



Kolme vuotta digitaalista pitkäaikaissäilytystä

5.11.2018

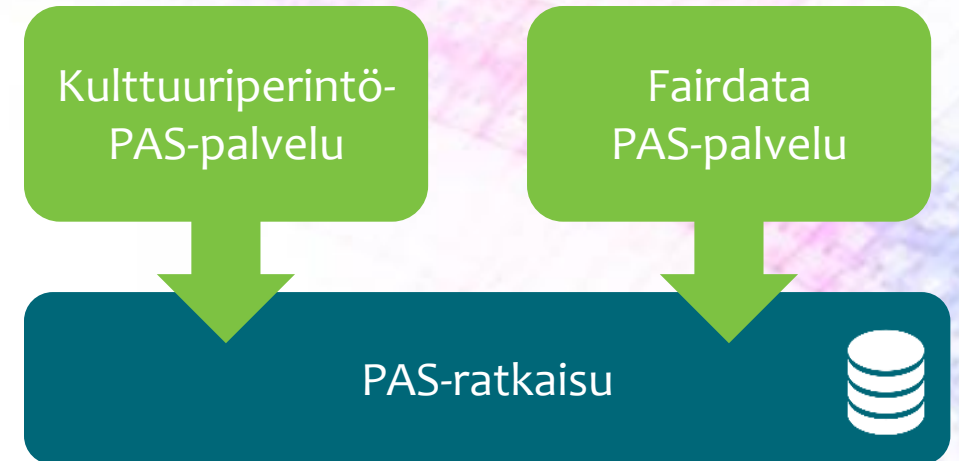
Kimmo Koivunen



CSC – Suomalainen tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon ICT-osaamiskeskus

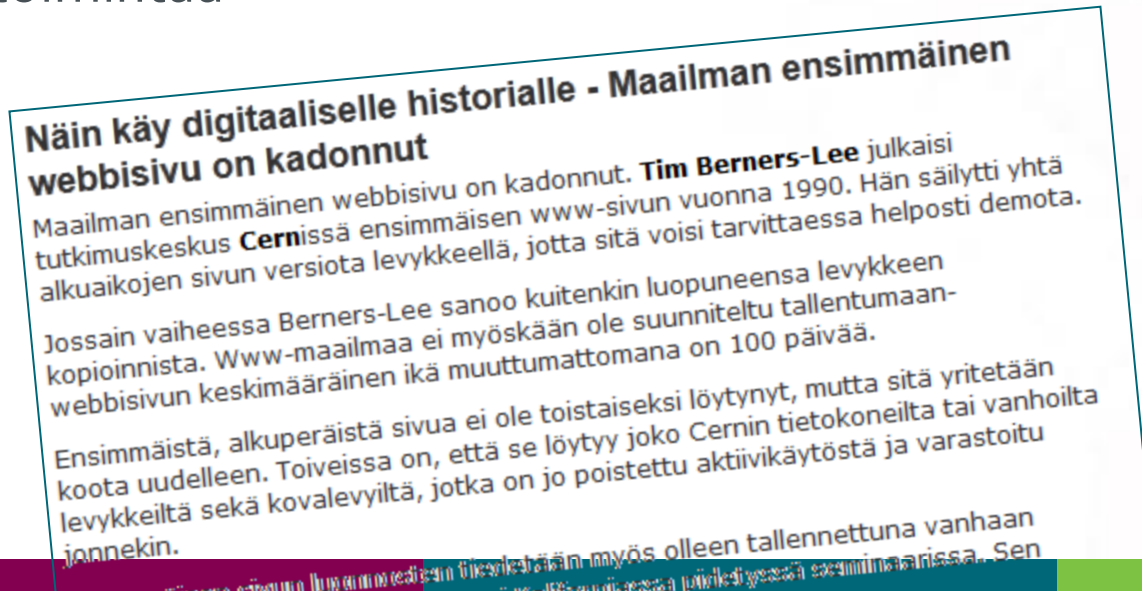
Säilyttämisen palvelut ja PAS-ratkaisu

- CSC tuottaa OKM:lle keskitettyjä PAS-palveluita
 - kulttuuriperinnön säilyttämiseen (tuotannossa vuodesta 2015)
 - Tutkimusaineistojen säilyttämiseen (tavoitteena aloittaa säilytys vuoden 2019 alussa)
- Pohjautuvat yhteiseen bittitason säilytyksen ratkaisuun
- Taustalla OKM:n KDK- ja ATT-hankkeet, joissa laaditut arkkitehtuurit tunnistavat säilyttämisen tarpeet ja asemoivat PAS-palvelut
 - Kulttuuriperinnön säilyttämiseen on monilla organisaatioilla lakisääteinen velvollisuus
- Tavoitteena on mahdollistaa digitaalisten aineistojen käyttökelpoisuus, autenttisuus ja eheys kymmenien tai jopa satojen vuosien ajan
- CSC on ottanut säilytysvastuun noin 150 teratavusta ja yli miljoonasta tietopaketesta
- PAS-palvelut sisältyvät CSC:n ISO/IEC 27001:2013-auditoinnissa tarkasteltuihin palveluihin



Pitkäaikaissäilytys = pitkäaikaissaatavuus = Digital Preservation

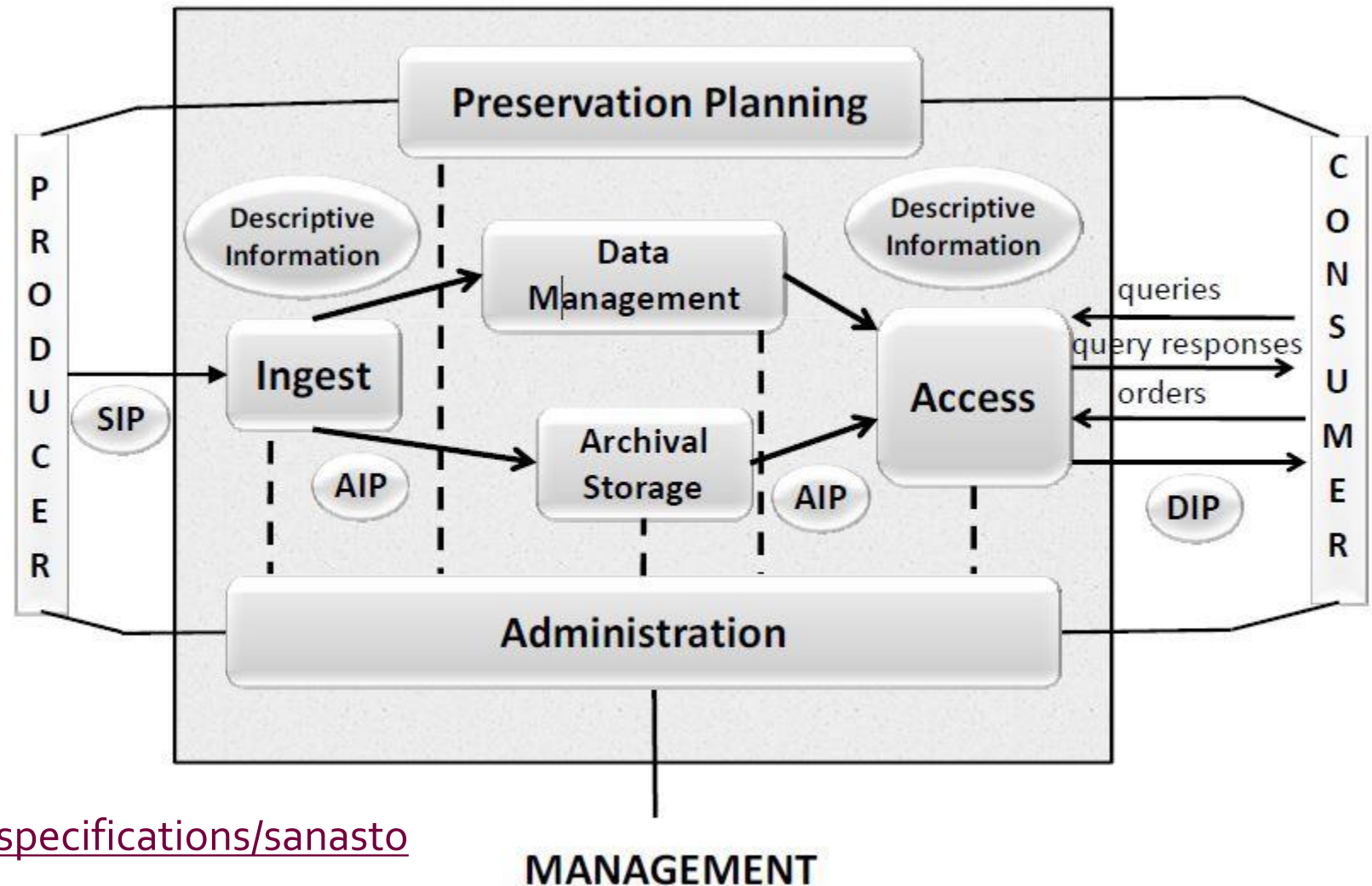
- Tarkoituksena säilyttää digitaaliset aineistot käyttökelpoisina useiden vuosikymmenien tai satojen vuosien ajan
- Mahdollistaa digitaalisten aineistojen hyödyntäminen tuleville sukupolville
- Toteutuksen näkökulmasta korostaa jatkuvuussuunnittelua
 - toiminnan elinkaari ylittää kaikkien toteutuksen osakokonaisuuksien käyttöajan
- Säilyttäminen on aktiivista toimintaa



Pitkäaikaissäilytyksen OAIS-viitemalli

ISO 14721:2012 (suomennos SFS 5972)

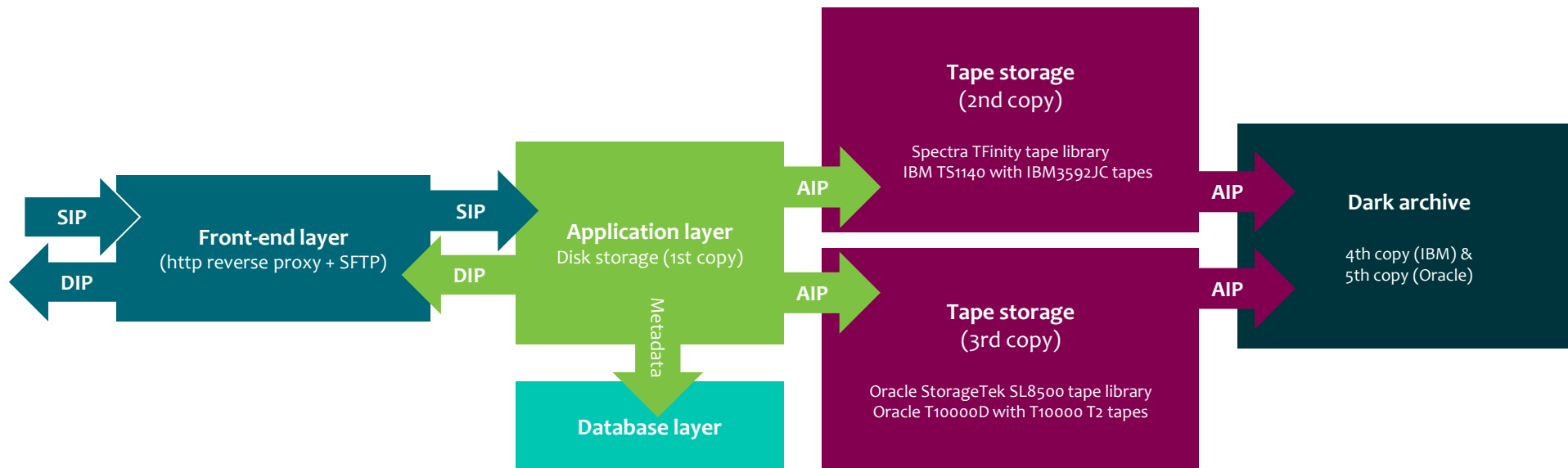
- Keskeisiä käsitteitä:
 - Siirtopaketti (SIP)
 - Säilytyspaketti (AIP)
 - Jakelupaketti (DIP)
 - Kohdeyhteisö (designated community)



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

Uhat ja niihin varautuminen

- Säilytysmedian viat
 - Datasta useampi kopio erilaisilla medioilla
- Toimittajakohtaiset systemaattiset viat
 - Datasta kopiot eri toimittajien ratkaisussa
- Toiminnallinen tai hallinnollinen virhe
 - Dataa hallinnoidaan vähintään kahdessa järjestelmässä
- Luonnonkatastrofit
 - Säilytys toteutetaan maantieteellisesti hajautettuna
- Pahantahtoinen käyttäjä
 - Säilytystä varmennetaan ns. pimeällä arkistolla
- Osaamisen ja henkilöstön riittämättömyys
 - Toimintaa keskittämällä ja suunnitelmallisuudella varmistetaan osaamisen kehittyminen ja riittävä henkilöstö



NDSA: Levels of digital preservation

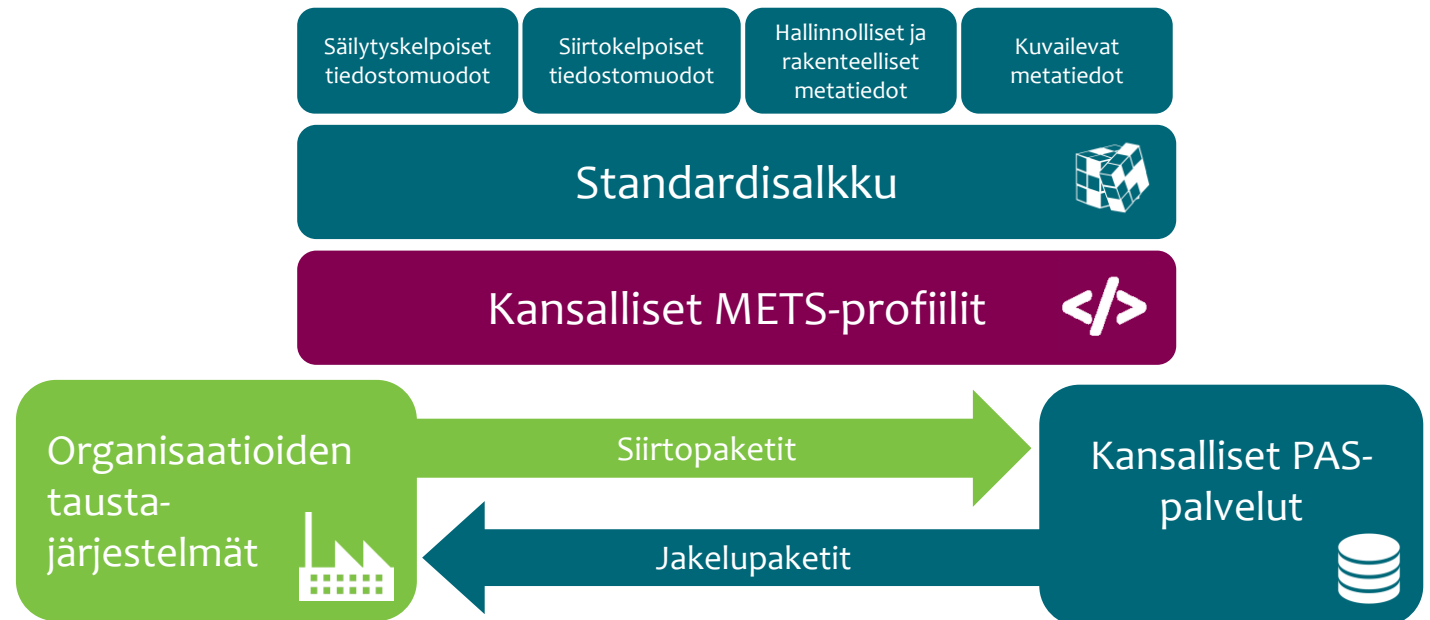
- storage and geographic location
- file fixity and data integrity
- information security
- metadata
- file formats

Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	<ul style="list-style-type: none"> - Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system 	<ul style="list-style-type: none"> - At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them 	<ul style="list-style-type: none"> - At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media 	<ul style="list-style-type: none"> - At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	<ul style="list-style-type: none"> - Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content 	<ul style="list-style-type: none"> - Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content 	<ul style="list-style-type: none"> - Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content 	<ul style="list-style-type: none"> - Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	<ul style="list-style-type: none"> - Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files 	<ul style="list-style-type: none"> - Document access restrictions for content 	<ul style="list-style-type: none"> - Maintain logs of who performed what actions on files, including deletions and preservation actions 	<ul style="list-style-type: none"> - Perform audit of logs
Metadata	<ul style="list-style-type: none"> - Inventory of content and its storage location - Ensure backup and non-collocation of inventory 	<ul style="list-style-type: none"> - Store administrative metadata - Store transformative metadata and log events 	<ul style="list-style-type: none"> - Store standard technical and descriptive metadata 	<ul style="list-style-type: none"> - Store standard preservation metadata
File Formats	<ul style="list-style-type: none"> - When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs 	<ul style="list-style-type: none"> - Inventory of file formats in use 	<ul style="list-style-type: none"> - Monitor file format obsolescence issues 	<ul style="list-style-type: none"> - Perform format migrations, emulation and similar activities as needed

NDSA: Levels of digital preservation

- storage and geographic location
- file fixity and data integrity
- information security
- metadata
- file formats

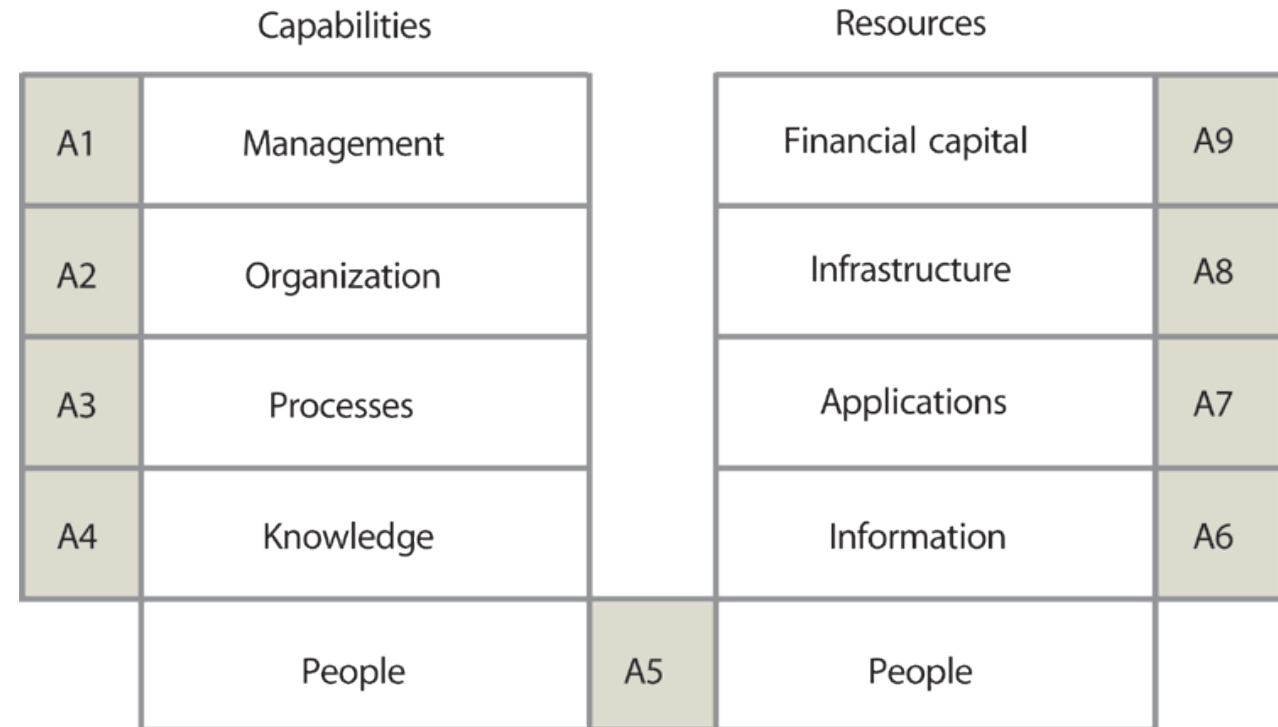


Yhteistyö mahdollistaa aineistojen pitkäkestoisen hyödyntämisen

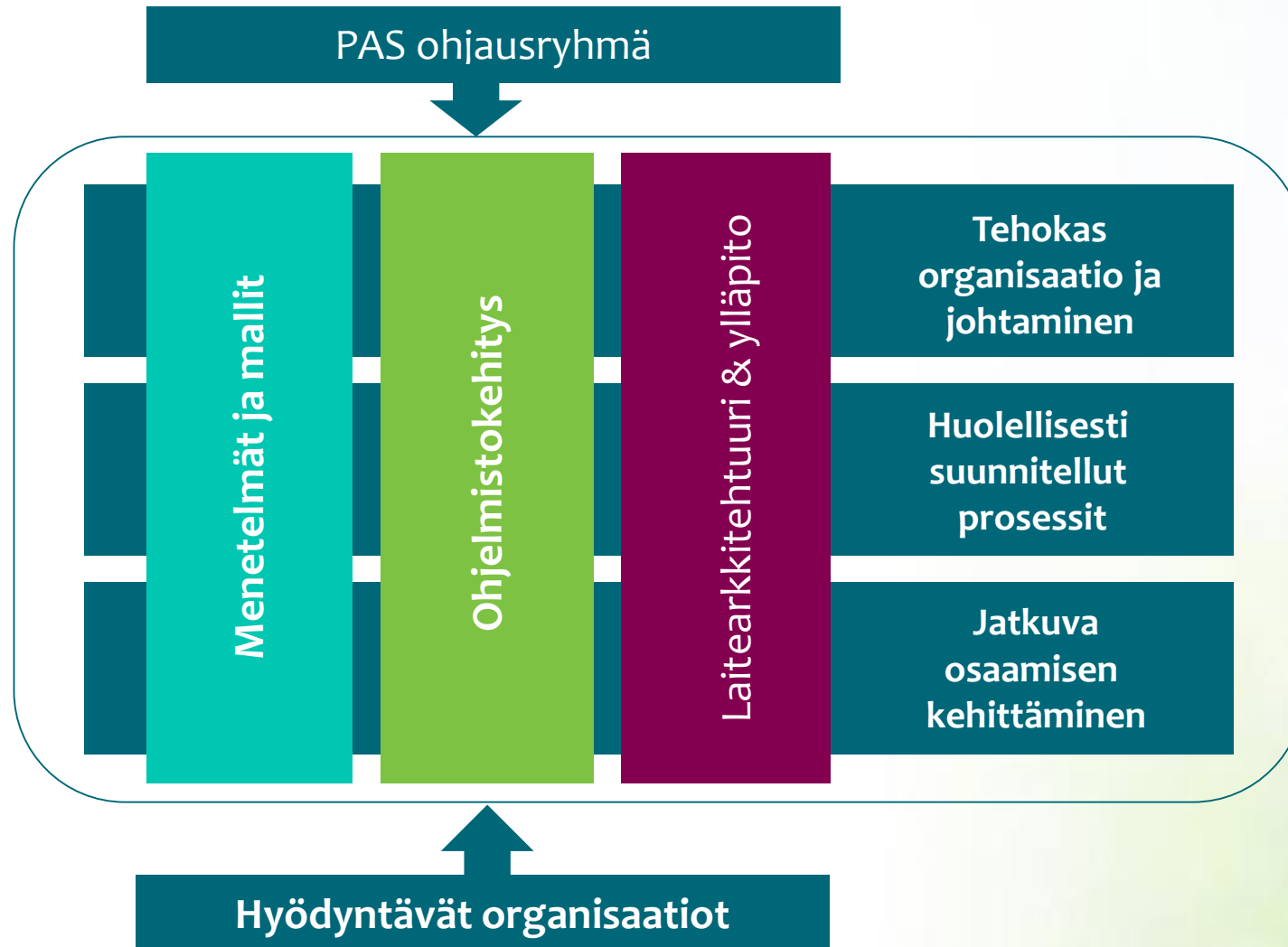


Miten tehdään PAS-työtä? (1/2)

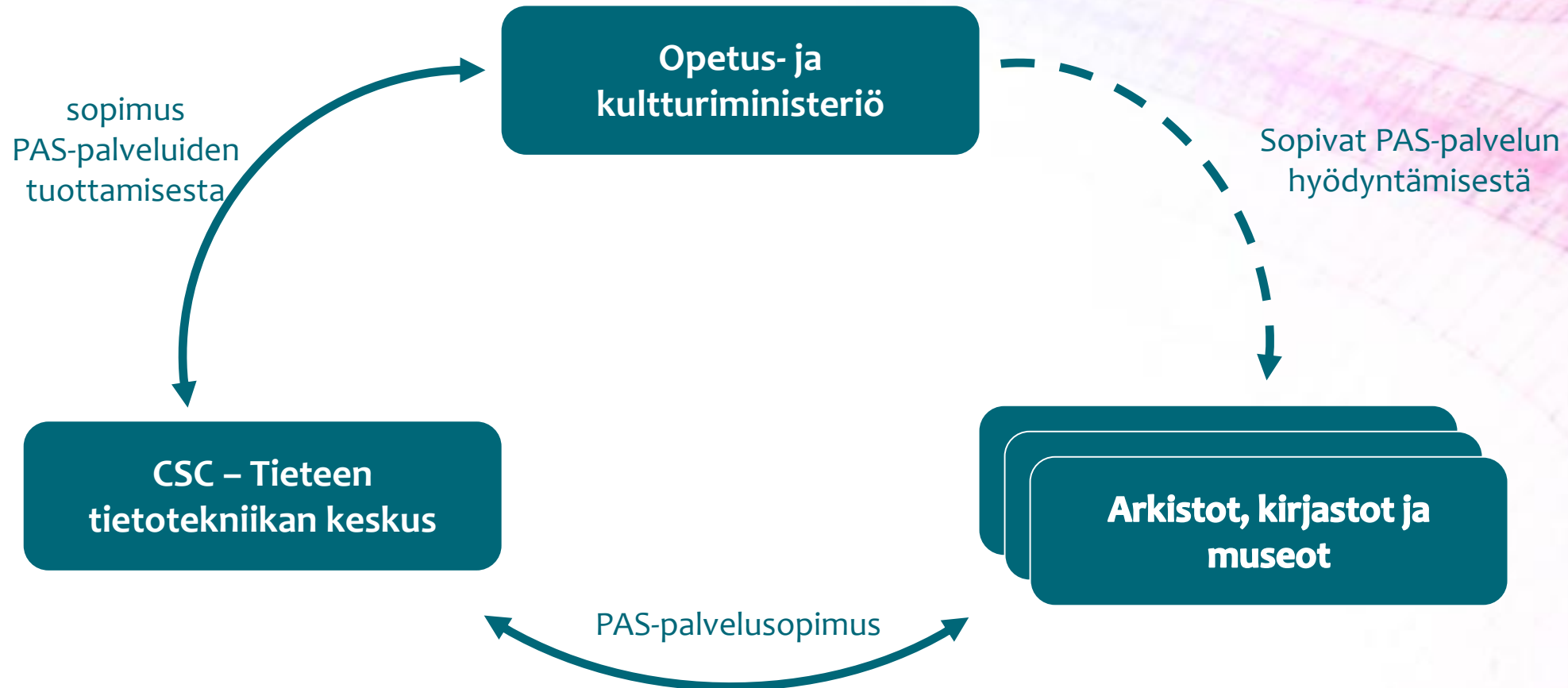
- Kehitetään toimintaan tarvittavia kyvykkyyksiä
- Keskitytään
 - Prosessien hiomiseen
 - Osaamisen kehittämiseen
 - Toimivaan työn organisointiin
- Toiminnan pitkäkestoisuus huomioitava kaikessa tekemisessä
- Henkilöriippumattomuus yksi ohjaava tekijä



Miten tehdään PAS-työtä? (2/2)



Pitkäaikaissäilytyksen sopimukset



PAS-palvelun hyödyntäjät

Organisaatio	Käyttötarkoitus	Kapasiteetti
Kansallinen audiovisuaalinen instituutti	Valikoitu osa kotimaisen elokuvan digitoitavista aineistoista	320 TT
Kansallisarkisto	Kansallisarkiston vastaanottamat alkujaan digitaaliset valtionhallinnon asiakirjalliset aineistot	41 TT
Kansallisarkisto	VAPA-järjestelmään siirretyt tietoaineistot	1 TT
Kansalliskirjasto	Kansalliskirjaston digitoimat kulttuuriperintöaineistot	175 TT
Kansalliskirjasto	Kulttuuriaineistolain nojalla kerätyt aineistot	165 TT
Kotimaisten kielten keskus Kotus	Kotuksen kielentutkimus- ja kulttuuriperintöaineistojen pitkäaikaissäilytys	60 TT
Museovirasto	Kulttuuriympäristön tutkimusraportit	1 TT
Musiikkiarkisto	Musiikkiarkiston pitkäaikaissäilytettävät aineistot	70 TT
Svenska Litteratursällskapet SLS	SLS:n pitkäaikaissäilytettävät aineistot	50 TT
Yhteiskuntatieteellinen tietoarkisto	Tietoarkiston arkistoimien tutkimusaineistojen kokoelman pitkäaikaissäilytys	1 TT

Mitä hyödyntäminen edellyttää organisaatiolta?

- PAS-palvelun hyödyntämisestä sopimista OKM:n kanssa
- Säilytettävän aineistokokonaisuuden tunnistamista
- Käyttöönottoprosessin läpikäymistä yhdessä CSC:n kanssa
- Aineistojen riittävää kuvailua
- Aineistojen paketointia PAS-määritysten mukaisesti
- Aineistojen siirtämistä PAS-palveluun
- Henkilöresursseja palvelun käyttöönottoon/käyttöön

PAS-palvelu
tukee

PAS-palveluiden tulevaisuus

- Laajennetaan säilytystä OKM:n päätöksien mukaisesti
- Jatketaan yhteistyötä organisaatioiden kanssa
 - Nykyiset ja tulevat hyödyntävät organisaatiot sekä muut halukkaat
- Täydennetään loogisen säilyttämisen ominaisuuksia organisaation tarpeiden mukaisesti
 - Mm. migraatioiden hallinta
- Auditoidaan PAS-palvelut laajemmin
- Parannetaan organisaatioille tarjottavaa tukea
- Vuonna 2019:
 - Kulttuuriperintö-PAS-palvelun kapasiteetti noussut 800 teratavuun
 - Säilytys on hajautettu kolmeen sijaintiin (nykyisen kahden sijaan)
 - Ensimmäiset laitetason virkistykset suoritettu
 - Kehitetään palvelua organisaatioiden tarpeiden mukaisesti (mm. tiedostomuotojen tuki)
 - Paljon lisää aineistoja säilytykseen

Kiitos!

PAS-support@csc.fi
digitalpreservation.fi