



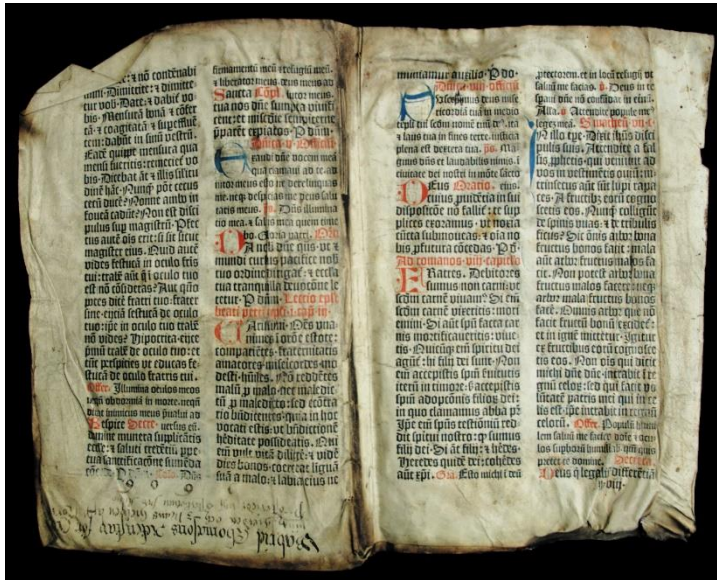
Miten digitaalinen tieto säilyy

lokakuu 2022

Kimmo Koivunen



Kumpi säilyy seuraavat sata vuotta?



Missale Aboense 1488

Vanhin Suomea varten painettu kirja

<http://jyk.jyu.fi/jyk/kokoelmat/missale/>



Tietokoneen kovalevy 2018

Kuva: [William Warby](#)

Digitaalinen pitkäaikaissäilytys?

- Tarkoituksena säilyttää digitaaliset aineistot eheinä, autenttisia ja käyttökelpoisina useiden vuosikymmenien tai satojen vuosien ajan
- Mahdollistaa digitaalisten aineistojen hyödyntäminen tuleville sukupolville
- Toteutuksen näkökulmasta korostaa jatkuvuussuunnittelua
 - Toiminnan elinkaari ylittää kaikkien toteutuksen osakokonaisuuksien käyttöajan
- Pitkäaikaissäilyttäminen on aina aktiivista toimintaa



Digitaalisiin aineistoihin kohdistuvat uhat ja niihin varautuminen

- Säilytysmedian viat ja vanhentuminen
 - Datasta useampi kopio erilaisilla medioilla, medioiden virkistäminen
- Toimittajakohtaiset systemaattiset viat ja vanhentuminen
 - Datasta kopiot eri toimittajien ratkaisussa, ratkaisut mahdollisimman avoimia, kokonaisuuden elinkaaren hallinta
- Toiminnallinen tai hallinnollinen virhe
 - Data hallinoidaan vähintään kahdessa järjestelmässä
- Luonnonkatastrofit ja muut sijaintiin liittyvät uhat
 - Säilytys toteutetaan maantieteellisesti hajautettuna
- Pahantahtoinen käyttäjä
 - Säilytystä varmennetaan ns. pimeällä arkistolla
- Osaamisen ja/tai henkilöstön riittämättömyys
 - Toiminnan keskittäminen, kyvykkyyksien tunnistaminen ja kehittäminen

Functional Area	Level			
	Level 1 (Know your content)	Level 2 (Protect your content)	Level 3 (Monitor your content)	Level 4 (Sustain your content)
Storage	<p>Have two complete copies in separate locations</p> <p>Document all storage media where content is stored</p> <p>Put content into stable storage</p>	<p>Have three complete copies with at least one copy in a separate geographic location</p> <p>Document storage and storage media indicating the resources and dependencies they require to function</p>	<p>Have at least one copy in a geographic location with a different disaster threat than the other copies</p> <p>Have at least one copy on a different storage media type</p> <p>Track the obsolescence of storage and media</p>	<p>Have at least three copies in geographic locations, each with a different disaster threat</p> <p>Maximize storage diversification to avoid single points of failure</p> <p>Have a plan and execute actions to address obsolescence of storage hardware, software, and media</p>
Integrity	<p>Verify integrity information if it has been provided with the content</p> <p>Generate integrity information if not provided with the content</p> <p>Virus check all content; isolate content for quarantine as needed</p>	<p>Verify integrity information when moving or copying content</p> <p>Use write-blockers when working with original media</p> <p>Back up integrity information and store copy in a separate location from the content</p>	<p>Verify integrity information of content at fixed intervals</p> <p>Document integrity information verification processes and outcomes</p> <p>Perform audit of integrity information on demand</p>	<p>Verify integrity information in response to specific events or activities</p> <p>Replace or repair corrupted content as necessary</p>
Control	<p>Determine the human and software agents that should be authorized to read, write, move, and delete content</p>	<p>Document the human and software agents authorized to read, write, move, and delete content and apply these</p>	<p>Maintain logs and identify the human and software agents that performed actions on content</p>	<p>Perform periodic review of actions/access logs</p>
Metadata	<p>Create inventory of content, also documenting current storage locations</p> <p>Backup inventory and store at least one copy separately from content</p>	<p>Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)</p>	<p>Determine what metadata standards to apply</p> <p>Find and fill gaps in your metadata to meet those standards</p>	<p>Record preservation actions associated with content and when those actions occur</p> <p>Implement metadata standards chosen</p>
Content	<p>Document file formats and other essential content characteristics including how and when these were identified</p>	<p>Verify file formats and other essential content characteristics</p> <p>Build relationships with content creators to encourage sustainable file choices</p>	<p>Monitor for obsolescence, and changes in technologies on which content is dependent</p>	<p>Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed</p>

Toiminnalliset alueet:

- Tallennus
- Eheys
- Valvonta
- Metatieto
- Aineisto

Suomeksi saatavilla:

www.digitalpreservation.fi/specifications/ndsalevels

Muuta esimerkki “Säilyttämisen tasot”-mallista

Metatieto	Taso 2 (Suojaa aineistosi)
	Tallenna riittävästi metatietoa tietääksesi, mitä aineisto on (tämä voi sisältää jonkin yhdistelmän seuraavista: säilytysmetatieto sekä hallinnollinen, tekninen, kuvaileva ja rakenteellinen metatieto)

Tarvitaan muutakin kuin kuvailevaa metatietoa

Taso 4 (Ylläpidä aineistoasi)

Pidä ainakin kolmea kopiota eri maantieteellisissä sijainneissa, joissa onnettomuusriskit ovat keskenään erilaiset

Maksimoi tallennustapojen vaihtelevuus välttääksesi yksittäisten vikaantumiskohteiden syntymisen

Suunnittele ja toteuta tallennuslaitteistojen, -ohjelmistojen ja -alustojen vanhenemisen ratkaisemiseksi tarvittavat toimenpiteet

Maantieteellinen hajauttaminen

Teknologian heterogeenisuus

Jatkuvuussuunnittelu

Tärkeistä aineistojen ohjelmistojen

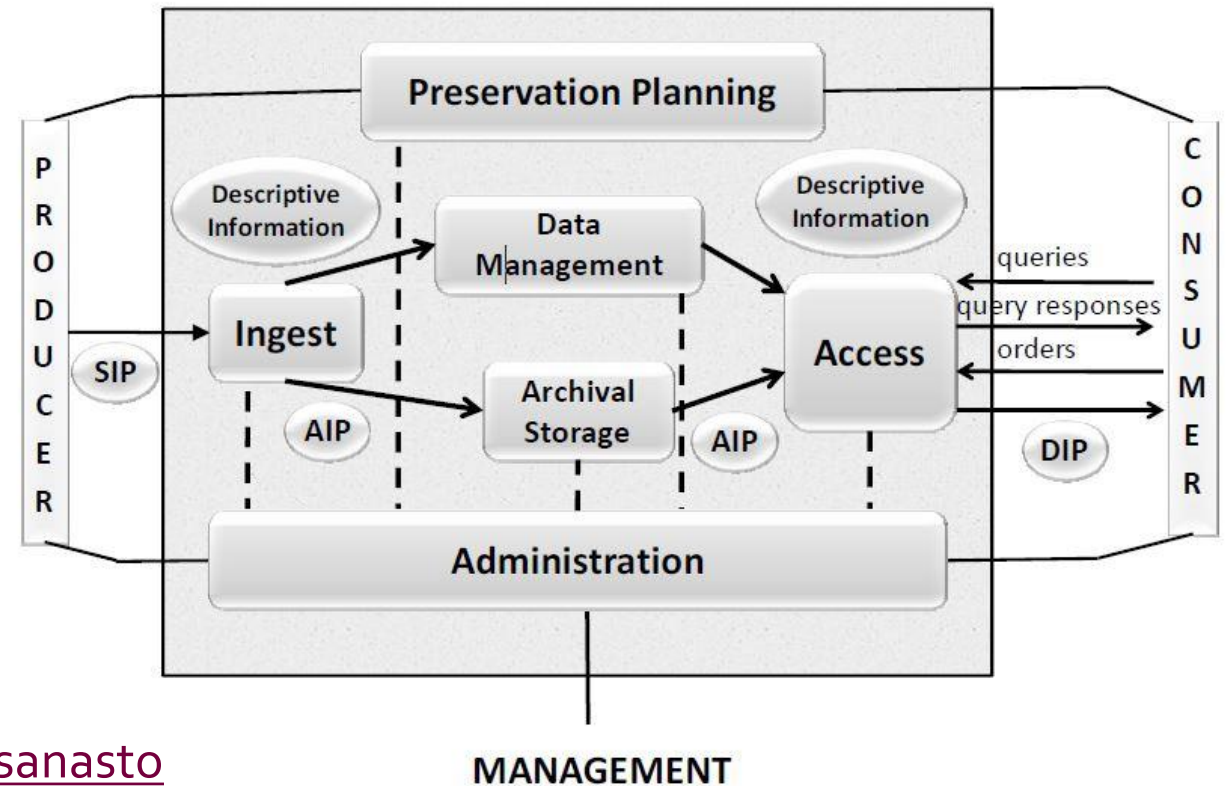
Pitkäaikaissäilytyksen OAIS-viitemalli

ISO 14721:2012 (suomennos SFS 5972)

- Keskeisiä käsitteitä:
 - Siirtopaketti (SIP)
 - Säilytyspaketti (AIP)
 - Jakelupaketti (DIP)
 - Kohdeyhteisö (designated community)

- PAS-sanasto:

- <http://digitalpreservation.fi/specifications/sanasto>



Säilyttäminen on aktiivista toimintaa

- Aineistojen laatu tarkistettava
 - Määritellään ns. säilyttämisen lähtötaso
- Aineistojen eheyttä tarkkaillaan jatkuvasti
 - Eheyden palauttaminen normaalia palvelutoimintaa
- Aktiivinen teknologioiden seuranta
 - Tiedostomuodot, ohjelmistot, ...
- Laitteistojen elinkaaren hallinta tärkeää
 - Levy- ja nauharikoista toipuminen normaalia toimintaa
- Hyödyntävien organisaatioiden tukeminen
- Tietoturva ja tietosuoja huomioitava toiminnassa
 - PAS-palvelut sisältyvät CSC:n ISO/IEC 27001:2013-auditoinnissa tarkasteltuihin palveluihin

Säilytyksessä
oleva
aineistomäärä
kasvaa noin 50
teratavua
kuukaudessa

Aineistot
tallennettu
kolmelle eri
mediatyypille
(levy ja kaksi
nauhaa)

Yhteistyö mahdollistaa aineistojen pitkäkestoisen hyödyntämisen

