

Sähköisen arkiston tekninen varmentaminen ja migraatio

Osmo Palonen
Mikkelin Ammattikorkeakoulu

Sisältö

- Arkiston rakenteen standardit ja Sähke normi
- Käyttäjähallinnan ratkaisut
- Teknisen varmentamisen perusratkaisut
- Formaattien ja medioiden hallinta
- Emulointi
- Migraatio
- Migraation edellytykset
- Migraation toteutusmallit
- 10 kohdan ohjelma sähköisen arkiston toimijalle

OAIS-malli

- OAIS = Open Archival Information System
- Luotu alunperin USA:n avaruushallinnon (NASA) tarpeisiin
- Arkistoyksikkö (content) tallennetaan muuttumattomana yhdessä kuvailevan tiedon kanssa
- Kansainvälisesti yleisesti hyväksytty

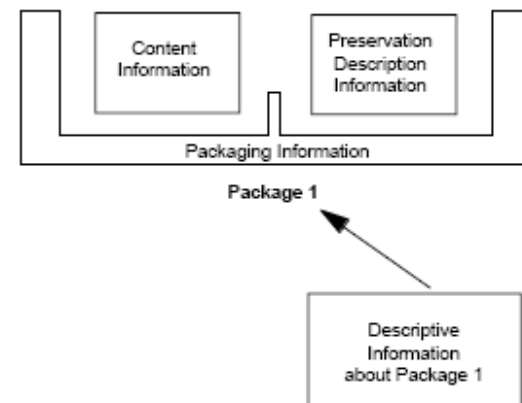


Figure 2-3. Information Package Concepts and Relationships

- <ftp://nssdcftp.gsfc.nasa.gov/standards/nost/isoas/int08/CCSDS-650.0-W-5.pdf>

OAIS-malli

- SIP = Submission information package = siirtotiedostopaketti
- AIP = Archival Information Package = arkistopaketti
- DIP = Dissemination Information Package = jakelupaketti

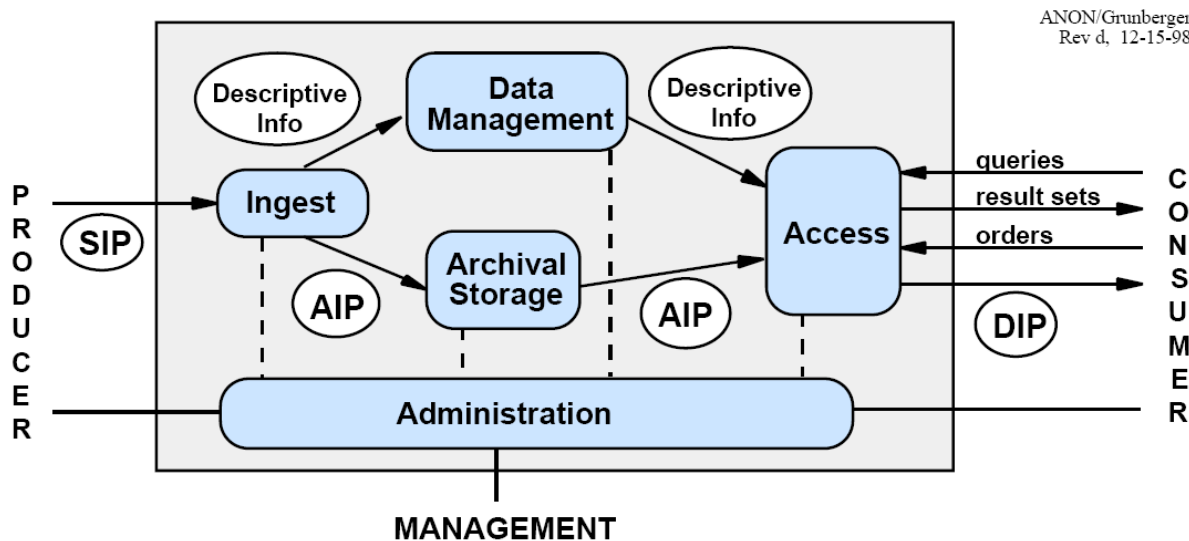
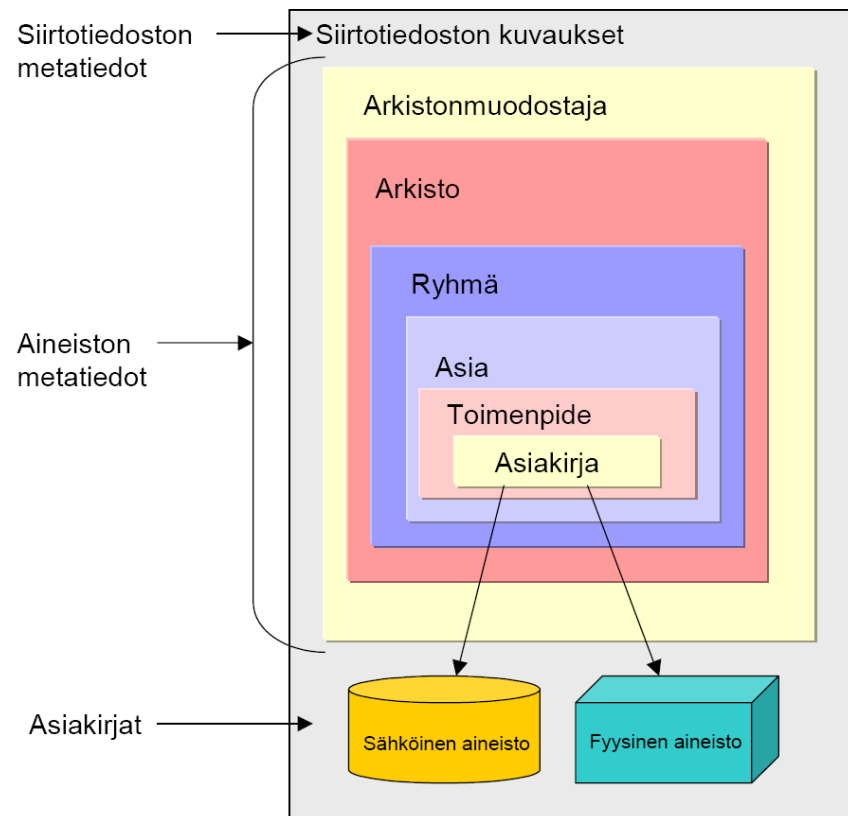


Figure 4-1. OAIS Functional Entities

Sähke

- Normi asiankäsittelyjärjestelmien tuottaman tiedon pysyvään sähköiseen arkistointiin
- Muiden järjestelmien normeja tulossa.
- Siirtotiedosto:
 - TIFF-kuva asiakirjasta
 - XML-skeeman mukainen metatieto
 - semanttisesti eheä tekstitiedosto UTF-8
 - järjestelmäkuvaus

<http://www.narc.fi/sahke/Aineisto/SAHKE-Tekninen-V2.0.pdf>



Kuva 1. Siirtotiedoston looginen rakenne.

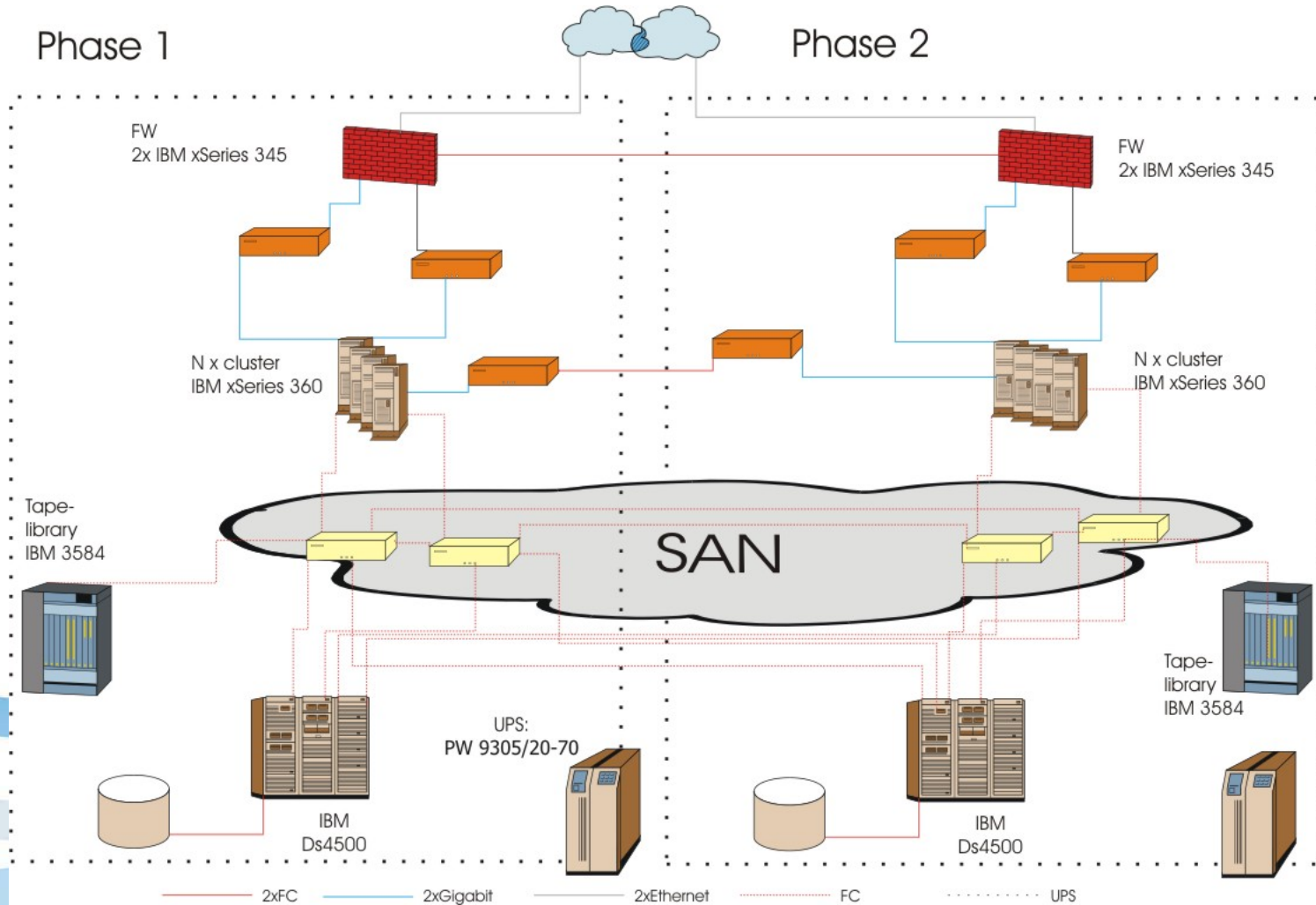
Käyttäjähallinnan ratkaisut

- Tietoa vain sille jolla on siihen oikeus
 - Käyttäjätunnukset ja vaihtuvat salasana
 - Käyttäjän tunnistus PKI-varmenteella
 - Sähköisessä arkistossa käyttöoikeudet ulottuvat dokumenttitasolle
 - Tekijänoikeuksien hallinta ja lunastaminen
- Lupatietojärjestelmät
 - Kenellä on oikeus mihinkin asiakirjaan
 - Järjestelmien käyttöönottovaihe menossa

Tekninen varmentaminen 1

- Tieto sijaitsee vähintään kahdella tietovälineellä
- Varmennus toisilla tietovälineillä (nauhat)
- Tietojen tallennus useassa paikassa
- Arkistokopiot säilytetään järjestelmien ulkopuolella
- Menetelmät kuvattu ja hyväksytty sähköiseen pysyvään ja pitkäaikaissäilytykseen:
 - Luotettava sähköinen arkisto (Trusted digital repository)
- Mikrofilmi?
- Arvokkaimman aineiston hajautus kansainvälisesti

Mikkelin Amk:n arkisto vaihe 2



Tekninen varmentaminen 2

- Tiedon aitouden, eheyden ja todistusvoimaisuuden varmentaminen
 - sähköinen allekirjoitus
 - tarkistussummat
 - vesileimat
- Metatietoihin tallennetaan
 - käsittelyhistoria (audit trail)
 - aineiston tuottamisessa käytetty tuotantoympäristö, laitteisto, käyttöjärjestelmät ja sovellukset versioineen
- Arkistoon tallennetaan vain arkistoformaateissa olevaa aineistoa

Formaatit

- Arkistokelpoisia tallennusformaatteja
 - XML
 - TIFF
 - PDF ?
 - PDF-A (Archive)
 - PDF-E (Engineering)
 - JPEG ?
 - JPEG 2000
 - Wave (BWF)
 - Motion JPEG2000
 - Normitetut ISO-SQL tietokannat
- Ei-arkistokelpoisia tallennusformaatteja
 - Valmistajakohtaiset formaatit, kuten
 - MS Office
 - Autocad DWG
 - Autocad DXF
 - Windows media
 - Valmistajakohtaiset tietokannat

Mediat

- Kiintolevyt eivät yksin riitä tallennukseen
 - LOCSS ja Distarnet hajauttavat aineistot useisiin palvelimiin
 - Arkistokappaleet järjestelmien ulkopuolelle
- Nauhaformaateissa oltava useamman toimijan tuki
 - LTO
 - IBM, HP, Quantum kehittäjät, tekniikka julkaistu
 - Useita toimittajia
- CD toimii siirtomediana kohtuullisen hyvin, jos
 - aineisto mahtuu
 - Poltetaan vähintään 2 koneella 2 eri levyille
- DVD-formaatit muutoksessa
 - uudet eivät välttämättä alaspäin yhteensopivia

Emulointi

- Rakennetaan kulloiseenkin ympäristöön virtuaalinen kone, joka toimii kuten aiempi ympäristö
- Esimerkiksi VMWare (useita käyttöjärjestelmiä)
- UVC = Universal Virtual Computer (IBM, Hollannin kansalliskirjaston projekti PADI)
 - Kaikki ohjelmistotoimittajat tekevät ohjelmistoistaan UVC-version
- Emulointi toimii osana säilytysstrategia
- Alaspäin yhteensopivuuden säilyttäminen pitkäaikaisesti mahdotonta koska se rajoittaa kehitystä

Emulointi



Migraatio

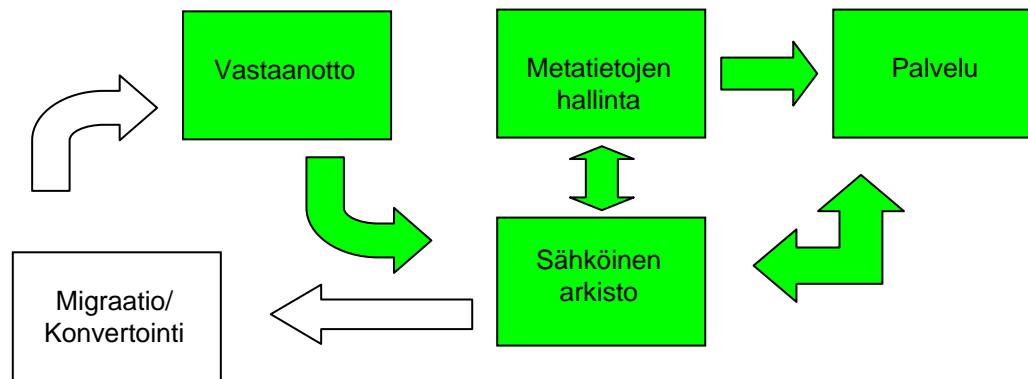
- Yleisimmin hyväksytty säilyttämisstrategia
- Migraatiossa arkistoaineisto siirretään vanhenevista formaateista uusiin, käyttöön tulleisiin formaatteihin
- Ulkoasu vai semanttinen eheys tärkeää?
- Migraatioprosessissa:
 - aineisto on käsiteltävä niin ettei sen sisältö muutu
 - kerättävä talteen kaikki tapahtumat audit trailiin
- Aineisto jaoteltava metatietojen ja formaatin tunnistuksen perusteella

Migraation edellytykset

- Migraatioprosessin on oltava automaattinen
- Käytettävä tarkastustekniikoita, manuaalinen tarkastaminen ei ole mahdollista
- Otantaan perustuva tarkastaminen?
- Metatietojen on sisällettävä riittävästi teknistä metatietoa
- Formaateista ja ohjelmistoista tarvitaan rekistereitä,
 - Pronom
<http://www.nationalarchives.gov.uk/pronom/>
- Migraatio on keskitetty toimintatapa, siksi arkistojen on oltava keskitettyjä
- Tarvitaan migraatiosuunnitelma ja seuranta

Migraation perusmalli

Sähköinen arkisto ja migraatio



Sähköisen arkiston tekninen varmentaminen ja migraatio, osa 2: 10 kohdan ohjelma

Osmo Palonen
Mikkelin Ammattikorkeakoulu



1. sääntö: IT on renki, ei isäntä

- Arkistoissa on kehitetty vuosisatojen aikana metodeja joilla voidaan säilyttää arkistoaineisto turvallisesti
- Sähköisessä arkistossa nämä menetelmät, sovellukset ja säännöt on siirrettävä IT-ympäristöön
- MIKSI on aina hyvä kysymys
- IT-arkkitehtuurin tai järjestelmän "ominaisuuksien" takia ei sääntöjä pidä muuttaa - ratkaisu on löydettävä yhteistyössä
- IT:n tuomia mahdollisuuksia ja etuja pitää toki käyttää , mutta sivuvaikutukset on arvioitava
- Insinöörejä ei ole koulutettu arvioimaan sivuvaikutuksia – se on sinun velvollisuutesi

2. sääntö: Opi perusasiat

- Huolehdi henkilöstön arkistokoulutuksesta
- Tee AMS
- Hanki osaamista sähköiseen arkistointiin
- Tarjoa mahdollisuuksia opiskeluun
 - www.mikkeli.amk.fi
 - >Erikoistumisopinnot
 - >Tekniikka ja IT
 - >Sähköisen aineiston hallinta ja arkistointi

3. sääntö: Tee oma säilytysstrategia

- Säilytysstrategia on tehtävä itse ja pidettävä omissa käsissä
- Hanki riittävästi tietoa niiltä, jotka ovat jo tehneet
- Kerää strategioita ja käytä hyväksi kokemukset: unohda huonot ja valitse hyvät käytännöt
- Älä koskaa hyväksy järjestelmää, menetelmää tai käytäntöä arvioimatta sitä
- Seuraa jatkuvasti kehitystä maailmalla, tutki, kysy, tutustu ja arvioi
- Tee myös suunnitelmat B ja C

4. sääntö: Älä jää yhden loukkuun

- Älä koskaan jätä aineistojasi yhdestä toimittajasta riippuvaiseksi
- Yksi toimittaja on liian helppo ratkaisu. Kun kuherruskuukausi on ohi, myös "kokonaisratkaisun" huonot puolet tulevat esiin
- Partnership voi merkitä myös sitä, että eräänä päivänä luet lehdestä, että partnerisi on myyty ja tuotteesi ajetaan alas – hyvässä tapauksessa tuki jatkuu, mutta muuttuu huonommaksi ja tieto lopulta katoaa
- Toinen vaihtoehto: valitse parhaat työkalut kuhunkin tarpeeseen ja kirjoita integraatio hankintasopimukseen
- Rehellisesti sanoen: ideaaliratkaisua on vaikea löytää, sinun pitää tehdä kompromisseja

5. sääntö: Käytä avoimia standardeja

- Vaikeuksissakin voit selviytyä kun valitset tuotteita, jotka perustuvat avoimiin standardeihin ISO, OSF ja W3C ovat avinlyhenteitä
- Muista kuitenkin että standardi ei ole oikeasti standardi, ellei sitä kukaan käytä
- Toisaalta laajalti käytetyt "teollisuusstandardit" ovat joskus parhaita vaihtoehtoja – on aika paljon yrityksiä, jotka toimivat Microsoft-ympäristössä ja sen käyttäminen on usein taloudellinen ratkaisu

6. sääntö: Virheen kestävät ratkaisut

- Rakenna ratkaisusi virheen kestäväksi
- Virheen kestävän laitteisto- ja ohjelmistokokonaisuuden lisäinvestoinnin voi säästää yhdessä yössä
- Järjestelmän seisokit käyvät usein kalliimmaksi kuin järjestelmän rakentaminen yhden virheen sietäväksi

7. sääntö: Luota vain tallennusosaajiin

- Luota vain sellaisiin IT-yrityksiin, jotka ovat sitoutuneet tallennusjärjestelmiin
- Vakavasti otettava arkistointi vaatii SAN-järjestelmää
- Selvitä toimittajan kyky palvella: jos järjestelmään tulee pieni vika, sinulla voi olla suuri ongelma
- Tallennusympäristön rakentamisessa tarvitaan toisenlainen strategia kuin sovelluksia valittaessa
- Tallennusjärjestelmä on sähköisen arkiston perustus

8. sääntö: Tiedosta median ongelmat

- Kovalevyt käyvät yhä halvemmiksi, mutta...
- Tarvitaan toinen tallennuspaikka arkistokopioille
- Käytä nauhaa, jota tukee useampi kuin yksi maailmanlaajunen toimittaja
- Unohda DAT, CD ja DVD, niitä voi käyttää vain levitykseen
- Muista että media on yksi osa migraatiostrategiaa, se tarvitsee jatkuvaa virkistystä ja konvertointia

9. sääntö: Et voi ulkoistaa vastuuta

- Sinulla on vastuu arkistoaineistosta, vaikka virheen tekeekin palvelun toimittaja
- Jos aineisto katoaa, tiukat sopimukset ja korkeat sakot eivät tuo takaisin sitä.

10. sääntö: Valitse oikein

- Kannattaako sittenkään tehdä itse?



MIKKELIN AMMATTIKORKEAKOULU
MIKKELI POLYTECHNIC

YLLÄTTÄVÄN HYVÄ.