossa on näkyvissä kaikkia valtiopäiväasiakirjoja koskeva hakulomake.

Hakuehtojen syöttämisen ja hae-painikkeen painamisen jälkeen haun tuloksena tulee näyttöön hakutuloluettelo, jossa on korkeintaan 50 ensimmäistä kyseisillä hakuehdoilla löydettyä asiaa. Kerralla selattavan hakutuloluettelon kooksi voi vaihtaa 50, 100 tai 500 asiaa/sivu.

Hakutuloluettelo on muodostettu siten, että uusimmat asiat ovat luettelon alussa. Luettelon järjestys voidaan vaihtaa käänteiseksi, jolloin arkistojärjestelmä tuottaa uuden listan, jossa vanhimmat asiat ovat listan alussa.

Jos hakutuloksen kokonaismäärä on yli 5000 asiaa, ei järjestelmä pysty tuottamaan lajiteltua luettelo. Tällaisessa tapauksessa käyttäjälle annetaan asiasta varoitus. Tästä ei ole kuitenkaan

haittaa tavallisten, tiettyyn asiaan kohdistuvien hakujen yhteydessä, koska yleensä haku voidaan rajata niin, että hakutuloksen koko jää alle tuon lajitteluun liittyvän ylärajan.

Ongelma haittasi jonkin verran tätä pdf-tiedostojen selvitystä, kun etsittiin kaikkia tiettyyn asiar ryhmään kuuluvia asioita. Tehtäessä haku ryhmästä *Kaikki valtiopäiväasiakirjat* ja jättämällä hakukentät tyhjiksi saadaan hakutulokseksi 67435 asiaa.

3.1.3 Pdf ja PDF/A selvitysmenetelmät

Pdf-tiedostojen eri versioiden ominaisuuksia ja mahdollisuuksia selvitettiin tutkimalla pdf-tiedostoihin liittyvää kirjallisuutta ja Internet-sivustoja. Lisäksi perehdyttiin Amsterdamissa 10.–11.4.2008 järjestetyn kansainvälisen PDF/A-konferenssin aineistoon, johon kuului luentokalvomonisteiden lisäksi muun muassa kirja *PDF/A in a Nutshell*. Kyseiseen kirjaan on kerätty PDF/A-standardiin liittyvää tietoa useista eri lähteistä.

3.2 Käyttäjäkyselyn kautta selvinnyttä

Kyselyyn osallistui 23 henkilöä 26.6.–6.10.2008 välisenä aikana kahdenkeskisten keskusteluiden muodossa. Kyselyn myötä selvisi, että he osallistuvat mitä moninaisimpien julkaisujen, tiedotteiden ja valtiopäiväasiakirjojen tuottamiseen.

Tässä vaiheessa selvisi muun muassa syy joihinkin uusienkin pdf-tiedostojen ongelmiin. Joissakin asiakirja-arkiston asiakirjoissa on tulostusjälki selvästi alkuperäistä tulostetta huonompaa. Varsinkin joidenkin tavutettujen, kaksipalstaisten pdf-tiedostojen rivien viimeiset merkit saattavat mennä päällekkäin. Tätä selvitetään tarkemmin kohdassa 3.2.2. Liitteessä 3 on koostettu kaikkien kyselyiden vastaukset.

Valiokunnille on kehitetty yhteinen tietopankkijärjestelmä, mutta kaikissa valiokunnissa sitä ei ole otettu käyttöön. Käyttämättömyyden syyksi ilmenivät epäilyt sinne tallennettujen tietojen luottamuksellisuuden säilymisestä. Tietopankkiin tallennetaan tietoja eri tiedostomuodoissa, ja näin ollen siellä olevia tietoja voitaisiin jopa muuttaa ilman, että tätä huomattaisiin. Näin voisi tapahtua aivan vahingossakin.

Pdf-tiedostoihin liittyvää tietoutta henkilöillä oli melko hyvin, mutta heidäthän oli valittu kyselyyn heidän työtehtäviensä vuoksi, joten tämä oli odotettavissa. Samalla kävi myös ilmi, että pdf-tiedostojen tuottamistavan vaikutusta syntyvän pdf-tiedoston laatuun ei kuitenkaan tunnettu kovinkaan hyvin. Tämä tuli esille ennen kaikkea skannaamalla tehtyjen pdf-asiakirjojen rakenteen tuntemisessa, johon liittyvät ongelmat saatettiin yleistää koskemaan kaikkia pdf-tiedostoja.

3.2.1 Julkaisujen tuottaminen

Eri julkaisutyyppjä oli runsaasti artikkeleista, esitteistä ja verkkotiedotteista aina kirjoihin ja Eduskunnan julkaisusarjan julkaisuihin saakka. Eduskunnan julkaisusarjassa julkaistaan keskimäärin kymmenen julkaisua vuosittain. Verkkotiedotteita lukuun ottamatta julkaisut painetaan paperille, joko yksinkertaisesti tulostamalla monisteeksi tai lehtiseksi. Osa julkaisuista painetaan sidotuksi tai liimatuksi kirjaksi. Painettavat julkaisut toimitetaan painettaviksi yleensä pdf-tiedostoina. Laajimmat pdf-julkaisut ovat n. 1 000 sivuisia valtiopäiväasioitten arkistokirjoja, joita tuotetaan 50–60 kpl vuodessa.

Julkaisujen tuottamiseen käytetään pääsääntöisesti tekstinkäsittely- tai taitto-ohjelmia. Tekstinkäsittelyohjelmista pdf-tiedostot tehdään pääsääntöisesti tulostamalla ne joko Adobe Distiller tai PDF-XChange -tulostusohjainta käyttäen. Vaativimmat julkaisut tehdään yleensä Adobe InDesign -ohjelmalla, josta pdf-tiedosto tehdään suoraan pdf-export-toiminnolla. Adoben ohjelmilla pdf-tiedostoja tehtäessä osattiin kiinnittää paremmin huomiota pdf-tiedostojen laatuun, kuten asiakirjaan liitettävien kuvien väriavaruudesta.

3.2.2 Valtiopäiväasiakirjojen tuottaminen

Valtiopäiväasiakirjojen tuottaminen tapahtuu valtaosin tarkkaan määritettyjen työprosessien mukaan, eikä niissä käyttäjillä ole mitään valinnan mahdollisuuksia asiakirjojen rakenteen, asetusten tai niistä tehtävien pdf-tiedostojen suhteen. Valtiopäiväasiakirjojen tuottaminen vaihtelee asiakirjatyypistä riippuen, mutta suurin osa niistä tuotetaan Adobe Framemaker 5.5 -ohjelmalla.

Joissakin tilanteissa valtiopäiväasiakirja tehdään suoraan Word-tekstinkäsittelyohjelmasta. Tällöin käyttäjä ei kuitenkaan itse määrittele mitään asetuksia, vaan pdf-tiedoston tuottaminen tapahtuu painamalla Tallenna tekstiarkistoon -painiketta. Tästä käynnistyy Wordin VBA-makron ohjaama tulostusprosessi. Jos asiakirjan teksti on tavutettu ja kahdella palstalla, näyttävät joidenkin rivien lopussa viimeiset kirjaimet menevän päällekkäin. Tämä näkyy sekä katsottaessa pdf-tiedostoa näytöltä että tulostettaessa paperille. Todennäköisesti kyse on alkuperäisen merkkilajin korvautumisesta toisella merkkilajilla, jossa merkkien välistys poikkeaa alkuperäisen merkkilajin vastaavasta.

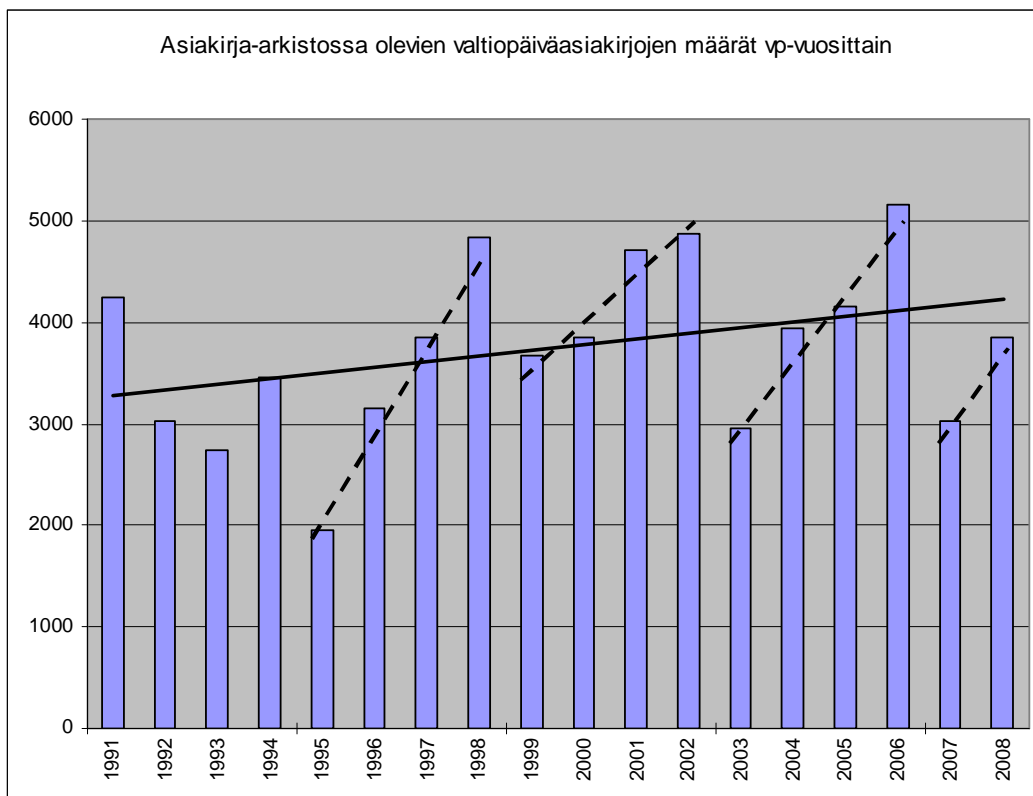
3.3 Asiakirja-arkiston aineistosta tehdyt havainnot

Asiakirja-arkistossa on kaikkiaan noin 73 500 asiakirjaa tai tarkemmin ilmaistuna eri asiaan liittyvää asiakirjaa. Kuhunkin asiakirja-arkiston asiaan voi löytyä useampi tiedosto. Joihinkin, erityisesti vanhempiin asiakirjoihin saattaa löytyä pelkästään tekstitiedosto, joka on tallennettu html-muotoon. Kaikkiaan 39 606 asiaan löytyy pdf-muotoinen asiakirja, mikä on runsas puolet kaikista asiakirjallisista asioista.

3.3.1 Asiakirja-arkiston asiakirjojen kokonaismäärä

Liitteessä 4 on asiakirjojen tarkat lukumäärät. Lukumäärissä on joitakin epätasällisyyksiä. Jos asiakirjoja hakee kohdasta kaikki valtiopäiväasiakirjat, saadaan kokonaismääräksi 67462, mutta jos lukumäärät haetaan asiakirjatyypeittäin ja lasketaan ne yhteen, saadaan yhteismääräksi 73 574. Tämä johtunee siitä, että kaikista valtiopäiväasiakirjoista haettaessa haku käsittelee rakenteisten ja ei-rakenteisten asiakirjojen arkistossa olevia asiakirjojen eri tavalla kuin haettaessa asiakirjatyypin mukaisilla hakulomakkeilla.

Asiakirja-arkiston aineisto elää samaa vaalikausittain toistuvaa 4-vuotissykliään kuin muukin eduskunnan toiminta. Vaalikauden toiminnan kiihtyminen kohti vaalikauden viimeisiä valtiopäiviä kohti tulee esiin myös asiakirja-arkiston aineistomäärissä, kuten seuraavan kuvion katkoviivoilla piirretyt trendiviivat osoittavat.



Kuvio 5. Valtiopäiväasiakirjojen määrät valtiopäivävuosittain vuodesta 1991 alkaen.

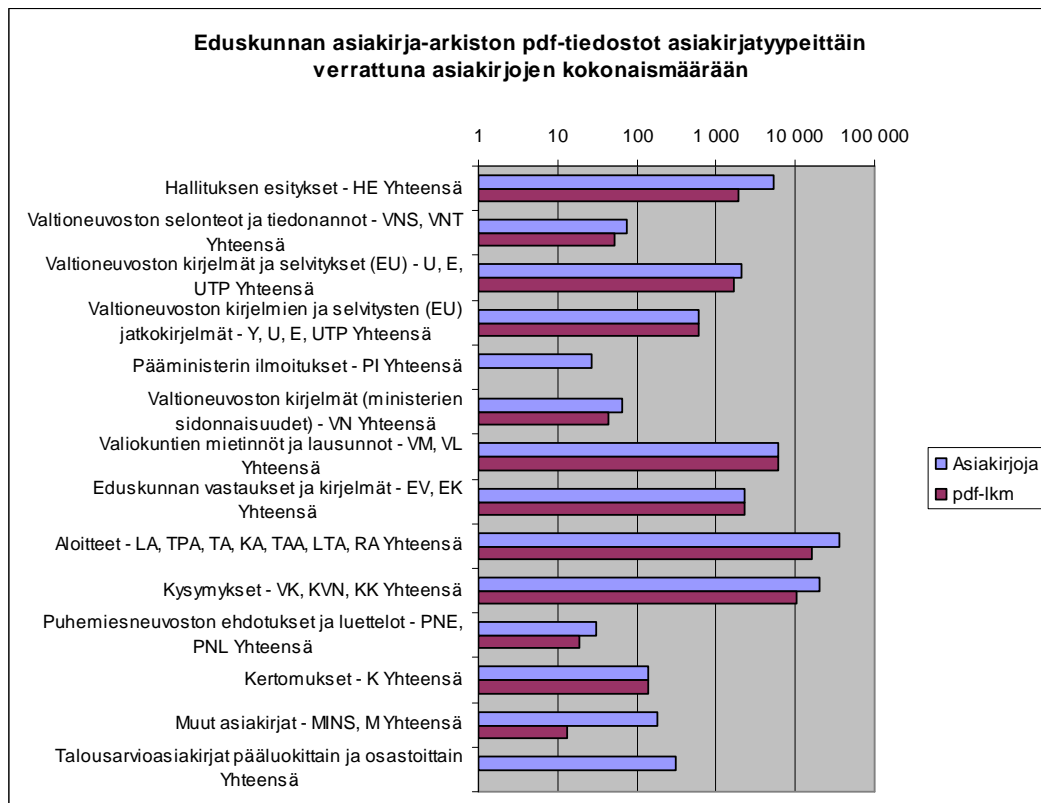
Asiakirja-arkiston ensimmäinen vaalikausi tekee poikkeuksen normaaliin 4-vuotissykliin. Kyseisen vaalikauden ensimmäisten valtiopäivien aikana yksittäisen kansanedustajan talousasioihin vaikuttamisen välineenä oli vielä raha-asia-aloite, jonka jättämiseen liittyvät rajoitukset olivat huomattavasti väljemmät, kuin vuonna 1992 käyttöön otetun talousarvioaloitteen. Kaikista asiakirja-arkiston vuosittaisista asiakirjamääristä laskettu trendiviiva osoittaa valtiopäivä asiakirjojen määrän tasaista kasvua, vaikka vaalikausien välillä onkin suuria eroja.

3.3.2 Pdf-tiedostojen määrät

Haettaessa asiakirjoja asiakirja-arkiston www-käyttöliittymän avulla yleisinä hakukriteereinä ovat hakusana, asiakirjan tyyppi ja tunnus sekä vaalikausi tai valtiopäivät. Vasta hakutuloksena saatavasta luettelosta selviää, missä tiedostomuodossa löydetty asiakirja on. Toisin sanoen asiakirjoja ei voi etsiä tiedostotyyppin perusteella. Pdf-tiedostojen kokonaismääräksi saatiin vähän yli 39 000 tiedostoa.

Seuraavassa kuviossa on esitetty asiakirja-arkiston asioiden kokonaismäärä asiakirjatyypeittäin sekä kuinka moni näistä on pdf-muotoisia. Valtioneuvoston jatkokirjelmät, Valiokuntien mietinnöt ja lausunnot, Eduskunnan vastaukset ja kirjelmät sekä Kertomukset ovat asiatyypejä,

joissa pdf-tiedosto löytyy jokaiseen asiaan. Toisaalta Pääministerin ilmoituksista tai Talousarvioasiakirjoista ei löydy yhtään pdf-tiedostoa.



Kuvio 6. Eduskunnan asiakirja-arkiston pdf-tiedostojen lukumäärä asiakirjatyypeittäin suhteessa asiakirja-arkiston asia/asiakirjamäärään. Huomaa, että asteikko on logaritminen.

Yllä oleva kuvio on esitetty logaritmisena, jotta pdf-tiedostojen osuus saataisiin näkyviin yhtä aikaa kaikissa asiakirjatyypeissä. Jos kuvio esitettäisiin lineaarisella asteikolla, siinä näkyisi käytännössä vain Valiokuntien mietinnöt, Aloitteet ja Kysymykset, koska niiden määrät ovat aivan eri suuruusluokkaa.

3.4 PDF/A:n mahdollisuudet ja tulevaisuus

PDF/A-standardista on kaksi eri versiota 1a ja 1b. Useat ohjelmat tuntuvat tukevan 1b-versiota, mutta PDF/A-1a:n kanssa tuntuu toistaiseksi olevan vielä hankalaa. 1a-versiossa on tallennettu myös asiakirja rakenne, mutta sen kuvaamisen mahdollistavien välineiden ja menetelmien tarjonta tuotantokäyttöön tuntuu olevan vähäistä. PDF/A-standardista ollaan tekemässä uutta versiota PDF/A-2, jonka julkaisun ei kuitenkaan yleisesti odoteta valmistuvan vielä vuonna 2009. Se pohjautuu pdf-versioon 1.6 ja sen keskeisimpiä uusia ominaisuuksia ovat muun muassa: JPEG2000-kuvatiedostojen pakkaus, läpinäkyvyys, valinnaiset sisältökomponentit, PDF-paketointi.

4 Mitä pdf-asioista saatiin selville ja ehdotuksia jatkosta

4.1 Selvityksen keskeiset tulokset ja päätelmät

Asiakirjatoimistossa oli havaittu ongelmia tiettyjen valtiopäiväasiakirjojen ulkoasussa. Ainakin yksi tällainen syy-seuraus-suhde löydettiin selvityksen yhteydessä. Jos pdf-tiedosto tehdään MS-Word-ohjelmalla tulostamalla asiakirja Adobe Acrobat 5.0 -ohjelman PDFWriter-ohjaimella ja asiakirjassa on tavutettua tekstiä kahdella palstalla, saattavat rivien viimeiset merkit mennä osittain päällekkäin.

Adobe Acrobat 5.0 -version käyttö juontaa alkunsa rakenteisten asiakirjojen tuottamiseen käytetystä Adobe FrameMaker 5.5 -ohjelmasta. Kun rakenteisten asiakirjojen järjestelmää kehitettiin, lisättiin ohjelman yhteyteen runsaasti räätälöityä ohjelmakoodia, jonka avulla ohjelman käyttöä helpotettiin. Tästä oli kuitenkin seurauksena, että ohjelman päivittäminen tuli ongelmalliseksi. Räätälöinti olisi jouduttu ohjelmoimaan uudestaan tai jättämään pois, mikä olisi taas hankaloittanut käyttöä.

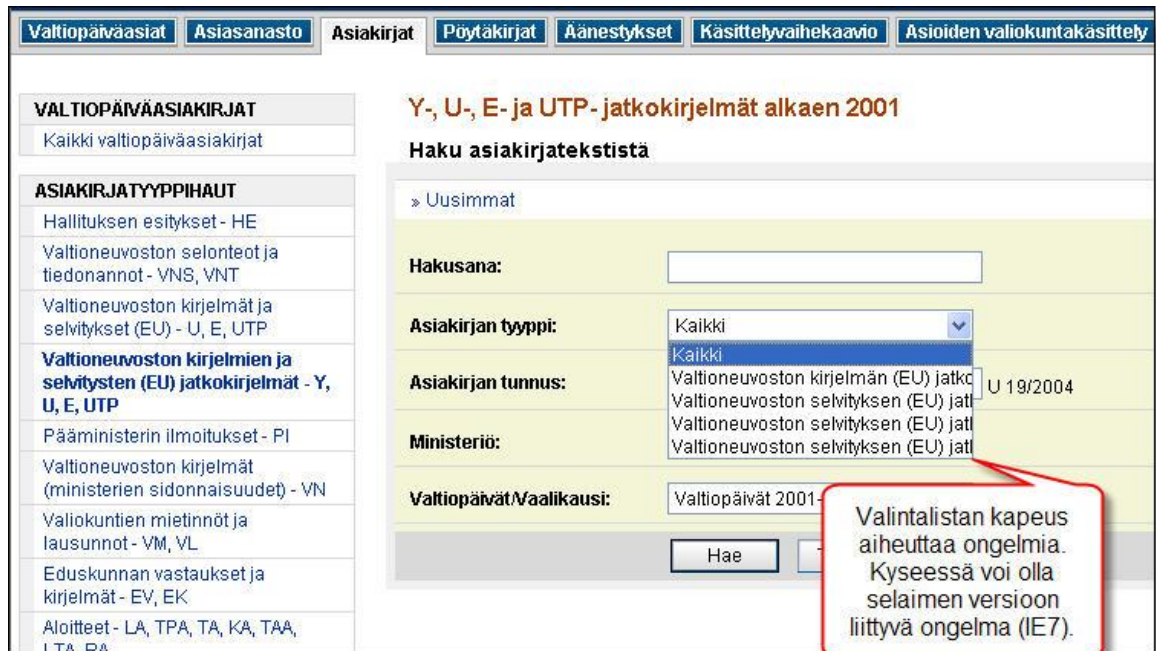
Kun FrameMaker 5.5 -ohjelman pdf-tuotanto perustui Adobe Acrobat 5.0 -versioon, ei sitäkään voitu päivittää uudempiin versioihin. Tätä itse asiassa kokeiltiin jossakin vaiheessa, mutta tuloksiin ei oltu tyytyväisiä.

Seuraavissa rakenteisten asiakirjojen kehitysprojekteissa kannattaa tarkoin harkita kuinka paljon ohjelmaan tehdään räätälöintejä ja millä menetelmillä ne toteutetaan. Ohjelmista on nykyään löydettävissä rajapinnat, joiden avulla vapaudutaan ohjelmariippuvuudesta, vaikka sovellukseen tehtäisiin suuriakin muutoksia.

4.1.1 Tietokannan pdf-tiedostojen hakemisen ongelma

Asiakirja-arkiston pdf-tiedostojen haku arkiston selainliittymän kautta aiheuttaa sen, että arkistossa olevia pdf-tiedostoja ei pysty hakemaan riittävällä varmuudella. Nykyisellä hakumenetelmällä haetaan lähinnä asioita, joihin liittyy mikä tahansa asiakirja. Hakua ei voi kohdistaa pelkästään niihin asioihin, joiden linkin kohteena on pdf-tiedosto, puhumattakaan, että haun voisi kohdistaa suoraan pdf-tiedostoon. Trip-tietokannan komentokielellä hakuja voisi mahdollisesti kohdistaa suoraan pdf-tiedostoihin, mutta sen käyttö on rajattu järjestelmän ylläpidon asiantuntijoille.

Internet Explorer 7 näyttää asiakirja-arkiston hakusivun valintaluettelon erilailla kuin aiemmat IE-versiot. Valintalistan kapeus aiheuttaa ongelmia, jos listan vaihtoehtojen alkuosat ovat samoja riittävän pitkälle. IE6-selainta käytettäessä valintalista näkyy siten, että valintalistan vaihtoehdot näkyvät kokonaisuudessaan. Tämä aiheuttaa ongelmia haluttaessa etsiä tiettyjen asiakirjatyypin asiakirjoja.



Kuvio 7. Esimerkki IE7-selaimessa valintalistan kapeuden aiheuttamasta ongelmasta.

Kuten yllä oleva esimerkki osoittaa, oli tiettyjen asiakirjatyypin haku IE7-selainta käyttäen hiukan hankalaa. Myöhemmin kävi kyllä ilmi, että esimerkiksi Google Chrome -selaimen kanssa valintalista näkyi aivan oikein. Onneksi kyseessä ei ollut kuitenkaan tehtävän toteuttamista estävä ongelma.

4.1.2 Tutkimusmenetelmien kehittäminen jatkotutkimuksia varten

Asiakirja-arkiston sisällön tutkimista varten olisi hyvä, jos tietokannan komentokieli olisi käytettävissä. On hyvin ymmärrettävää, että jatkuvasti käytössä olevaan tietokantaan ei voida antaa komentokielen käyttömahdollisuutta, koska seurauksena voisi olla koko järjestelmän kaatuminen.

Jos asiakirja-arkistosta halutaan täsmällisempää tietoa, olisi tietokantaan hyvä olla selainliittymää parempi saantikeino. Tietokannasta voitaisiin toden näköisesti ottaa kopio tiettyinä hetkenä, palauttaa se jollekin toiselle palvelimelle ja toteuttaa pdf-tiedostojen tutkiminen näin luodussa turvallisessa ympäristössä.

4.1.3 Tulosten yleistettävyyys

Selvityksen tulokset ovat yleistettävissä yleisten toimistoasiakirjojen osalta. Asiakirja-arkiston osalta tulokset ovat yleistettävissä vain, jos työskentely-ympäristö vastaa nyt selvityksen alla ollutta. Eduskunnan ohjelmistoympäristö lienee sellainen, että niitä ei esiintyne muualla.

4.2 Toimenpidesuositukset ja jatkotutkimusaiheet

4.2.1 Tallenna tekstiarkistoon -makron muuttaminen

Tehtäessä valtiopäiväasiakirjoja Wordillä tuotetaan pdf-tiedosto asiakirja-arkistoa varten painamalla makropainiketta *Tallenna tekstiarkistoon*. Tästä käynnistyy Wordin VBA-makro, joka tulostaa asiakirjan Acrobat PDFWriter -tulostinajurilla käyttäen pdf-versiota 1.2.

Näyttää siltä, että jos pdf-tiedosto on tehty PDFWriter-ajurilla pdf-versioon 1.2, asiakirjassa on tavutus päällä ja teksti on kaksipalstaista, saattavat rivien viimeiset kirjaimet sijoittua väärin ja osittain toistensa päälle. Tämän ongelman korjaamiseksi olisi kyseistä VBA-makroa muutettava esimerkiksi niin, että siinä käytettäisiin Acrobat Distiller -tulostusohjainta. Ongelma poistuu myös, jos käytetään Adobe Acrobat 6 -versiota tai uudempaa, sillä niissä ei ole enää PDFWriter-ajuria. Silloin on huomioitava, että Acrobat Distiller -tulostinajurin nimeksi on uusissa versioissa annettu Adobe PDF.

4.2.2 Pdf-tietouden lisääminen

Eduskunnan henkilöstöä pitäisi ohjeistaa ja kouluttaa pdf-tiedostojen käytössä. Käyttämällä pdf-muotoa alkuperäisformaatin asemesta, voidaan lisätä myös tietoturva. Monissa toimiston ohjelmissa voidaan asiakirjaan sisällyttää runsaastikin metatietoja, jolloin asiakirjassa voi olla piilossa sellaisia tietoja, joista käyttäjällä ei ole edes tietoa. Pdf-tiedoston tuottaminen voidaan tehdä sellaiseksi, että metatietoja ei oteta mukaan tähän luotavaan pdf-asiakirjaan.

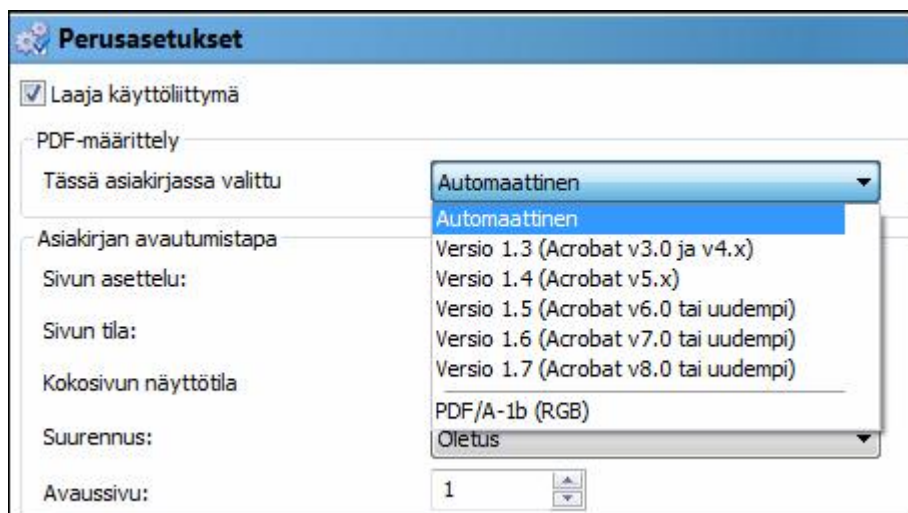
Joissakin ohjelmissa on mahdollista säilyttää muutoshistoria, jota tiedoston vastaanottaja on jopa voinut selata taaksepäin ja saada näin selville asioita, joita ei ollut lainkaan tarkoitettu hänen nähtäväkseen. Lisäksi on otettava huomioon, että vastaanottaja voi myös vahingossa muuttaa vastaanottamaansa tiedostoa, vaikkakin tämä on melko harvinaista.

Pdf-tiedostojen ominaisuuksia kannattaa tuki hyödyntää laajemminkin. Monissa tilanteissa on tarpeen suojata asiakirja ulkopuolisten katseilta tai vaikkapa sisällön kopioinnilta, jolloin kan-

nattaa käyttää asiakirjan suojausmahdollisuuksia. Joissakin tapauksissa saattaa olla tärkeää kyetä osoittamaan, että asiakirja on aito, että sen lähettäjä on juuri se kuin pitääkin ja että se on säilynyt juuri sellaisena, kuin lähettäjä oli sen lähettänyt.

Aikaisemmin pdf-tiedostojen laajemman käytön esteenä on ollut niiden tuottamiseen tarkoitettujen ohjelman puuttuminen eduskunnan perustyöasemasta. Pääsääntöisesti vain valtiopäiväasiakirjojen tuottamiseen ja julkaisujen tekoon tarkoitetuissa työasemissa on ollut Acrobat-ohjelma, jolla on voinut tehdä pdf-tiedostoja.

Eduskuntaan on hankittu vuonna 2004 PDF-XChange-ohjelman niin sanottu Site-lisenssi, joka sallii kyseisen ohjelman asentamisen ja käyttämisen kaikissa eduskunnan tietokoneissa. Ilman perusteellista selvitystä ja testaamista sitä ei kannata kuitenkaan ryhtyä käyttämään valtiopäiväasiakirjojen laadintaan eikä tähän ole mitään syytäkään, sillä nykyinen valtiopäiväasiakirjojen tuottamisjärjestelmä toimii nykyvaatimusten mukaan hyvin. PDF-XChange-ohjelman ominaisuudet riittävät kuitenkin oikein hyvin tavallisten pdf-tiedostojen tekoon, siis normaaleilla toimisto-ohjelmilla tehtävien, jakeluun tarkoitettujen asiakirjojen tekoon.



Kuvio 8. Kuva, josta ilmenee PDF-XChange 4 -ohjelman pdf-määrittelyn valintaikkuna, ja siinä olevat versiovaihtoehdot.

Nykyisin käytössä olevalla PDF-XChange 4 -ohjelmalla voi tehdä pdf-tiedostoja versiosta 1.3 versioon 1.7 ja jopa PDF/A-muotoakin, tosin vain sen 1b-versiota. Pdf-tiedostolle voi myös määritellä avautumistavan, jota käyttämällä voidaan helpottaa pdf-tiedoston avaajaa, kun pdf määritellään avautumaan juuri kyseiselle tiedostolle sopivassa koossa. ja tilassa.

4.2.3 PDF/A-version jatkoselvitys ja kokeilun käynnistäminen

PDF/A-version käyttömahdollisuuksien tutkimiseksi kannattaa käynnistää jatkoselvitys. Selvityksen tuloksena pitäisi saada selville mitä eri toimijoita PDF/A-formaatin yhteydessä on käytävissä ja tukevatko ne myös PDF/A-1a-standardiin pohjautuvaa rakenteellista pdf-tiedostoa. Jatkoselvityksellä pitäisi löytää, mitä eri sovelluksia ja konsepteja 1a-version hyödyntämiseksi on olemassa.

4.2.4 Pdf-tiedostojen tuottamisen automatisointi

Kun käyttäjä tekee pdf-tiedoston työasemaassa olevalla ohjelmalla, vaikuttavat monet työaseman asetukset syntyvän pdf:n laatuun. Tästä seuraa, että eri työasemilla tehdyt pdf-tiedostot saattavat poiketa toisistaan. Edelleen aina uuden työasemamallin käyttöönoton yhteydessä on tehtävä perusteelliset testaukset muun muassa pdf-tiedostojen laadun varmistamiseksi.

Työaseman ohjelmistolla tapahtuvan pdf-tiedostojen tuottamisen sijalle kannattaisi kehittää palvelin pohjainen pdf-tuotantojärjestelmä, joka eliminoisi työaseman vaikutukset pdf:n laatuun. Tähän on olemassa useita erilaisia ratkaisuja, joten tällaisen järjestelmän hankkimista varten olisi käynnistettävä oma selvitysprojektinsa.

Palvelin pohjaiset pdf-ratkaisut koskevat ennen kaikkea valtiopäiväasiakirjoihin liittyviä pdf-tiedostoja. Normaalien toimistoasiakirjojen pdf-versioiden tuottamiseen nykyiset työasema pohjaiset ratkaisut riittävät hyvin. Skannaamalla tuotettujen pdf-tiedostojen käyttöä voitaisiin parantaa tarjoamalla käyttäjille mahdollisuus muuntaa kuva-pdf-tiedostot tekstintunnistusohjelman avulla kopioitaviksi pdf-tiedostoiksi, joista tekstin saa tarvittaessa erotettua ja tuotua edelleen tekstinkäsittelyyn.

Lähteet

Drümmer, O. Oettler, A & von Seggern, D. 2007. PDF/A in a Nutshell. PDF/A Competence Center. Berlin.

Eduskunnan automaatiotyöryhmä. 1985. Eduskunnan lainsäädäntötyöhön liittyvän tietojenkäsittelyn kokonaistutkimus. Valtion painatuskeskus.. Helsinki.

Eduskunnan tietojärjestelmäarkkitehtuurin määrittelyvaihe. Eduskunnan kanslian julkaisu 13/2005, Eduskunta. Helsinki.

Katajajuuri, J-M. Loikkanen, T. 1999. Elektroninen painoviestintä. Valtion teknillinen tutkimuskeskus. Espoo.

Helkama, I. 2002. Eduskunnan julkaisutoiminta. Kuka, mitä, kenelle ja miten? Eduskunnan kanslian julkaisu 10/2002. Helsinki.

Kuronen, T. 1998. Hajautettu dokumenttien hallinta. 2. painos. Oulun yliopistopaino. Oulu.

Mustajärvi, O. 14.1.2009. Eduskunnan kanslian tietohallintotoimisto. Kehittämispäällikkö. Keskustelu. 14.1.2009.

Nurmeksela, R. Virtanen, M. Lehtinen, A. Järvenpää, M. & Salminen, A. 2006. Suomalaisen lainsäädäntötyön tiedonhallinta. Eduskunnan kanslian julkaisu 2/2006, Liite 8. Eduskunta. Helsinki.

PDF-yhteensopivuustasot. Luettavissa:

http://help.adobe.com/fi_FI/Acrobat/9.0/Professional/WS25210BC7-2345-4e30-A05C-80903A3B36EE.w.html. Luettu 12.1.2009

Prepressure.com. The history of PDF. Luettavissa:

<http://www.prepressure.com/pdf/basics/history>. Luettu: 6.1.2009.

Taskinen, H. 2006. Acrobat julkaisijan työvälineenä. WSOY Pro/Docendo-tuotteet. Jyväskylä.

Valtiopäiväasiakirjat eduskunnan asiakirja-arkistossa. Luettavissa:

http://www.eduskunta.fi/thwfakta/yht/ohje/akx_tekstiarkistot.htm#sisalto. Luettu 8.1.2009

1.7.2008

Kyselyn tiedot

Ajankohta:
Henkilö:
Tentävä:

Pääkysymys	Minkälaisia	Käytetäänkö pdf-tiedostoja (kyllä, ei, en tiedä)	Millä ohjelmilla	Miten
Oletko mukana tuottamassa valtio-päiväasiakirjoja				
Tuotatko julkaisuja				
Tuotatko tiedotteita				
Tuotatko asiakirjoja muiden tarkastettaviksi				

47

Kysymys	Vastaus
Miten pdf-tiedostoja voi tehdä	
Miten pdf-tiedostoja voi lukea	
Mitä ominaisuuksia tai asetuksia pdf-tiedoille voi tehdä	
Oletko tehnyt pdf-tiedostoja omaan käyttöösi	
Missä muodoissa saat tiedostoja	
Mitä ongelmia	
Muuta	

Haastateltava	Puhelin	Yksikkö	Ajankohta	Status	Yhteyspvm	Muuta
Neittaanmäki Pertti	4119	Avustaja	26.06.2008	13:00 ok		pk ok
Virta Päivi	2161	AKT	30.06.2008	10:00 ok		pk ok
Kojo Kirsti	2572	THT	30.06.2008	11:00 ok		pk ok
Jääskeläinen Maija-Liisa	2589	THT	30.06.2008	14:00 ok		pk ok
Heikkinen Ulla	2116	PKT	01.07.2008	10:00 ok		pk ok
Paasio Leena	2143	AKT	01.07.2008	11:00 ok		pk ok
Hietala Raija	3429	Kirjasto	01.07.2008	14:00 ok		pk ok
Kaakkola Jaana	2145	AKT	02.07.2008	10:00 ok	01.07.2008	pk ok
Katri Rauhala	2040	VK	02.07.2008	11:00 ok	30.06.2008	pk ok
Inga Luukko	2063	VK	03.07.2008	13:00 ok	30.06.2008	pk ok
Sanna Turpeinen	2053	VK	03.07.2008	14:00 ok	30.06.2008	pk ok
Juutilainen Helena		Avustaja	03.07.2008	15:00 ok	03.07.2008	pk ok, puhelimitse
Karjalainen Leena	3439	Kirjasto	04.07.2008	11:00 ok	30.06.2008	pk ok
Nyström Tuula	2576	THT	08.07.2008	13:00 ok	01.07.2008	pk ok
Holmberg Jaana	3325	Kirjasto	08.07.2008	14:00 ok	07.07.2008	pk ok
Graae Kristiina	3464	Kirjasto	09.07.2008	10:00 ok	07.07.2008	pk ok
Apilo Ari	2141	AKT	09.07.2008	14:00 ok	07.07.2008	pk ok
Nummenpää Kirsti	2162	AKT	10.07.2008	10:00 ok	08.07.2008	pk ok
Haukka Hilppa	3467	Kirjasto	14.07.2008	10:00 ok	10.07.2008	pk ok
Karhula Päivikki	3485	Kirjasto	17.07.2008	13:00 ok		pk ok
Ahloth Asta	3235	Ryhmät	11.08.2008	13:00 ok		pk ok
Vehmas Antti	4858	AKT	12.08.2008	10:00 ok	03.07.2008	pk ok
Juusti Riitta	2147	AKT	06.10.2008	10:00 ok	03.10.2008	pk ok
Räty Teuvo	2115	PKT		auki	11.08.2008	
Siitonen Pertti	2111	PKT		auki	11.08.2008	Sairaslomalla
Vekuri Tuija	2091	VK		auki	11.08.2008	
Salo Kari	2511	Muut		auki	25.08.2008	

8.12.2008

Kyselyn tiedot

Ajankohta: 26.6.–6.10.2008

Henkilö: 23 kpl

Tehtävä: Asiakirjatoimisto, Kirjasto, Tietohallintotoimisto, Valiokunnat, Avustajat, Ryhmät, Pöytäkirjatoimisto

1 Pääkysymys	a) Minkälaisia	b) Käytetäänkö pdf-tiedostoja (kyllä, ei, en tiedä)	c) Millä ohjelmilla	d) Miten
1.1 Tuotatko julkaisuja	<ol style="list-style-type: none"> 1 artikkelit 2 eduskunnan julkaisusarja 3 esitteitä 4 heteka 5 kirjaston julkaisusarja (n. 10 kpl) 6 kirjoja 7 kokousohjelmia 8 Neuvonnan ja lainauksen käsikirja 9 raportteja 10 register 11 sammandrag 12 sisäisiä lehtiä 13 verkkotiedotteita 14 VN ministerien puheita 15 vp-asiakirjasarja 	<ol style="list-style-type: none"> 1 Frame-versiona Editaan 2 kirjapainoon pdf-tiedostona 3 kirjaston julkaisusarja (v. 2003 alkaen myös pdf-muodossa) 	<ol style="list-style-type: none"> 1 Acrobat 2 CorelDRAW 3 Distiller 4 InDesign 5 Photoshop 6 Ventura 7 Word 8 WordPerfect 	<ol style="list-style-type: none"> 1 lähettää tiedostoja kirjapainoon 2 pdf-export 3 tulostamalla PDF-XChangeen

8.12.2008

1 Pääkysymys	a) Minkälaisia	b) Käytetäänkö pdf-tiedostoja (kyllä, ei, en tiedä)	c) Miillä ohjelmilla	d) Milten
1.2 Oletko mukana tuottamassa valtiopaivaasiakirjoja	<ol style="list-style-type: none"> 1 arkistointi 2 arkiston kokoaminen 3 atk-tukipyyntöjen kautta 4 edustajan aloitteet ja kysymykset 5 hallituksen esitykset 6 Istunto PK 7 kertomukset 8 kirjalliset kysymykset 9 kirjaston kertomus 10 lakialoitteet 11 mietinnöt ja lausunnot 12 ollut laatimassa asiakirjojen tuotantoprosessia 13 painatusvaiheessa 14 päiväjärjestykset 15 sammandrag 16 selonteot 17 u-asiat 18 valiokunnan mietinnöt ja lausunnot 19 valiokunnan pöytäkirja 	<ol style="list-style-type: none"> 1 netissä kertomuksen pdf-versio tulee paperilla, johon kerätään korjaukset. VK:n osastosihteeri tekee korjaukset Frameen. 2 	<ol style="list-style-type: none"> 1 1. vaiheen pdf-tiedosto tehdään ptj:stä makrolla. M-L:n avulla poistetaan. Lopullinen tulee Editasta ja laitetaan väliaikaisen tilalle. Tässä vaiheessa vain pdf on lopullinen versio. 2 EFI Fiery (väritulostimen työnkulku) 3 Framemaker 4 Framemaker 5.5 + SGML 5 pdf-writer 6 Pitstop (pdf-tarkastus) 7 Quite Imposing (arkkiaseointi) 8 Word 9 Xerox FreeFlow (työnkulku = ps-tiedoston käsittely RIP:n kautta kuvaksi) 	<ol style="list-style-type: none"> 1 Adobe-painike 2 Arkistokirja: Framen Book-toiminto. Bookin pdf-tallennus tekee automaattisesti kirjaan valituista tiedostoista pdf-tiedostot. 3 prosessin mukaisesti 4 tallenna nimellä 5 Tallenna tekstiarkistoon -makro tuottaa pdf-tiedoston. Acrobat Writer Pdf 1.2. Poikkeustilanne (kartat): Tallenna nimellä -valikosta pdf:ksi. 6 tulostamalla
1.3 Tuotatko asiakirjoja muiden tarkastettaviksi	<ol style="list-style-type: none"> 1 Istunto PK 2 pöytäkirjoja intranettiin 3 puheita ja kannanottoja 4 pöytäkirjoja 	<ol style="list-style-type: none"> 1 web-sivuille pdf-versio 	<ol style="list-style-type: none"> 1 PDF-XChange 2 Acrobat 3 Word 	<ol style="list-style-type: none"> 1 tiedosto tulosta
1.4 Tuotatko tiedotteita	<ol style="list-style-type: none"> 1 painopiste 2 asiakastiedotteet 3 lehdistötiedotteet 4 käyttöohjeita 5 sisäisiä tiedotteita 6 sähköposti tekstikentässä 	<ol style="list-style-type: none"> 1 pdf:nä verkkoon 2 pdf-tiedostoina postitse ja verkkoon 	<ol style="list-style-type: none"> 1 InDesign 2 Word, jotta muutkin voivat päivittää 3 WordPerfect 	<ol style="list-style-type: none"> 1 pdf-export 2 tulostamalla

8.12.2008

	Vastaus
2 Kysymys	
2.1 Miten pdf-tiedostoja voi tehdä	<ol style="list-style-type: none">1 Acrobat Distiller2 edustajien CV Framemakerilla3 Framemaker4 Framemaker tallentaminen, tallentaminen kirjana5 InDesign6 pdf-export7 pdf-tools8 PDF-XChange9 PP-esityksen sijasta pdf-esitys, jossa voi olla sekä pysty että vaakarivejä10 skannaamalla11 Wordin Acrobat-painikkeet
2.2 Miten pdf-tiedostoja voi lukea	<ol style="list-style-type: none">1 Acrobat 52 Adobe Reader 83 Illustrator4 oletusasetus5 PDF-XChange Viewer
2.3 Mitä ominaisuuksia tai asetuksia pdf-tiedostoille voi tehdä	<ol style="list-style-type: none">1 aukeamistapa2 bookmarkit3 fonttien tallentaminen4 kaksitasoinen salasana: luku/muutos5 kopioinnin esto6 painoasetuksia options-tiedostolla, kokomuunnoksia7 salasanasuojaukset8 sisällysluettelo9 tulostuksen esto10 versio11 väri- ja grafiikka-asetukset

8.12.2008

	Vastaus
2 Kysymys	
2.4 Oletko tehnyt pdf-tiedostoja omaan käyttösi	<ol style="list-style-type: none">1 arkistointiin2 ei yleensä pdf3 esimerkiksi pp-esityksistä4 kyllä5 skannaamalla6 vedoskäyttöön
2.5 Missä muodoissa saat tiedostoja	<ol style="list-style-type: none">1 asiantuntijoiden lausunnot joko doc tai rtf2 eps3 Excel4 Framemaker5 gif6 InDesign-taittoja7 jpg8 pdf9 tiff10 VN lähettää Wordinä, mutta pdf voisi olla parempi11 Word12 WordPerfect

8.12.2008

	Vastaus
2 Kysymys	
2.6 Mitä ongelmia	<ol style="list-style-type: none">1 Adobe Reader 7 integroituminen selaimen kun käytössä on myös Acrobat 5 Professional2 Automaattipäivitysten ja automaattikenttien kanssa joskus ongelmia3 CMYK-muunnos4 ei ongelmia5 epävarmuuksia pdf:n tekemisessä6 eri versioiden avautuminen7 fonttien heittäminen8 Framemakerin jäykkyys, attribuuitit ja elementit9 Joskus kuva katoaa tehtäessä wordista pdf-tiedostoa PDF-XChange-ohjelmalla. Lisäksi ongelmia linkkien kanssa (linkit toiseen pdf-tiedostoon)10 joskus pdf-kuvana, joskus tekstinä (tämän pitäisi onnistua lukijan ikkunasta11 kirjapainojen erilaiset pdf-asetukset12 kuvan resoluutio13 kuvissa riittämätön resoluutio14 monia eri tapoja tuottaa pdf-tiedostoja => laatu vaihtelee15 pdf:n liittäminen Wordiin16 pdf:n tuottaminen pitäisi siirtää työasemalta palvelimelle, jolloin laatu olisi tasaisempaa ja versioylläpito helpompaa17 pdf:stä kopioitu teksti ei asemoidu Wordiin oikein18 pdf-linkkien toimimattomuus PDF-XChangeassa19 pdf-versio-ongelmia (ei rak -makrolla Possu tai Naama ruots.) Maija-Liisa tai TietoEnator20 teknisuontoisia ongelmia pdf-versiohallinnassa, fonteissa, sivukoossa, tuottamisessa ja tulostamisessa21 tiedosto voi sisältää ominaisuuksia, joita lukija ei tunnista22 tulut ulkoa suojattu pdf-tiedostoja, joita ei voi tulostaa eikä kopioida23 tulostuspohjan koko ei vastaa oikeaa24 wordin oleteustulostimen vaikutus pdf-tulostukseen25 vääriä fontteja

8.12.2008

	Vastaus
2 Kysymys	
2.7 Muuta	<ol style="list-style-type: none">1 Abbyy-ocr-ohjelmalla kuva-pdf:stä wordiksi2 Acrobat Prof:n Text toolilla voi korjata suoraan pdf:ään3 Blogger (H. Rajala)4 Ilustratorilla voidaan muokata pdf-tiedostoja5 kerroskuva tehty PSE:llä kuvaksi, josta PDF. Tämä voisi kai olla myös kerroksittainen pdf, jota voisi editoida.6 miten hakukoneet hakevat pdf-tiedostojen sisältä7 ohjeet asiakirjojen tuottamisesta, Arkistokirjaohje8 olisi hyvä julkaista vaikka Kuukauden vinkki –palstaa9 painotyössä sivujen olisi oltava saman suuntaisia10 pdf/a:ta silmällä pitäen kannattaisi pysytellä pdf-tasolla 1.411 PDF:n muokkausta (lisähuomion kirjoittaminen) Acrobat-ohjelmalla12 PDF:n tuottamien vasta lukuaiheessa?13 pdf-arkiston säilyvyys, alkuperäisversion säilyttäminen14 pdf-komenttien käyttö15 pdf-lomakkeiden käyttö voisi olla näppäriä16 pdf-tiedostoihin liittyvän koulutuksen lisääminen17 pdf-tiedostojen muokkaaminen tekstiksi (Abbyy)18 saatava sellaista laatua, jota voidaan käyttää arkistoinnissa paperin asemesta ja jota voidaan jakaa sähköisessä muodossa19 Valiokuntien tietopankki: Wordilla tulleet kirjalliset lausunnot, joita voi muuttaa => lausunnon antajan oikeusturva20 Valiokuntien tietopankkia ei käytetä, koska epäillään tiedon luottamuksellisuuden säilymistä.21 versionumerointi: 1.0 ens. virheineen, 1.1 virheet korjattu, 1.2 virheet korjattu, vastaukset virheineen, 2.0 painovalmis

Eduskunnan valiopäiväasiakirja-asioiden lukumäärä ja pdf-tiedostojen määrä

Tyyppi	Asiakirjoja	pdf-lkm
Hallituksen esitykset - HE Yhteensä	5 313	1 913
Valtioneuvoston selonteot ja tiedonannot - VNS, VNT Yhteensä	75	53
Valtioneuvoston kirjelmät ja selvitykset (EU) - U, E, UTP Yhteensä	2 105	1 707
Valtioneuvoston kirjelmien ja selvitysten (EU) jatkokirjelmät - Y, U, E, UTP Yhteen:	602	602
Pääministerin ilmoitukset - PI Yhteensä	27	0
Valtioneuvoston kirjelmät (ministerien sidonnaisuudet) - VN Yhteensä	66	44
Valiokuntien mietinnöt ja lausunnot - VM, VL Yhteensä	5 954	5 954
Eduskunnan vastaukset ja kirjelmät - EV, EK Yhteensä	2 272	2 272
Aloitteet - LA, TPA, TA, KA, TAA, LTA, RA Yhteensä	36 181	16 333
Kysymykset - VK, KVN, KK Yhteensä	20 326	10 560
Puhemiesneuvoston ehdotukset ja luettelot - PNE, PNL Yhteensä	30	19
Kertomukset - K Yhteensä	136	136
Muut asiakirjat - MINS, M Yhteensä	181	13
Talousarvioasiakirjat pääluokittain ja osastoittain Yhteensä	306	0
	<u>73 574</u>	<u>39 606</u>