

# Digitaalinen pitkäaikaissäilytys

Arto Teräs <arto.teras@csc.fi>

Kirjastoverkkopäivä

Helsingin yliopisto 6.5.2009



# Esityksen sisältö

- **Mitä pitkäaikaissäilytys tarkoittaa?**
- **Säilyttämisen vaatimustasot**
- **Muutoksia viimeisen 30 vuoden aikana**
- **Standardit, lait ja normit**
  - OAIIS-viitemalli, TRAC-tarkistuslista
  - Metatietoihin liittyvät standardit
- **Pitkäaikaissäilytyksen looginen kokonaisuus**
  - Informaation luovutus, säilytys ja käyttö
- **Pitkäaikaissäilytyksen riskit ja ratkaisumalleja**
  - Vaatimukset pitkäaikaissäilytysjärjestelmälle
- **Mitä pitkäaikaissäilytys maksaa?**



# Mitä pitkäaikaissäilytys (PAS) tarkoittaa?

- **Digitaalisen informaation luotettavaa säilyttämistä useiden kymmenien tai satojen vuosien ajan**
  - Laitteet ja ohjelmistot, jopa tiedostomuodot vanhenevat, mutta informaation täytyy säilyä
  - Luotettava säilyttäminen edellyttää sisällön eheyden aktiivista valvontaa ja monenlaisiin riskeihin varautumista
- **Informaation, ei pelkästään bittien säilyttämistä**
  - Metatiedot tärkeässä asemassa
- **Aikajänne pitempi kuin joidenkin organisaatioiden elinaika**
  - Tarvitaan yhteistyötä, standardeja ja käytännöistä sopimista
- **Käyttö on tärkeä osa säilytystä => pitkäaikaissäilytyksen toimenpiteillä varmistetaan informaation jatkuva käytettävyys**
  - Käyttö myös edistää ymmärrettävyyden säilymistä



# Säilyttämisen vaatimustasot

- **Bittien säilyttäminen**
  - Alkuperäinen bittijono voidaan toistaa nykyaikaisilla laitteilla, mutta ei takeita siitä voidaanko sisältöä tulkita
  - Säilyttämisen perusta ja optio tulevaisuuteen: antaa mahdollisuuden nostaa vaatimustasoa myöhemmin jos tarvetta ja rahaa on
- **Sisällön ymmärrettävyyden säilyttäminen**
  - Teksti, kuvat ja muu olennainen sisältö voidaan tulkita
  - Ymmärrettävyys riippuu myös kohdeyleisöstä
  - Käytön ja uuden tiedon tuottamisen kannalta yleensä tärkein taso
- **Alkuperäisen käyttökokemuksen säilyttäminen**
  - Kalvoesityksen tehosteet toistuvat alkuperäisessä muodossaan
  - Vanha tietokonepeli näyttää ja kuulostaa samalta kuin 30 vuotta sitten
  - Vaativin, mutta historiantutkimuksen kannalta mielenkiintoinen taso



# Muutoksia viimeisen 30 vuoden aikana

- **1980-luku**
  - Mietittiin, mikä nauha tai levy kestää pisimpään
  - Tiedostomuodot vaihtuivat tiuhaan eivätkä olleet yhteensopivia
- **1990-luku**
  - Internet, tietokannat ja verkkopalvelut: Sisältöjä alettiin käyttää verkkoyhteyksien yli
- **2000-luku**
  - Tallennusvälineet muuttuvat edelleen nopeasti => asia hyväksytty, bitit siirretään vanhoilta levyiltä ja nauhoilta uudemmille
  - Tiedostomuodot muuttuvat hitaammin kuin aiemmin ja uudet ohjelmat osaavat yleensä lukea vanhoja muotoja => ongelma pienentynyt
  - Uusia julkaisumuotoja ja tiedon tuottajia (blogit, Internetin kuva- ja videopalvelut, sosiaaliset verkkopalvelut) => miten nämä talteen?
  - Pitkäaikaissäilytyksen standardit (OAIS, TRAC, ...)
  - Globaalit tietovarastot, saatavilla olevan tiedon määrän huima kasvu, hakutoimintojen korostuminen

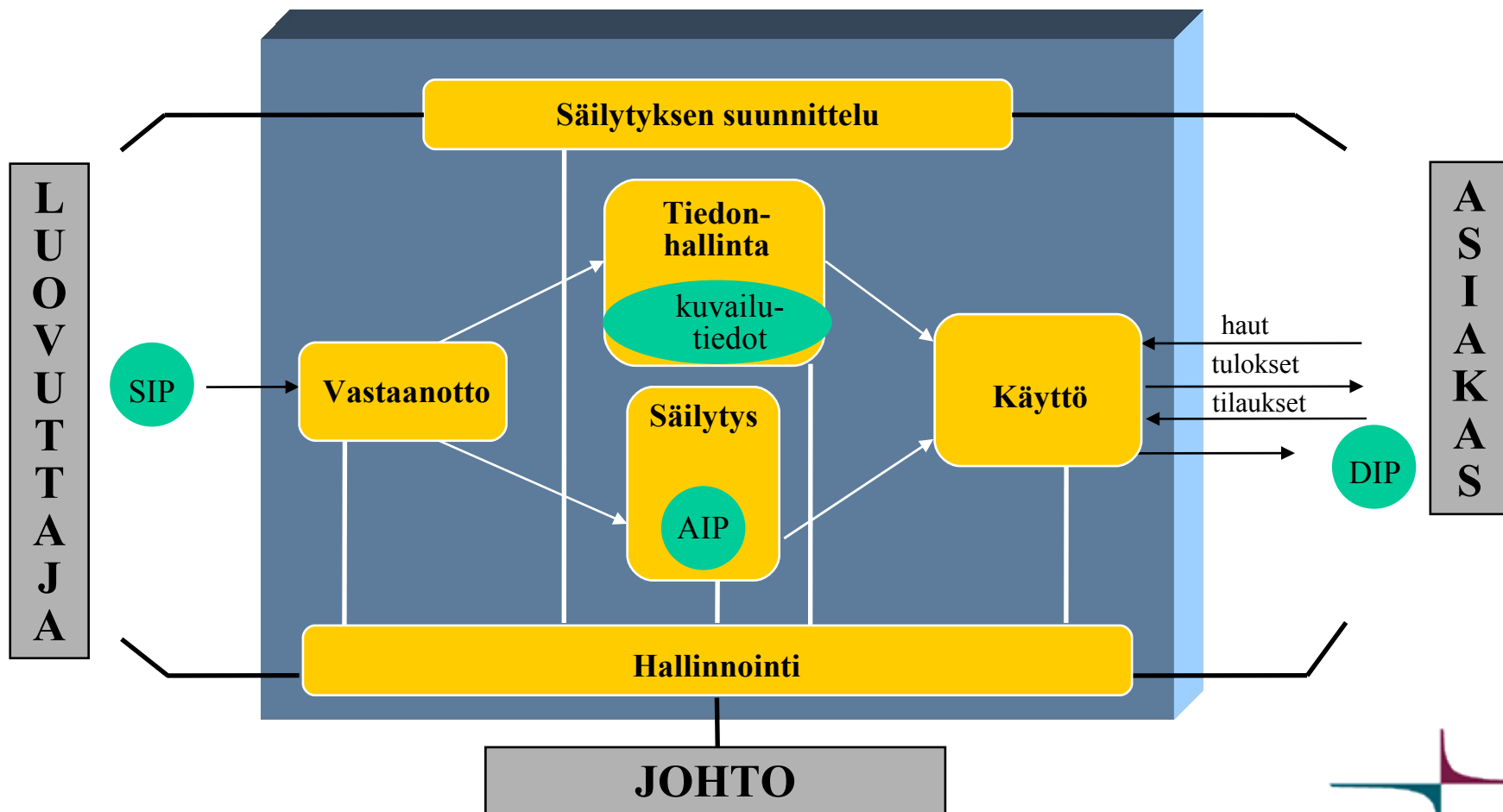


# Standardit, lait ja normit

- **Säilytysjärjestelmää ja organisaatiota koskevat standardit**
  - Open Archival Information System (OAIS) -viitemalli kuvaa loogista kokonaisuutta (OAIS on myös ISO-standardi ISO 14721:2003)
  - Trustworthy Repositories Audit & Certification (TRAC) -tarkistuslista määrittelee turvallisen säilyttämisen periaatteet
- **Metatietoihin liittyvät standardit**
  - Kuvailustandardit (mm. Dublin Core), asiasanastot, ontologiat
  - Teknisemmät kuvailu- ja paketoitistandardit: PREMIS ja METS
- **Kotimaiset lait ja normit**
  - Julkisuuslaki, henkilö- ja tietosuojalait, valmiuslaki, tekijänoikeuslaki, arkistolaki, museolaki, laki kulttuuriaineistojen tallettamisesta ja säilyttämisestä
  - VAHTI-tietoturvaohjeet ja -määräykset
- **Muut hyvät toteutuskäytännöt, mm. tietoturvan osalta**
  - Paljon materiaalia tarjolla, kannattaa valikoida ja hyödyntää!



# Pitkäaikaissäilytyksen kokonaisuus OAIS:n toiminnallisen mallin pohjalta

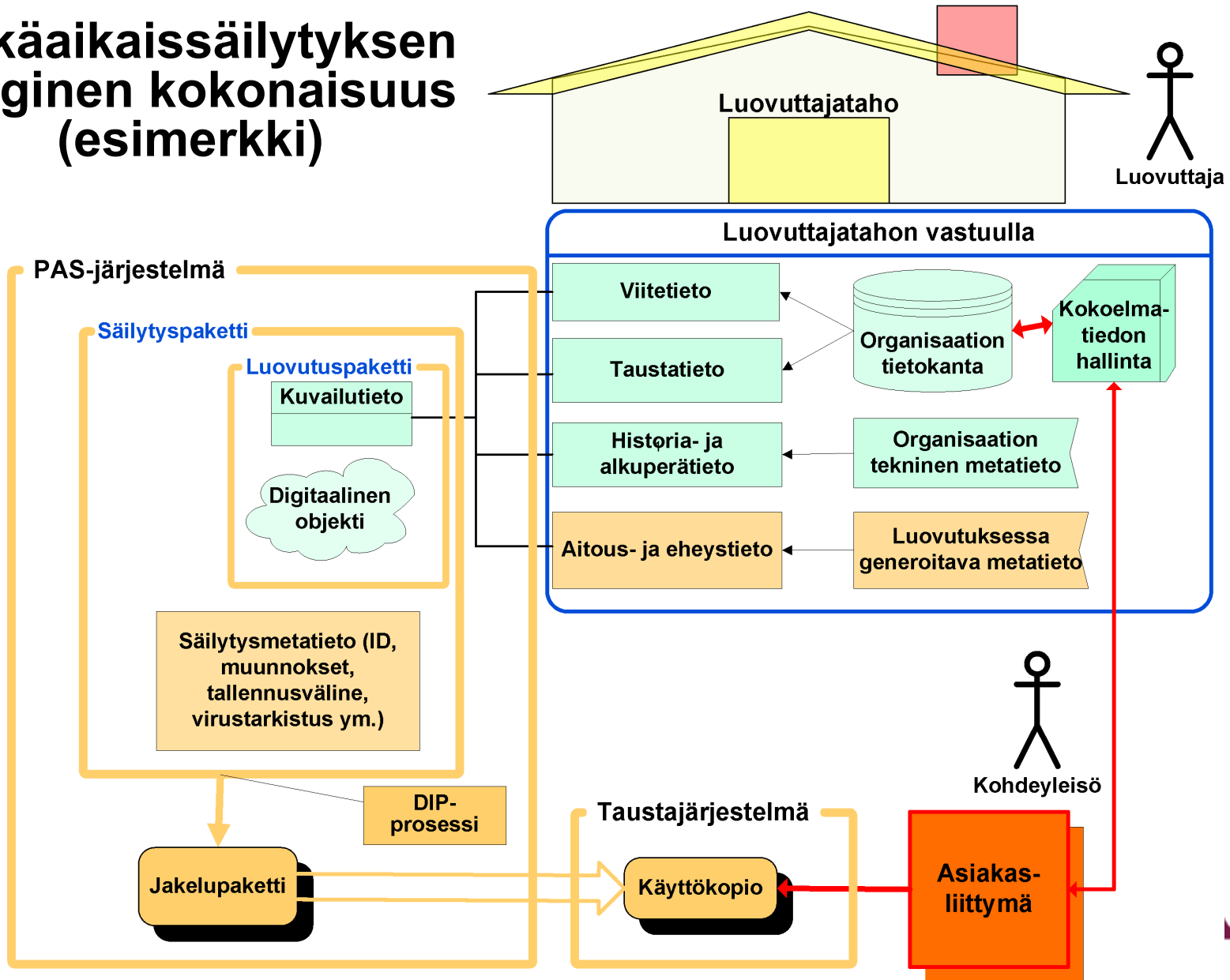


# OAIS-mallin mukaiset tietopaketit

- **Luovutuspaketti (Submission Information Package, SIP)**
  - Koostuu luovutettavasta aineistosta ja siihen liittyvistä metatiedoista
- **Säilytyspaketti (Archival Information Package, AIP)**
  - Pitkäaikaissäilytysjärjestelmässä (PAS-järjestelmässä) säilytettävä tietopaketti, joka muodostetaan yhden tai useamman luovutuspaketin pohjalta.
  - Koostuu sisällöstä ja siihen liittyvästä säilytysinformaatiosta
- **Jakelupaketti (Dissemination Information Package, DIP)**
  - Yhdestä tai useammasta säilytyspaketista koostettu tietopaketti, jonka asiakas (tai taustajärjestelmä) saa tehtyään aineistopyynnön PAS-järjestelmälle
- **Tietopaketti (Information Package)**
  - Yleistermi edellämainituista



# Pitkäaikaissäilytyksen looginen kokonaisuus (esimerkki)



# Pitkäaikaissäilytyksen riskit

- **Tekniset riskit**
  - Mm. tallennusmedian tai laitteiston rikkoutuminen, ohjelmistovirheet
- **Inhimilliset riskit**
  - Ylläpitäjän virhe
  - Hyökkäys järjestelmää vastaan (organisaation ulkoa tai sisältä)
- **Eryisesti pitkäaikaissäilytykseen liittyvät riskit**
  - Tiedon ymmärrettävyyden häviäminen
  - Piilevät virheet ("bit rot")
- **Katastrofit**
  - Luonnonilmiöt: tulva, tulipalo ym.
  - Yhteiskunnalliset: Sotatila, rahoituksen loppuminen ym.



# Keinoja riskien minimoimiseksi

- **Useampi kuin yksi tallennusmediatyyppe (esim. kiintolevy, nauha, optiset levyt, kertakirjoitettavat mediat)**
- **Automaattinen informaation eheyden tarkistus ja virheiden korjaus**
- **Avoimet rajapinnat, avoimen lähdekoodin ohjelmistot**
- **Suunnitelmat laitteisto- ja ohjelmistotoimittajasta toiseen siirtymiseksi (ns. exit-strategiat)**
- **Turvalliset verkko- ja tilaratkaisut**
- **Järjestelmän toiminnan valvonta, lokitietojen keruu ja analysointi**
- **Kopioiden hajautus maantieteellisesti eri paikkoihin, eri kopioiden tallennusprosessien riittävä itsenäisyys**
- **Aineiston käyttö ja käytettävyyden seuranta**
- **Järkevä rahankäyttö**

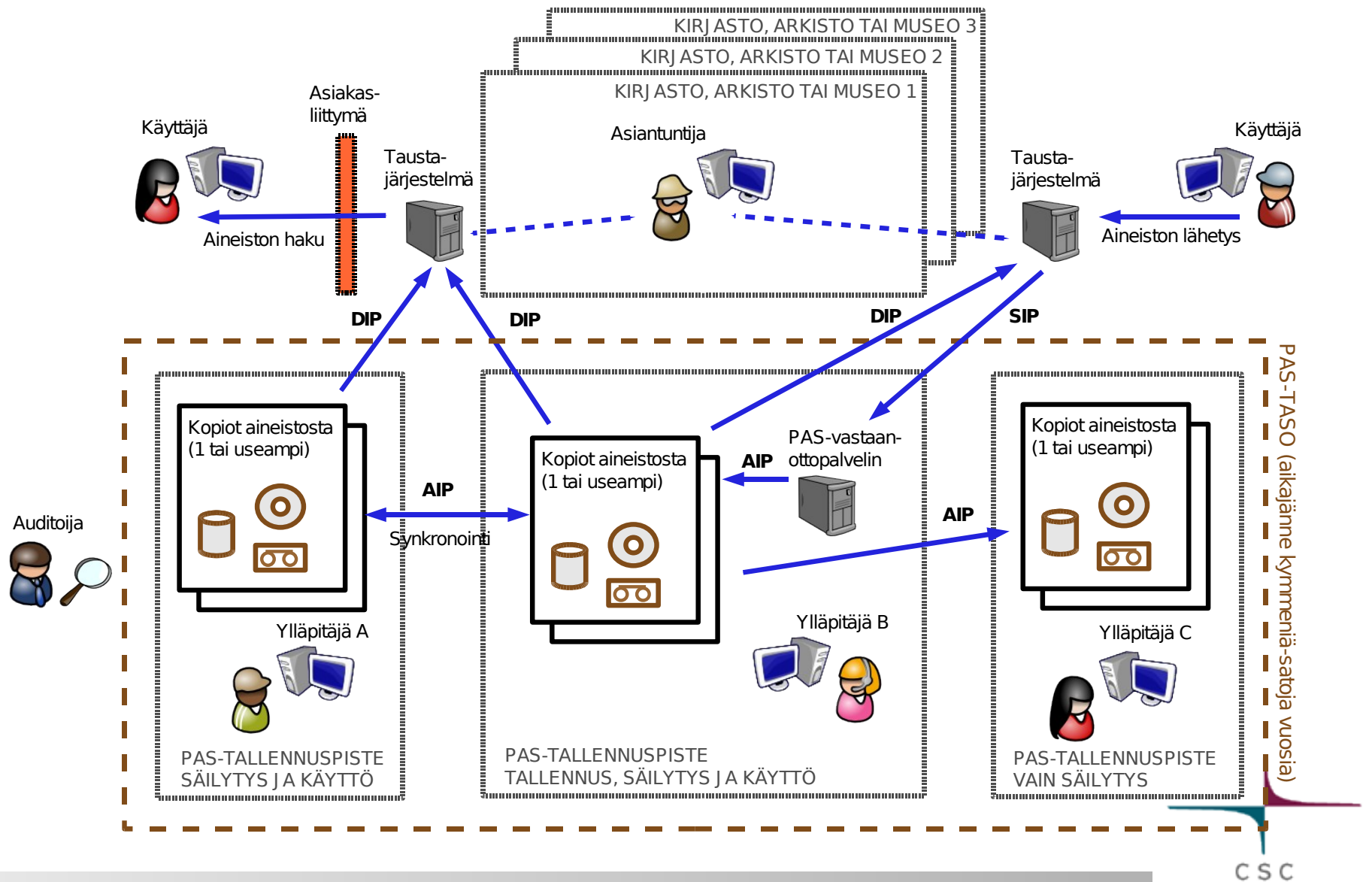


# Pitkäaikaissäilytysjärjestelmän tekniseltä infrastruktuurilta vaadittavat ominaisuudet

- **Järjestelmässä ei saa olla yksittäistä kohtaa, jonka pettäessä koko järjestelmä pettää (single point of failure)**
- **Järjestelmän täytyy toimia, vaikka mediat, ohjelmistot ja laitteistot vaihtuvat tasaisena virtana**
  - Järjestelmän täytyy tukea erilaisuutta ja välttää lukkiutumista tiettyjen laite- tai ohjelmistotoimittajien ratkaisuihin
- **Järjestelmän täytyy jatkuvasti seurata tietopakettien eheyttä**
  - Suurinta osaa järjestelmässä olevaa aineistoa käytetään harvoin, ja se aiheuttaa suuren riskin piilevien virheiden kerääntymiseen
- **Järjestelmän tulee olla ylläpidettävissä pääosin katkotta**
- **On mietittävä, halutaanko kaiken säilytettävän digitaalisen informaation olevan saatavilla viiveettä (esim. kiintolevyillä) vai riittääkö alhaisempi taso.**



# Esimerkki teknisestä infrastruktuurista



# Mitä pitkäaikaissäilytys maksaa?

- **Esimerkkejä saatavilla, muttei vielä yhtenäisiä malleja jotka tekisivät niistä keskenään vertailukelpoisia**
  - LIFE<sup>2</sup>-projektin (<http://www.life.ac.uk/>) elinkaarimalli ehkä lähimpänä tässä suhteessa
- **Tarvitaan jatkuvaa rahoitusta**
  - Laitteistojen ja ohjelmistojen uusiminen, henkilötyö
- **Säilytyksen vaatimustason vaikutus kustannuksiin**
  - Bittien, sisällön vai alkuperäisen käyttökokemuksen säilytys?
- **Kaikkea ei voida säilyttää: valikoitava**
- **Huolellinen valmistelu ja sisältöjen kuvailu kannattaa**
  - Yleensä aina halvempaa nyt kuin myöhemmin
  - Edistää sekä säilytystä että erityisesti käytettävyyttä
- **Mitä maksaa jos ei tehdä mitään?**



# Lisätietoja

- **Opetusministeriön Kansallinen digitaalinen kirjasto -hanke**
  - [http://www.minedu.fi/OPM/Kulttuuri/kulttuuripolitiikka/linjaukset\\_ohjelmat\\_ja\\_hankkeet/digitaalinen\\_kirjasto/](http://www.minedu.fi/OPM/Kulttuuri/kulttuuripolitiikka/linjaukset_ohjelmat_ja_hankkeet/digitaalinen_kirjasto/)
  - Pitkäaikaissäilytysjärjestelmän alustava toiminnallinen vaatimusmäärittely
- **Alan kirjallisuus, mm.**
  - Deegan, Tanner: Digital Preservation (ISBN 978-1-85604-485-1)
  - Gladney: Preserving Digital Information (ISBN 978-3-540-37886-0)
- **Runsaasti materiaalia verkossa, mm.**
  - Digital Preservation Tutorial (Cornell University Library):  
<http://www.icpsr.umich.edu/dpm/dpm-eng/>
  - DPC/PADI What's new in digital preservation:  
<http://www.dpconline.org/graphics/whatsnew/>
  - Blue Ribbon Task Force: Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation  
[http://brtf.sdsc.edu/biblio/BRTF\\_Interim\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf)

