

2x13 Vapaat tiedostomuodot - Viikon VALO #65

Vapailla ja avoimilla tiedostomuodoilla pyritään parantamaan tiedon saavutettavuutta noudattamalla avoimia standardeja.



Vapailla tai avoimilla tiedostomuodoilla tarkoitetaan tiedostomuotoja, joiden määrittely on kokonaan julkinen, mahdollisimman standardoitu ja avoin kenen tahansa toteutettavaksi. Vapaita tiedostomuotoja käyttämällä pyritään varmistamaan pitkäaikainen pääsy tietoon millä tahansa laitteistolla ja ohjelmistoalustalla ilman teknisten tai oikeudellisten esteiden luomia rajoituksia. Teknisiä rajoituksia

voi syntyä esimerkiksi silloin, jos tiedostomuoto on jonkin yksittäisen suljetun ohjelman sisäinen tallennusmuoto, jonka toiminnasta ei ole riittävästi dokumentaatiota taikka standardia. Tällöin ei ole mahdollista tai voi olla hyvin vaikeaa luoda muita ohjelmia, jotka osaisivat käsitellä kyseistä tiedostomuotoa. Tällöin tiedoston avaaminen vaatii, että käyttäjällä on käytössään juuri kyseinen ohjelma, ja tiedoston sisältämän tiedon käyttäjä on tavallaan kyseisen ohjelman valmistajan armoilla. Ei ole myöskään varmaa, onko ohjelma saatavissa ja siten tiedosto avattavissa vielä vuosien kuluttua. Vapaiden tiedostomuotojen kanssa sen sijaan dokumentaatio tiedostomuodon toteutuksesta on aina saatavilla ja siten tiedostojen käyttö on mahdollista vielä vuosien päästäkin. Oikeudellisia rajoitteita voi syntyä, jos tiedostomuodon toteutuksessa on käytetty esimerkiksi patentoituja tekniikoita. Näin on esimerkiksi joidenkin ääni- ja videotiedostomuotojen kanssa. Tällöin tiedostoja käsittelevien vaihtoehtoisten ohjelmien toteuttamista rajoittaa teknisten esteiden sijaan lainsäädäntö.

Vapaita tiedostomuotoja ja avoimia standardeja käyttävän on mahdollista vaihtaa käyttämänsä ohjelmat muihin samaa tiedostomuotoa tai standardia tukeviin ohjelmiin. Sen sijaan suljettua tiedostomuotoa käyttävä on helposti sidoksissa kerran tekemäänsä ohjelmistovalintaan.

Vuosittainen Document Freedom Day, eli tiedostojen vapauden päivä, jota vietetään vuonna 2012 maaliskuun 28. päivänä, juhlistaa tiedon vapautta ja saavutettavuutta sekä pyrkii lisäämään tietoisuutta avoimista standardeista.

HTML – www-sivujen tiedostomuoto

Www-sivujen sisällön kuvaava html-kieli on W3C-organisaation ylläpitämä avoin standardi, jolla tehtyjä sivuja pystytään lukemaan useiden eri valmistajien tekemillä selaimilla monenlaisilla laitteilla tietokoneista puhelimiin ja tabletilaitteisiin. Html-tiedostomuoto on erinomainen esimerkki vapaiden tiedostomuotojen ja avoimien standardien merkityksestä ja toimivuudesta. Aiemmin joidenkin www-selainten valmistajat tukivat huonosti standardia ja pyrkivät keksimään selaimiinsa omia laajennoksiaan ja epästandardeja ominaisuuksia, mistä seurasi, että sivustojen tekijät joutuivat näkemään paljon ylimääräistä vaivaa jotta saivat sivut toimimaan kaikilla hieman eri tavoin toimivilla selaimilla. Viime vuosina viimeisimmätkin selainvalmistajat ovat alkaneet noudattaa tarkemmin standardeja ja sivustojen ylläpitäjien tehtävät ovat helpottuneet huomattavasti. Eri selainten ja erilaisten laitteiden käyttäjät voivat nauttia samasta sisällöstä tasavertaisesti.



Hyvin standardeja seuraava www-selain on esimerkiksi vapaan lähdekoodin Firefox-selain.

ODF – toimisto-ohjelmien tiedostomuodot

ODF (Open Document Format) koostuu itse asiassa useammasta OpenDocument-standardin mukaisesta tiedostomuodosta. Näitä ovat seuraavia päätteitä käyttävät tiedostot:

- .odt tekstiasiakirjoille,
- .ods laskentataulukoille,
- .odp esitysgrafiikalle,
- .odg piirtografiikalle,
- .odc kaavioille,
- .odf matemaattisille kaavoille,
- .odb tietokannoille,
- .odi kuvatiedostoille ja
- .odm isäntäasiakirjoille, jotka yhdistelevät näitä kaikkia.



Tiedostomuodoille löytyy tuki useista toimisto-ohjelmistopaketeista. Vapaita toimisto-ohjelmistoja ovat ainakin: LibreOffice, OpenOffice, Calligra Suite, KOffice sekä NeoOffice. Muita ODF-muotoa tukevia toimisto-ohjelmistoja ovat ainakin IBM Lotus Symphony, Corel WordPerfect Office X4, Google Docs ja Zoho Office Suite. Myös Microsoftin Office-paketti tukee uudemmissa versioissaan ODF-tiedostoja ainakin jossain määrin.

ODF-tiedostomuotojen perheen yksittäisiä tiedostumuotoja tukevat myös vapaat ohjelmat Abiword (tekstinkäsittely) ja Gnumeric (taulukkolaskenta). ODF-muodoissa tallennettuja tiedostoja voidaan myös näyttää erilaisissa katseluohjelmissa, kuten Evince ja Okular ja niiden sisältämää informaatiota voidaan tuoda esimerkiksi vektoripiirto-ohjelma Inkscapeen ja sivuntaitto-ohjelma Scribukseen. Tietojen vaihto eri ohjelmien välillä siis onnistuu yhteisellä avointa standardia noudattavalla tiedostomuodolla.

ODF-tiedostomuotojen etuna esimerkiksi Microsoftin Office-paketin vastaaviin tiedostomuotoihin on, että niitä hyvin tukevia ohjelmia löytyy lähes jokaiselle laite- ja käyttöjärjestelmälustalle, kun Microsoftin Office-paketti on tarjolla vain Windows- ja Mac OS X -järjestelmiin. Toisaalta Microsoftin Office-tiedostomuodot ovat saavuttaneet niin vankan aseman, että niiden kanssa kilpaileminen on vaikeaa. Tärkeätä olisi kansalaisten tasa-arvon kannalta kuitenkin saavuttaa tilanne, jossa varsinkin viranomaisten tuottamat asiakirjat olisi mahdollista saada myös vapaissa tiedostomuodoissa, esimerkiksi ODF-tiedostomuodoissa. Hyvänä esimerkkinä avoimien standardien käyttöönotosta valtionhallinnossa on oikeusministeriön vuonna 2006 tekemä päätös ODF-tiedostomuotojen käyttöön siirtymisestä vuodesta 2007 alkaen.

Myös Norjan valtio teki vuonna 2009 päätöksen, että vuoden 2011 alusta lähtien he käyttävät sähköpostin liitteinä lähetettävien tiedostojen tallennusmuotoina joko pdf- tai odf-muotoja.

Tiedostonäytteet:

- (odt) JHS 169 Avoimen lähdekoodin ohjelmien käyttö julkisessa hallinnossa (jhs-suositukset.fi)
- (odp) Lubuntu

PDF – tulostusvalmiille asiakirjoille ja esitysgrafiikalle

PDF (Portable Document Format) on alun perin Adoben kehittämä tulostusvalmiiden tiedostojen esittämiseen soveltuva PostScript-kieleen pohjautuva tiedostomuoto. Tiedostojen päätteenä on .pdf. PDF-tiedostoissa sisältö on aseteltu sivuille ulkoasultaan kiinteään muotoon, joka on helppo tulostaa laitteesta ja ohjelmasta riippumatta juuri sen näköisenä kuin tiedoston tekijä on sen tarkoittanut. PDF-tiedostot sopivat hyvin tiedonvaihtoon silloin, kun vastaanottajan ei ole tarvetta muokata tiedoston sisältöä vaan tiedoston lukeminen, tulostaminen sekä kommentointi ovat riittävät toiminnot. PDF-tiedostoja käytetään muotoiltujen asiakirjojen lisäksi usein esitysgrafiikan tallentamiseen ja esittämiseen.



PDF-tiedostomuoto on hyvin tuettu erilaisilla alustoilla ja pdf-tiedostoille löytyykin Adoben Reader-ohjelman lisäksi useita vaihtoehtoisia lukijaohjelmia, joista PDFreaders.org-sivusto ylläpitää kattavaa luetteloa.

PDF-tiedostojen luominen onnistuu muun muassa LibreOffice-toimistopaketin ohjelmistoilla sekä Scribus-taitto-ohjelmalla.

Tiedostonäytteet:

- Konferenssiesitys

SVG – vektorigrafiikalle

SVG (Scalable Vector Graphics) on kuvauskieli kaksiulotteisen vektorigrafiikan esittämiseen ja se on HTML:n tavoin W3C-organisaation ylläpitämä standardi. SVG-tiedostojen päätteenä on .svg tai gzip-algoritmilla pakattuna .svgz. SVG-tiedosto voi pelkän staattisen kuvan lisäksi kuitenkin sisältää myös muun muassa animaatiota ja JavaScript-kielellä toteutettua toiminnallisuutta, joiden avulla avoimet SVG-tiedostot voivat sisältää monia samanlaisia toimintoja kuin suljetut Flash-tiedostot. SVG liittyykin läheisesti paljon huomiota ja suosiota saaneeseen HTML5:een, johon sitä voi upottaa suoraan. SVG-kuvien käyttö verkkosivuilla on vielä aluillaan, mutta HTML5:n yleistymisen sekä uusimpien Internet Explorerin versioiden tuki SVG:lle todennäköisesti kasvattavat myös SVG:n suosiota.



SVG-muotoisia kuvia voi katsella käytännössä millä tahansa modernilla www-selaimella, kuten Firefox, lukuun ottamatta joitain mobiiliselaimien malleja. SVG-kuvia voi käyttää esimerkiksi LibreOffice Writer-tekstinkäsittelyohjelmassa ja saman toimistopaketin esitystyökalussa

Impressissä. Myös taitto-ohjelma Scribus osaa hyödyntää SVG-kuvia. Kuvien luomiseen sopii vaikka Inkscape-piirtotyökalu. Muun muassa Adobe Illustrator sekä CorelDraw tukevat myös SVG-standardia.

Wikipedia on käyttänyt jo pitkään SVG-tiedostoja kaavioiden, karttojen sekä muiden piirtokuvien tallettamiseen ensisijaisena tiedostomuotona. Verkkotietosanakirjan sivuilla SVG-kuvat on tosin jouduttu toistaiseksi esittämään PNG-muotoon muunnettuina joidenkin yhä suosittujen vanhojen selainten puuttuvan SVG-tuen takia.

Wikipedian lisäksi vapaita SVG-muotoisia tiedostoja voi ladata itselleen Open Clip Art Library -palvelun sivuilta.

Tiedostonäytteet:

- Kuva Open Clip Art Libraryssa (Peace_and_love.svg)
- Kuva Wikimedia Commonsissa (2007-02-20_time_zones.svg)

JPEG – valokuville

JPEG (Joint Photographic Experts Group) on erityisesti valokuvien tallentamiseen tarkoitettu häviöllistä pakkausta käyttävä kuvantallennusmuoto. Häviöllisyys tarkoittaa sitä, että kuvan tallennuskoon pienentämiseksi kuvan pikseleiden sisältämä informaatio voi muuttua jonkin verran, eli niiden värisävyjä muutetaan hieman. Tyypillisesti valokuvissa tällä häviöllä ei ole suurta merkitystä ainakaan ellei pakkausalgoritmia säädetä liian aggressiiviseksi. Paljon jyrkkiä reunoja sisältäviin kuviin, eli piirtokuvaan ja teksteihin, JPEG ei kuitenkaan sovellu vaan tekee kuvan suttuisen näköiseksi. JPEG-tiedostojen päätteenä on .jpg, .jpe tai .jpeg.



JPEG-muotoisten rasterikuvien avaamiseen, käsittelyyn ja tallentamiseen sopivat vapaista ohjelmista vaikka GIMP, Pinta sekä Krita. JPEG-kuvia voi sisällyttää myös muun muassa LibreOfficella, Scribuksella ja Inkscapella tehtäviin tiedostoihin.

Tiedostonäytteet:

- Pingviini Wikimedia Commonsissa (Falkland_Islands_Penguins_49.jpg)
- Charlie Chaplin Wikimedia Commonsissa (Chaplin2.jpg)

PNG – piirretyille rasterikuville

PNG (Portable Network Graphics) on häviötön kuvantallennusmuoto, joka sopii JPEG-tiedostoja paremmin juuri piirroskuville ja paljon tekstiä tai muita jyrkkiä reunoja sisältäviin kuviin. PNG on häviötön, eli kuva ei tallennuksen yhteydessä muutu millään tavalla. PNG-tiedostojen päätte on .png. PNG tukee 16-bittistä läpinäkyvyyttä, eli kuvan pikselit voivat olla kokonaan tai osittain läpikuultavia tai täysin peittäviä. Tässä



suhteessa PNG on huomattavasti parempi kuin vanhempi GIF-tiedostomuoto, joka mahdollisti vain joko kokonaan läpinäkyvät tai täysin peittävät osat. PNG-tiedostot tukevat myös miljoonia värisävyjä verrattuna GIF-tiedostojen enintään 256 väriin. PNG-tiedostot ovatkin www-käytössä korvanneet GIF-kuvat lähes täysin. GIF-kuvien ainoaksi eduksi on jäänyt mahdollisuus tehdä animoituja kuvia.

JPEG-tiedostojen tavoin PNG-kuvien avaamiseen, muuntamiseen ja tallentamiseen voi käyttää GIMP-, Pinta- ja Krita-ohjelmia. PNG-kuvia voi myös sisällyttää muiden ohjelmien luomiin tiedostoihin aivan samoin kuin JPEG-tiedostojakin.

Tiedostonäytteet:

- Wikimedia Commonsissa (Donate-icon.png)

EPUB – sähköisille kirjoille

EPUB on vapaa tiedostomuoto sähköisten kirjojen tallentamiseen. EPUB-tiedostojen päätte on .epub. EPUB hyödyntää esimerkiksi muita vapaita tiedostomuotoja sillä käytännössä EPUB-tiedosto koostuu XHTML-muotoisesta sisällöstä, SVG-, JPEG-, PNG- ja GIF-muotoisista kuvista, CSS-tyylimäärittelyistä sekä metatiedon sisältävästä XML-tiedostosta. Nämä tiedostot on pakattu yhteen ZIP-tiedostoon, jonka nimi on muutettu .epub-päätteiseksi.



Suurin osa sähköisten kirjojen lukijalaitteista ja -ohjelmistoista osaa lukea ja näyttää laitteisto- ja ohjelmistoriippumattomaksi suunniteltuja EPUB-tiedostoja juuri tiedostomuodon vapauden ja avoimuuden vuoksi. Vapaista ohjelmista EPUB-tiedostoja näyttävät ainakin FBReader sekä Calibre. EPUB-tiedostoja voi puolestaan luoda joko suoraan tekemällä sisältö tekstieditorilla ja kuvankäsittelyohjelmilla sekä pakkaamalla ZIP-tiedostoksi taikka käyttämällä ohjelmaa nimeltä Sigil.

EPUB-muotoisia kirjoja löytyy muun muassa vapaita sähkökirjoja jakavasta Project Gutenberg -palvelusta.

Tiedostonäytteet:

- The Memoirs of Sherlock Holmes by Sir Arthur Conan Doyle

OGG/Vorbis – äänitiedostoille

Ogg on avoimen standardin multimediasäilötiedostomuoto. Multimediasäilöt ovat tiedostoja, joiden sisälle säilötään itse media, ääni, video tai molemmat, jollain pakkausmenetelmällä. Yleisimmin OGG-säilön kanssa käytetty äänenpakkausmenetelmä on avoimeen lähdekoodin perustuva häviöllinen Vorbis. Yhdistelmästä käytetään nimitystä Ogg Vorbis ja tällaisten tiedostojen päätteeksi on yleensä .ogg tai .oga (ogg audio). Vorbis on MP3-pakkausten tavoin häviöllinen, eli Vorbis-koodekilla pakkaaminen hukkaa alkuperäisestä äänitiedosta osan, esimerkiksi ihmiskorvan kuuloalueen ulkopuolelle jääviä osuuksia. Ogg Vorbis on vaihtoehto MP3-tiedostomuodolle, jonka ongelmana ovat patenttien luomat rajoitteet. Esimerkiksi MP3-soittimien valmistajat maksavat laitekohtaisia lisenssimaksuja Fraunhofer-instituutille.

Ogg-säilön lisäksi Vorbis-ääntä voidaan käyttää myös muiden säilöjen yhteydessä, esimerkiksi ääniraitana DivX-pakatun videon kanssa.

Ogg Vorbis -tiedostoille on tuki useissa multimediaohjelmissa, esimerkiksi Winampissa ja avoimen lähdekoodin VLC-soittimessa. Esimerkiksi Android-laitteet tukevat Ogg Vorbis -ääntä. Myös Firefox- ja Chrome-selaimet tukevat Ogg Vorbis -ääntä suoraan tiedostoina tai html5:n audio-tagin kautta avattuina. Wikipedian ääninäytteet ovat myös Ogg Vorbis -muodossa.

Ogg Vorbis -tiedostomuotoon voi tutustua tarkemmin osoitteessa Vorbis.com.

Tiedostonäytteet:

- Jabberwocky.ogg (Wikimedia Commons)
- The Adventures of Sherlock Holmes: 01 - A Scandal in Bohemia by Sir Arthur Conan Doyle (LibriVox)

OGG/Theora – videotiedostoille

Ogg-säilöön voidaan äänen lisäksi säilöä myös videota. Tyypillisimmin Ogg-säilössä tallennetaan Theora-muodossa. Theora on Ogg-projektin yhteydessä kehitetty avoin ja patenteista riippumaton videokoodekki. Se on tarkoitettu vaihtoehdoksi suljetulle MPEG-4-koodekille, jonka ongelmana ovat patentit, kuten mp3-tiedostojenkin yhteydessä. Ogg Theora -tiedostojen päätteenä on yleensä .ogg tai .ogv (ogg video) ja ääniraitana käytetään tavallisesti Vorbis-ääntä.

Ogg Vorbiksen tavoin Firefox- ja Chrome-selaimet tukevat suoraan myös Ogg Theora -tiedostoja. Samoin vapaan lähdekoodin VLC-mediasoitin toistaa Ogg Theora -videot mukisematta. Vapaana tiedostomuotona Ogg Theora onkin myös Wikipedia-projektin käyttämä tiedostomuoto tietosanakirjan videonäytteille.

Tiedostonäytteet:

- J.F. Kennedy "Ich bin ein Berliner" -puhe
- Big buck bunny

FLAC – korkeatasoiselle äänelle

FLAC (Free Lossless Audio Codec) on pakattu, mutta häviötön, sekä patenteista ja lisenssimaksuista vapaa tiedostomuoto digitaaliselle äänelle. FLAC-muotoon tallennettaessa äänestä ei siis poisteta lainkaan informaatiota vaan pakkaaminen säilyttää äänen täsmälleen alkuperäisenä. Tehokkaasti pakattuna, mutta samalla häviöttömänä, FLAC on erinomainen tiedostomuoto, kun on tarkoitus editoida äänitiedostoa mahdollisimman korkealla laadulla. Tuki FLAC-tiedostoille löytyykin monista ohjelmistoista ja



laitteistoista. Avoimen lähdekoodin Audacity sopii hyvin FLAC-tiedostojen käsittelyyn. VLC-mediasoitin soittaa erinomaisesti myös FLAC-tiedostot.

Tiedostonäytteet:

- Jono Bacon: Free software song (25,3 Mt)

Document Freedom Day 2012

- <http://documentfreedom.org/>

Toimii seuraavilla alustoilla

- Vapaiden ja avoimien tiedostomuotojen tarkoitus on, että niiden on mahdollista toimia kaikilla mahdollisilla alustoilla.

VALO-CD

- VALO-CD-kokoelmasta löytyy useita ohjelmia, joilla on mahdollista käyttää ja muokata vapaissa tiedostomuodoissa olevaa sisältöä.

Teksti: Pesasa

Lisenssi: CC-BY-SA