



Kansallinen digitaalinen kirjasto

## KDK/PAS: Mitä kuvien luetteloinnissa tulee ottaa huomioon pitkäaikaissäilytyksen näkökulmasta?

Valtakunnalliset kuva-arkistopäivät  
22.11.2011  
Vesa Hongisto

### Esityksen sisältö



Kansallinen digitaalinen kirjasto

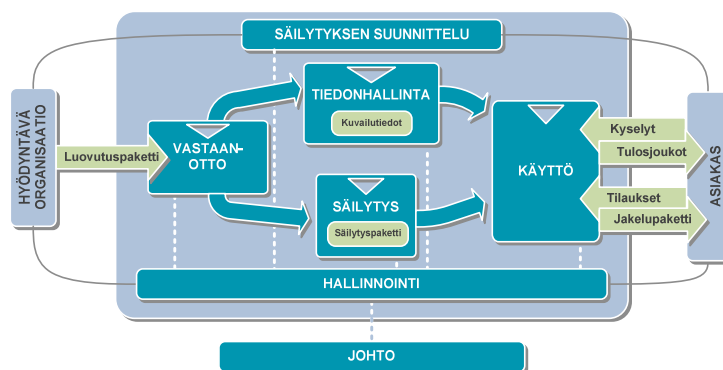
- › Mitä on digitaalinen pitkäaikaissäilytys
- › Pitkäaikaissäilytyksen kokonaisuus OAIS-mallin mukaan
- › Aineiston siirto pitkäaikaissäilytykseen, aineiston paketointi ja standardit
- › Pakolliset metatietoelementit
- › Suositeltavat metatietoelementit
- › Tiedostomuotojen pysyvyyden arviointi
- › Säilytuskelpoiset tiedostomuodot
- › Siirtokelpoiset tiedostomuodot



# Mitä digitaalinen pitkäaikaissäilyttäminen on

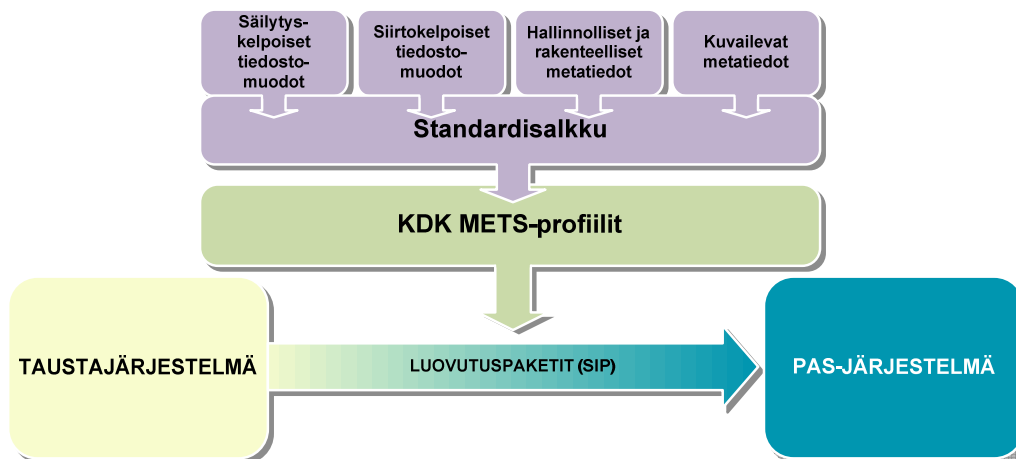
- Pitkäaikaissäilytys tarkoittaa digitaalisen informaation luotettavaa säilyttämistä useiden kymmenien tai jopa satojen vuosien ajan.
  - › Laitteet, ohjelmistot ja tiedostomuodot vanhenevat, mutta informaation täytyy säilyä.
  - › Digitaalisessa muodossa olevan aineiston pitkäaikaissäilytys ei ole yksinomaan tekninen haaste, vaan sen hoitaminen menestyksellisesti vaatii myös toiminnallisia, tiedollisia, taidollisia, taloudellisia ja juridisia valmiuksia.
- Digitaalisten kulttuuriperintöaineistojen säilyttämisestä vastaavien organisaatioiden käytössä olevat tietojärjestelmät eivät sisällä pitkäaikaissäilytyksen edellyttämiä toiminnallisuuksia (mm. eheyden tarkkailu, migraatiot, virkistäminen). Vastaavasti organisaatioilta puuttuvat myös digitaaliseen pitkäaikaissäilytykseen keskittyvät prosessit. Yhteisellä pitkäaikaissäilytysratkaisulla ratkaistaan useiden organisaatioiden tarpeet ja mahdollistetaan aineistojen tuleva hyödyntäminen.

## Pitkäaikaissäilytyksen kokonaisuus OAS-mallin mukaan



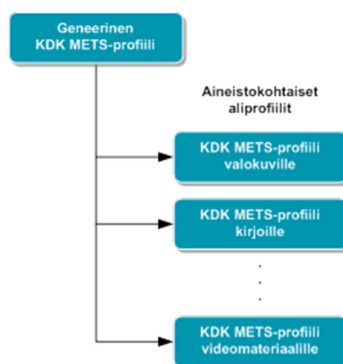
- › Open Archival Information System (OAIS) -viitemalli on keskeisin kansainvälinen pitkäaikaissäilytyksen standardi. Se määrittelee mm. pitkäaikaissäilytyksen keskeiset käsitteet, toiminnot sekä järjestelmissä käsiteltävät tieto-objektit. Järjestelmien tekniseen toteutukseen malli sen sijaan ei ota kantaa.
- › OAIS-viitemallin mukaisia toimintoja ovat vastaanotto, säilytys, tiedonhallinta, käyttö, hallinnointi ja säilytyksen suunnittelu.

# Standardisalkun, METS-profiilin ja luovutuspaketin suhde



- › KDK:ssa käytettävät standardit, tiedostomuodot sekä pakolliset ja suositeltavat metatiedot määrittellään standardisalkussa
- › Standardisalkun mukaisesti aineisto siirretään taustajärjestelmästä PAS-järjestelmään käyttäen METS vaihtomuotoformaattia (Metadata Encoding and Transfer Standard)

# Geneerinen KDK METS-profiili ja mahdollisia aliprofiileja



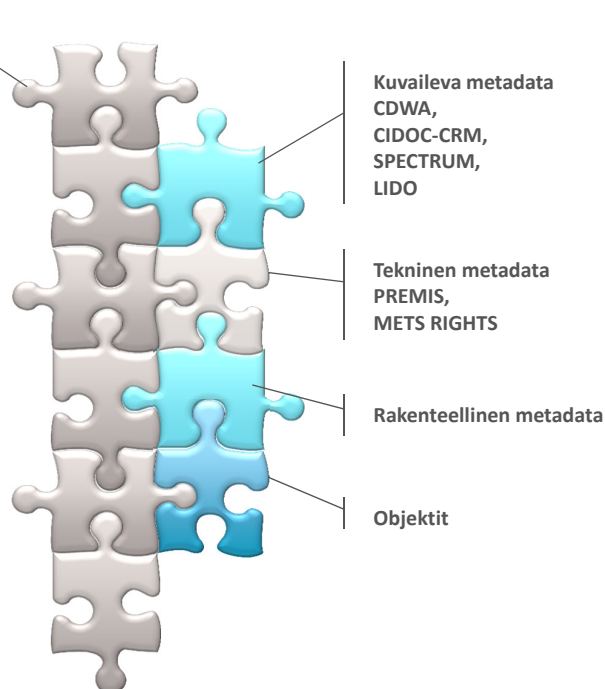
- › KDK:ssa on määritelty geneerinen KDK METS-profiili, joka on yhteinen kaikille luovutuspaketeille riippumatta siitä mitä tai millaista aineistoa siirretään. Profiili kuvaa kaikille luovutuspaketeille yhteiset piirteet ja kuinka ne kirjataan luovutuspakettiin METSiä hyödyntäen.
- › Myöhemmin voidaan erityyppisille aineistoille määrittellä aliprofiileja, jotka määrittelevät tarkemmin kyseiseen aineistoon liittyviä piirteitä.



- › Aineisto tallennetaan PAS:iin luovutuspaketteina. Luovutuspaketin tekninen rakenne määrittelee muun muassa sen, mitä (meta)tietoja varsinaisen säilytettävän aineiston lisäksi luovutuspaketissa on oltava.
- › Näiden tietojen avulla varmistetaan esimerkiksi se, että luovutuspaketti on siirtynyt PAS-järjestelmään muuttumattomana, varmistutaan paketin lähettäjistä ja tarkistetaan että paketti täyttää hankkeen tekniset vaatimukset.



Luovutuspaketti  
METS



# Pakolliset metatietoelementit (1)

Elementti	Määrittäminen
Tekniset piirteet	<p>Pakolliset tekniset piirteet sisältävä seuraavat elementit</p> <ul style="list-style-type: none"><li>- Tiedostomuoto ja sen versio</li><li>- Aineiston syntyhistoria</li><li>- Digitoidut aineistot: millä laitteistolla digitoitu, millä ohjelmistolla ja sen versiolla digitoitu, millä ohjelmistolla ja sen versiolla käsitelty.</li><li>- Alkujaan digitaaliset aineistot: millä ohjelmistolla aineisto on muodostettu, ohjelmiston versio.</li></ul> <p>Tiedostomuodon ja sen version tulee sisältyä luovutuspakettiin, koska työkalut, joilla tiedostomuotoa voidaan selvittää kuten JHOVE 2, ovat puutteellisia. Työkalut eivät kata kaikkia KDK:n kannalta relevantteja tiedostomuotoja, eivätkä ne välttämättä tunnista tuttuakaan tiedostomuotoa jos tiedoston otsakkeessa on virhe.</p> <p>Aineiston syntyhistorian tulee sisältyä luovutuspakettiin, koska aineiston todistusvoimaisuus syntyy osittain aineiston syntyhistorian myötä. Ilman syntyhistoriaa tiedetään ainoastaan aineiston historia siitä lähtien kun se on vastaanotettu PAS-järjestelmään.</p>
Eheystieto ja sen algoritmi	<p>Jokaiselle tiedostolle on laskettava eheystieto mahdollisimman aikaisessa vaiheessa jotta tiedoston muuttumattomuus voidaan tarkistaa. Eheystieto (esimerkiksi MD5-tarkistussumma) on tallennettava luovutus- tai jakelupaketin ko. tiedostoa koskeviin teknisiin metatietoihin. Siirron jälkeen vastaanottaja voi tarkistaa tiedostojen muuttumattomuuden tarkistussummaa hyödyntäen. Myös luovutus- ja jakelupaketille on laskettava eheystieto, jotta voidaan varmistua METS-dokumentin eheydestä.</p>
Kuvailuformaatti	<p>Kuvailuformaatin määrittely on välttämätöntä, jotta tiedetään miten metatietotietuetta pitää tulkitella. Jos käytetään organisaation omaa kuvailuformaattia, siitä on lähetettävä tiedot aineiston mukana. Kuvailutiedot on kuitenkin aina toimitettava myös KDK-standardisalkussa määritellyssä kuvailuformaattissa.</p> <p>Kuvailu- ja luettelointisääntöjen tarkoituksena on aineiston käytettävyyden parantaminen hakemistojen tietosisällön kehittämällä sekä hakemistojen tietosisällön ja tietojen esittämistavan yhdenmukaistaminen.</p>

# Pakolliset metatietoelementit (2)

Elementti	Määrittäminen
Tunniste paketille	<p>Luovutuspaketilla on oltava myös taustajärjestelmän tai luovutuspaketin rakentaneen sovelluksen luoma järjestelmäkohtainen uniikki tunniste. Organisaation nimen olisi suositeltavaa sisältyä tunnisteeseen esimerkiksi ao. organisaation ISIL-tunnisteena [ISIL].</p> <p>Tunniste mahdollistaa ao. paketin yksilöimisen. Ongelmatapauksissa paketti tulee voida jäljittää aineiston lähettäjäorganisaatioon, missä kyseinen luovutuspaketti tulee voida tarvittaessa rakentaa uudelleen.</p>
Paketin luonti- tai muokausaika	<p>Luovutus- tai jakelupaketin luontiaika on aina ilmoitettava paketin metatiedoissa.</p>
Tiedostojen ja metatietotietueiden luontiaika	<p>Luovutuspaketissa on oltava siihen sisältyvien tiedostojen luonti- ja/tai muokausajankohdat sekä ajankohdat jolloin paketin sisältämiä metatietotietueita on viimeksi muokattu.</p>
Rakennekartta	<p>METS-standardin mukaan rakennekartta on pakollinen tieto. Vaikka luovutus- tai jakelupaketti sisältäisi vain yhden tiedoston, on dokumentin rakenne silti kuvattava. Tällöin rakennekartta kertoo sen, että paketin sisältämä dokumentti koostuu yhdestä tiedostosta, sekä sen, mistä tämä tiedosto löytyy.</p> <p>Jos dokumentti koostuu useista tiedostoista, rakennekartan avulla kerrotaan missä järjestyksessä ne (esim. digitoidun kirjan sivukuvat) on esitettävä. Näin aineistosta saadaan ymmärrettävä ilman sisällöllistä analyysia. Rakennekartan avulla voidaan myös pitää samassa paketissa olevat digitaalisen objektin eri manifestaatiot toisistaan erillään.</p>
Käyttö- ja näyttörajoitukset	<p>Jos aineistoon liittyy käyttö- tai näyttörajoituksia, ne on aina esitettävä luovutus- tai jakelupaketissa. Jos samassa paketissa on useita objekteja joiden käytettävyydessä on eroja, rajoitukset tulee kohdentaa niihin objekteihin, joita ne koskevat.</p>

# Suosittelavat metatietoelementit (1)

Elementti	Määrittäminen
Tunniste tietueille	Taustajärjestelmissä päivitetään metatietotietueita, jotka pitää viedä PAS-järjestelmään. Vastaavasti PAS-järjestelmässä päivitetään migraatioita tehtäessä hallinnollista metatietoa, joka tulee lähettää uusien dokumenttien mukana jakelupakettina taustajärjestelmiin. Tietueilla on yleensä kaikissa tietojärjestelmissä sisäiset tunnisteet, mutta jos ne eivät ole KDK-tasolla uniikkeja, tulee ne muuntaa ennen siirtoa KDK-ympäristöön sopiviksi.
Tallennusmedia- ja tapahtumahistoria	Tapahtumahistoria koskee niin aineiston digitointihistoriaa (kuka on digitoinut ja missä) kuin myös aineistolle suoritettuja ylläpitotoimenpiteitä kuten migraatioita. Tapahtumahistorian olennaisimpia tietoja ovat: Koska dokumenttia on muutettu, kuka tai mikä on vastannut muutoksesta ja mitä dokumentin tietosisällölle tai ulkoasulle on tällöin tapahtunut. Tyypillisin esimerkki tietosisältöön vaikuttavasta tapahtumasta on tiedoston tiedostomuutos. Tallennusmedia- ja tapahtumahistorian avulla voidaan jäljittää kaikki sellaiset aineistolle tehdyt muutokset, joilla on voinut olla vaikutusta dokumenttien ulkoasuun tai intellektuaaliseen sisältöön. Jos hyödyntävä organisaatio ei välitä näitä tietoja, muutoksia voidaan seurata vain luovutuspaketin PAS-järjestelmään siirtämisestä alkaen.
Kuvailukäytäntö ja kuvailun taso	Kuvailukäytännön ja kuvailun tason osoittaminen helpottaa luovutus- tai jakelupakettiin tallennetun metatiedon käsittelyä ja hyödyntämistä. Kuvailukäytänteellä tarkoitetaan tässä esimerkiksi organisaation omaa aineiston kuvailuohjetta tai kansainvälistä kuvailusäännöstöä kuten RDA:ta tai AACR2:sta.

# Suosittelavat metatietoelementit (2)

Elementti	Määrittäminen
Tekniset piirteet	Pakollisten teknisten piirteiden lisäksi on olemassa aineistokohtaisia teknisiä piirteitä, joita ei esiinny muilla aineistoilla mutta joiden kuvaaminen helpottaa tämänkaltaisten objektien tulkintaa tai on sen kannalta oleellista (esim. audiovisuaaliseen aineistoon liittyvät tekniset piirteet, kuten näytteenottotaajuus). Suositeltavilla teknisillä piirteillä tarkoitetaan yleisesti sellaista teknistä tietoa, joka parantaa tiedon käytettävyyttä ja säilytettävyyttä. Esimerkiksi tiedoston koko voi olla tarpeellinen tekninen metatieto (jos tiedostot ovat esim. poikkeuksellisen isoja).
Tunniste tiedostolle	Jos hyödyntävä organisaatio käyttää omassa järjestelmässään tunnisteita tiedostoille, niitä voidaan hyödyntää PAS-järjestelmässä. Tunnisteet mahdollistavat yksittäisen tiedoston nopean etsimisen luovutuspaketista tai säilytyspaketista, jos tiedosto tallennetaan sellaisenaan. Lisäksi tunnisteet helpottavat a.o. tiedoston päivittämistä taustajärjestelmässä mahdollisen PAS-järjestelmässä tehdyn migraation jälkeen, sekä linkkien luontia alkuperäisen ja migroitujen tiedostojen välille. Jos tiedostojen tunnisteet puuttuvat, ne generoidaan PAS-järjestelmässä. Migroitujen tiedostojen päivittäminen taustajärjestelmiin tapahtuu dokumentin tunnisteiden avulla.
Tunniste dokumenttien manifestaatioille	Yhdessä luovutus- tai jakelupaketissa voi olla sama dokumentti useina manifestaatioina, esimerkiksi digitaalinen kirja PDF- ja OOXML-muodossa. Tällöin jokaisella manifestaatiolla tulee olla oma tunniste, eli esimerkkitapauksessa ISBN. Eri manifestaatiot tulisi osoittaa teostason metatiedoissa, joihin tulee sisällyttää teoksen tunniste, eli esimerkiksi tekstimuotoisen teoksen tunniste ISTC tai audiovisuaalisen aineiston ISAN sekä tunnisteisiin pohjautuvat linkit manifestaatiotason metatietoihin.

# Tiedostomuotojen pysyvyyden arviointi

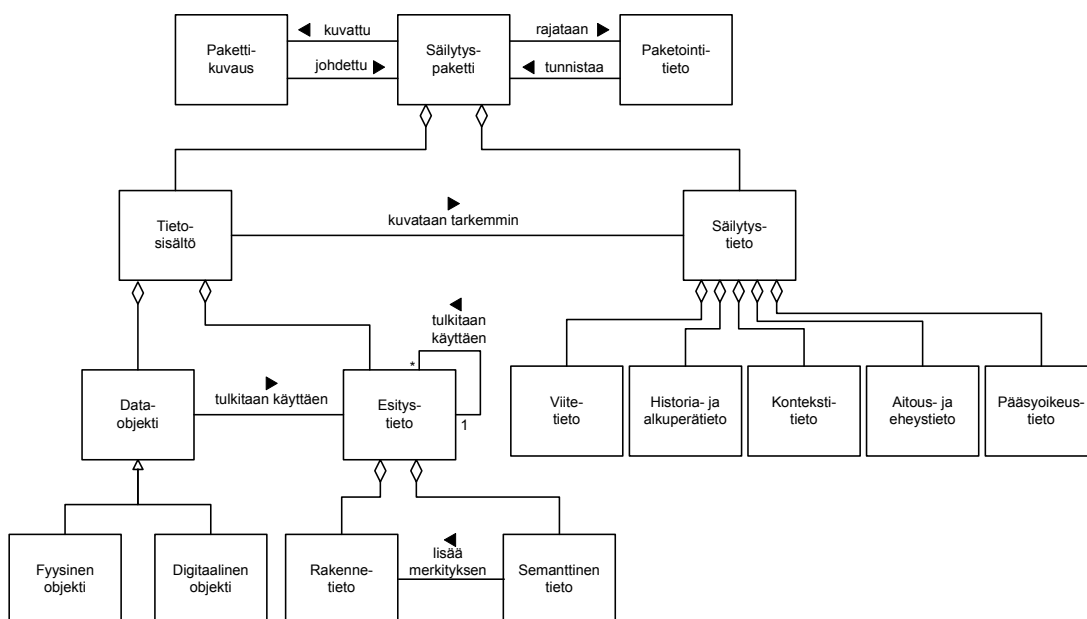
Kriteeri	Selitys
Avoimuus	Kuinka helppoa tiedostomuodosta on saada tietoja?
Käyttö PAS-standardina	Missä määrin tiedostomuoto on muodollisesti hyväksytty pitkäaikaissäilytyksen välineeksi kansalliskirjastoissa, kansallisarkistoissa ja muissa alan laitoksissa?
Vakaus / yhteensopivuus	(a) Missä määrin tiedostomuoto on eteen- ja taaksepäin yhteensopiva? (b) Missä määrin tiedostomuoto on suojattu tiedoston korruptoitumista vastaan? (c) Kuinka usein tiedostomuodosta julkaistaan korvaavia versioita?
Riippuvuudet / yhteentoimivuus	Missä määrin tiedostomuoto on sidottu esimerkiksi tiettyyn laitteistoon tai ohjelmistoon?
Standardisuus	Missä määrin tiedostomuoto on käynyt läpi perusteellisen standardointiprosessin?

# Säilytyskelpoiset tiedostomuodot

Tiedostotyyppi	Tiedostomuodot	Versiot
Teksti	<ul style="list-style-type: none"><li>- Electronic Publications (EPUB)</li><li>- Extensible Hypertext Markup Language (XHTML)</li><li>- Extensible Markup Language (XML)</li><li>- Hypertext Markup Language (HTML)</li><li>- Open Document Format (ODF)</li><li>- PDF for long-term preservation: PDF-Archive (PDF/A)</li><li>- Tekstiedosto (Plain text)</li></ul>	EPUB versio 2.0.1 XHTML versio 1.0; XHTML versio 1.1 XML versio 1.0 HTML versio 4.01 .odt versio 1.0; .ods versio 1.0; .odb versio 1.0; .odp versio 1.0; .odg versio 1.0 PDF/A-1a; PDF/A-1b; PDF/A-2 Tekstiedosto merkistöllä ISO 8859-1; Tekstiedosto merkistöllä UNICODE (UTF-8-, UTF-16- tai UTF-32-koodattuna)
Ääni	<ul style="list-style-type: none"><li>- Audio Interchange File Format (AIFF), PCM-koodattu</li><li>- Broadcast Wave Format (BWF)</li><li>- Free Lossless Audio Codec (FLAC)</li><li>- MPEG-4 AAC – Advanced Audio Coding (AAC)</li><li>- Waveform Audio Format (WAV)</li></ul>	AIFF BWF Version 2 FLAC MPEG-4 Part 3 WAV
Elävä kuva	<ul style="list-style-type: none"><li>- Motion JPEG 2000</li></ul>	Motion JPEG 2000 Part 3
Kuva	<ul style="list-style-type: none"><li>- Joint Photographic Experts Group (JPEG)</li><li>- Joint Photographic Experts Group JPEG 2000 (JP2)</li><li>- Tagged Image File Format (TIFF)</li></ul>	JPEG part 1 JPEG 2000 part 1 TIFF 6.0
Verkoarkisto	<ul style="list-style-type: none"><li>- Web ARChive Format (WARC)</li></ul>	WARC

Tiedostotyyppi	Tiedostomuodot	Versiot
Teksti	- Microsoft Office Suite  - Portable Document Format (PDF)	Word 97 (8.0); Word 98 (8.5); Word 2000 (9.0); Word 2002 (10.0); Word 2003 (11.0); Word 2007 (12.0); Word 2010 (14.0); Excel 97 (8.0); Excel 99 (9.0); Excel 2001 (10.0); Excel 2003 (11.0); Excel 2007 (12.0); Excel 2010 (14.0); Powerpoint 1997 (8.0); Powerpoint 1999 (9.0); Powerpoint 2001 (10.0) PDF 1.2; PDF 1.3; PDF 1.4; PDF 1.5; PDF 1.6; PDF 1.7
Ääni	- Audio Interchange File Format (AIFF-C) - Moving Pictures Expert Group (MPEG) MPEG-1 layer-3, MPEG-2 layer-3 (MP3) - Windows Media Audio (WMA)	AIFF-C MPEG-1 layer-3; MPEG-2 layer-3  WMA 9
Elävä kuva	- Digital Video (DV) - Moving Pictures Expert Group (MPEG) - Windows Media Video (WMV)	DV MPEG-1; MPEG-2; MPEG-4 WMV 9
Kuva	- Digital Negative (DNG) - Encapsulated postscript (EPS) - Graphics interchange format (GIF) - Portable network graphics (PNG)	DNG 1.3.0.0 EPS versio 3.0 GIF 1987a; GIF 1989a PNG

## Säilytyspaketin looginen rakenne







Kansallinen digitaalinen kirjasto

---

Kiitos!

[www.kdk.fi](http://www.kdk.fi)