

METS ja digitaaliset aineistot

Karo Salminen, Kansalliskirjasto

Digitoinnin työpaja: Videot

5.3.2010

Alkuun yksinkertaistus

- Aineiston tarjoaminen pääpiirteissään:
 - Aineistoa digitoidaan, tuotetaan tai vastaanotetaan
 - Sille etsitään sopiva käyttöympäristö eli järjestelmä
 - Palvelua markkinoidaan ja asiakkaat löytävät tiensä aineiston ääreen
- Tarinan ei pitäisi loppua tähän...

Aineistoa on totuttu pimittämään

- Tai ainakin tarjoamaan rajatussa käyttöympäristössä...
- Avoimet rajapinnat ja tiedon liikkuminen ovat nykyaikaa
 - Valtion laitoksilla ja virastoilla on paljon laadukasta aineistoa ”piirongin laatikoissa”
 - Kova paine tietovarantojen julkistamiseen
 - Avoin data ja sen tarjoamat mahdollisuudet

Järjestelmä on vain väline

- Aineisto on ensisijaista, sitä käyttävä järjestelmä toissijainen
- Järjestelmä voidaan vaihtaa, aineistoa ei
- Järjestelmissä olevat tietomassat on pystyttävä siirtämään
 - Tähän ei ole ollut yleensä tarvetta (järjestelmien vaihto poikkeuksena)
 - Viimeistään KDK-hankkeessa huomattiin, että aineiston siirto ja liikuteltavuus pitäisi ottaa huomioon, jos aineistoa halutaan säilyttää pitkäaikaisesti

Tietojen kokoaminen yhteen

- Järjestelmien sisäinen tietorakenne ja toteutus voi olla minkälainen vain
- Tietolähteet voivat vaihdella
 - Tieto voi olla siroteltuina eri järjestelmiin, kunhan se on olemassa ja saatavilla
- Kun aineisto halutaan siirtää järjestelmästä toiseen, kaikki tarvittava tieto kerätään yhteen ja paketoidaan esim. METS:iin

METS on siirtämistä varten

- Tietomassojen siirtämiseksi yksi ratkaisu on METS (Metadata Encoding and Transmission Standard)
- METS
 - on säiliö tai paketti digitaaliselle aineistolle
 - nivoo yhteen metadatan ja tiedostot suhteellisen riippumattomaksi kokonaisuudeksi
 - Kuvaileva, tekninen ja hallinnollinen metadata
 - mahdollistaa aineiston rakenteen kuvaamisen
 - Jako voi olla fyysinen tai looginen
 - DVD:n kappaleet (chapters)
 - Useasta kappaleesta koostuva albumi

Esimerkki: lyhytelokuva

- Siirrettävän kokonaisuuden koostumus voi yksinkertaisimmillaan olla seuraavanlainen:
 - Nimi
 - Julkaisija
 - Päiväys
 - Videotiedosto(t)

Toinen esimerkki: digitoitu haastatteluvideo

- Nimi, julkaisija, päiväys, kieli, kesto
- Haastattelija, haastateltavat, tuottaja, kuvaaja
- Yleiskuvaus sisällöstä
- Video jaettu osiin
 - videoleikkeiden kuvaus
 - videoleikkeiden aikaleimat
- Asiasanoitus
- Tekniset tiedot alkuperäisestä videosta
 - esim. betacam-nauha
- Tiedot digitoinnista
 - esim. käytetty laitteisto, asetukset ja ohjelmistot, näytteenottotaajuus, digitoinnin suorittaja
- Tekniset tiedot digitaalisista videoleikkeistä
 - Esim. tiedostomuoto, koodekki (pakkausmuoto), resoluutio, tiedonsiirtonopeus
- Tekijänoikeustiedot

Tiedot yhtenäiseen muotoon

- Jatkokäyttöä hyödyttää, jos tiedot on siirtovaiheessa jäsennetty johonkin yleisesti käytettyyn muotoon
- Esimerkkejä metadataformaateista:
 - Dublin Core
 - PB Core (Public Broadcasting Metadata Dictionary)
 - MODS/MARCXML
- Kannattaa silti huomata, että metadataformaatejakin voi käyttää vaihtelevin tavoin

Kuriositeetin vuoksi

<http://www.jiscdigitalmedia.ac.uk/images/mets.xml>

<http://lcweb2.loc.gov/diglib/ihas/loc.natlib.ihas.200031106/mets.xml>