



Fairdata PAS-palvelu

Juha Lehtonen

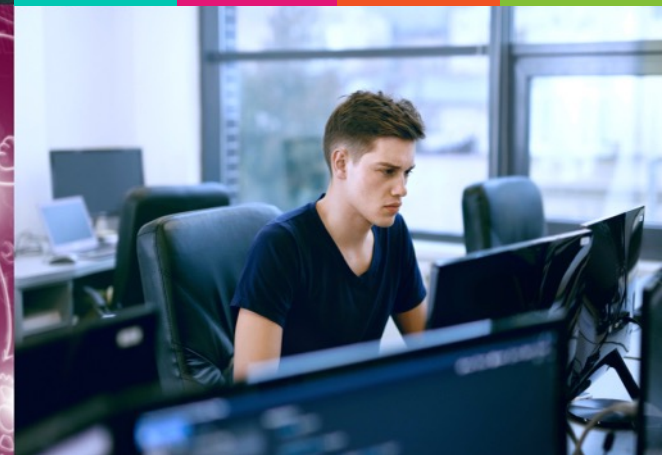
2019-11-15



Sisällys

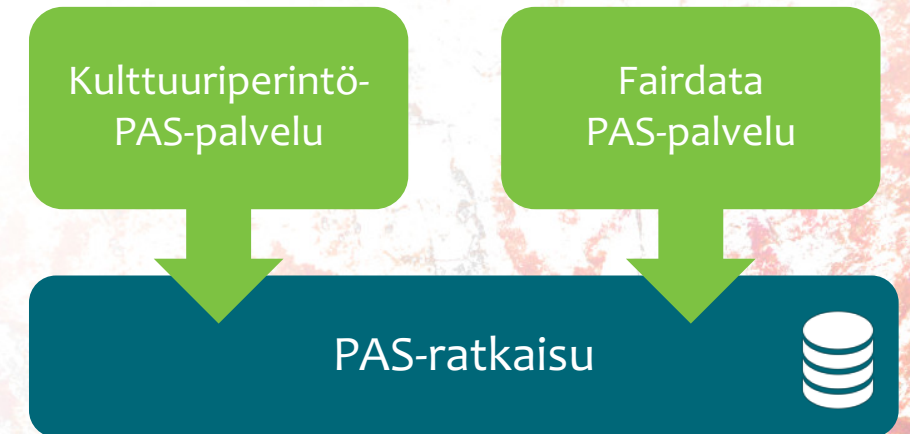
- Pitkäaikaissäilytys ja PAS-palvelut
- Vaihtoehdot aineistojen paketointiin ja siirtämiseksi PAS-palveluun
- Pitkäaikaissäilytyksen määritykset
- Työkalut
 - Paketointikomponentti
 - Hallintaliittymä

Pitkäaikaissäilytys ja PAS-palvelut



Säilyttämisen palvelut ja PAS-ratkaisu

- CSC tuottaa OKM:lle keskitettyjä PAS-palveluita
 - Kulttuuriperinnön säilyttämiseen (tuotannossa vuodesta 2015)
 - Tutkimusaineistojen säilyttämiseen (tuotannossa vuoden 2019 aikana)
- Pohjautuvat yhteiseen bittitason säilytyksen ratkaisuun
- Taustalla OKM:n KDK- ja ATT-hankkeet, joissa laaditut arkkitehtuurit tunnistavat säilyttämisen tarpeet ja asemoivat PAS-palvelut
- Tavoitteena on mahdollistaa digitaalisten aineistojen käyttökelpoisuus, autenttisuus ja eheys kymmenien tai jopa satojen vuosien ajan
- CSC on ottanut säilytysvastuun lähes 500 teratavusta ja 1,4 miljoonasta tietopaketesta
- PAS-palvelut sisältyvät CSC:n ISO/IEC 27001:2013-auditoinnissa tarkasteltuihin palveluihin

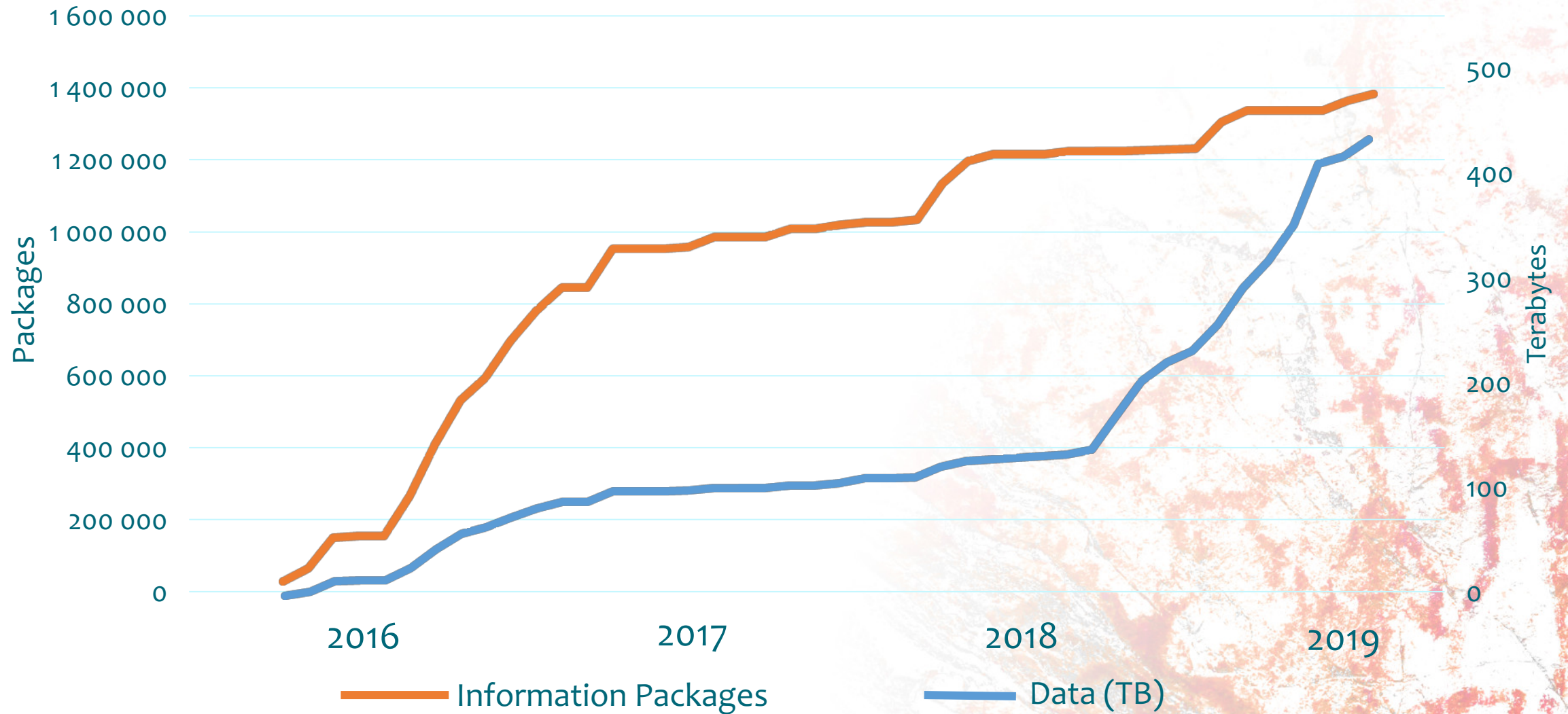


OKM:n PAS-palveluita hyödyntävät organisaatiot

- Celia
- Kansallinen audiovisuaalinen instituutti
- Kansallisarkisto
- Kansalliskirjasto
- Kotimaisten kielten keskus Kotus
- Museovirasto
- Musiikkiarkisto
- Svenska litteratursällskapet i Finland (SLS)
- Yhteiskuntatieteellinen tietoarkisto

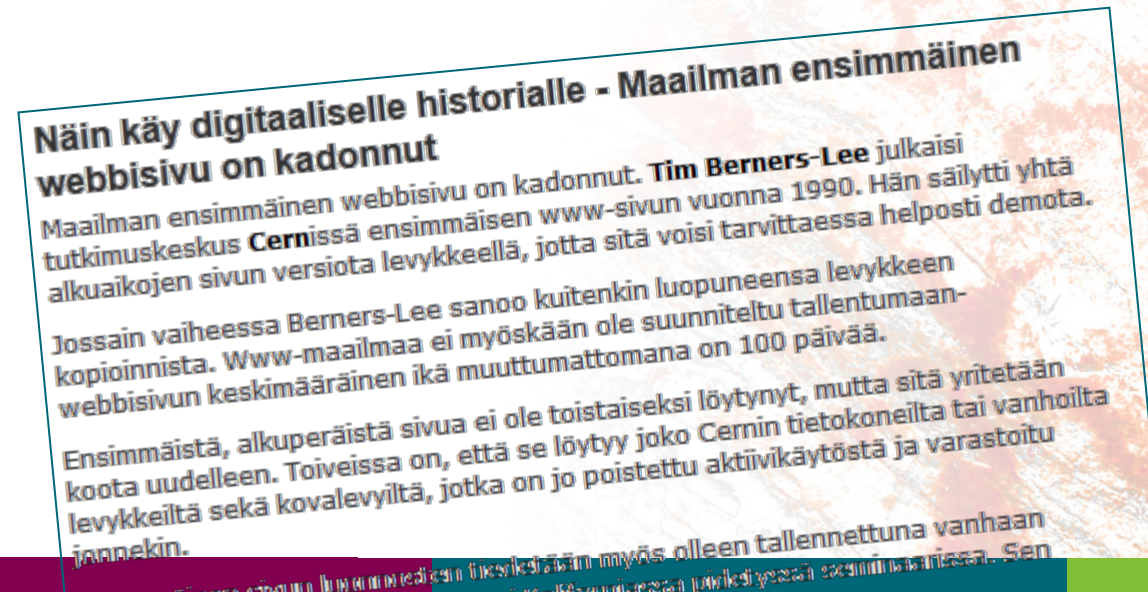
Vuoden loppuun mennessä aloitetaan tutkimusaineistojen säilyttäminen, mikä täydentää listaa

Data in digital preservation 2016-2019



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



Pitkääikaissäilytyksen uhkia

- Media failure
- Hardware failure
- Software failure
- Communications error
- Failure of network services
- Media & hardware obsolescence
- Software obsolescence
- Operator error
- Natural disaster
- External attack
- Internal attack
- Economic failure
- Organizational failure
- ...

Source: David Rosenthal, Thomas Robertson, Tom Lipkis, Vicky Reich, Seth Morabito. Requirements for Digital Preservation Systems: A Bottom-Up Approach. D-Lib Magazine 11(11), 2005.

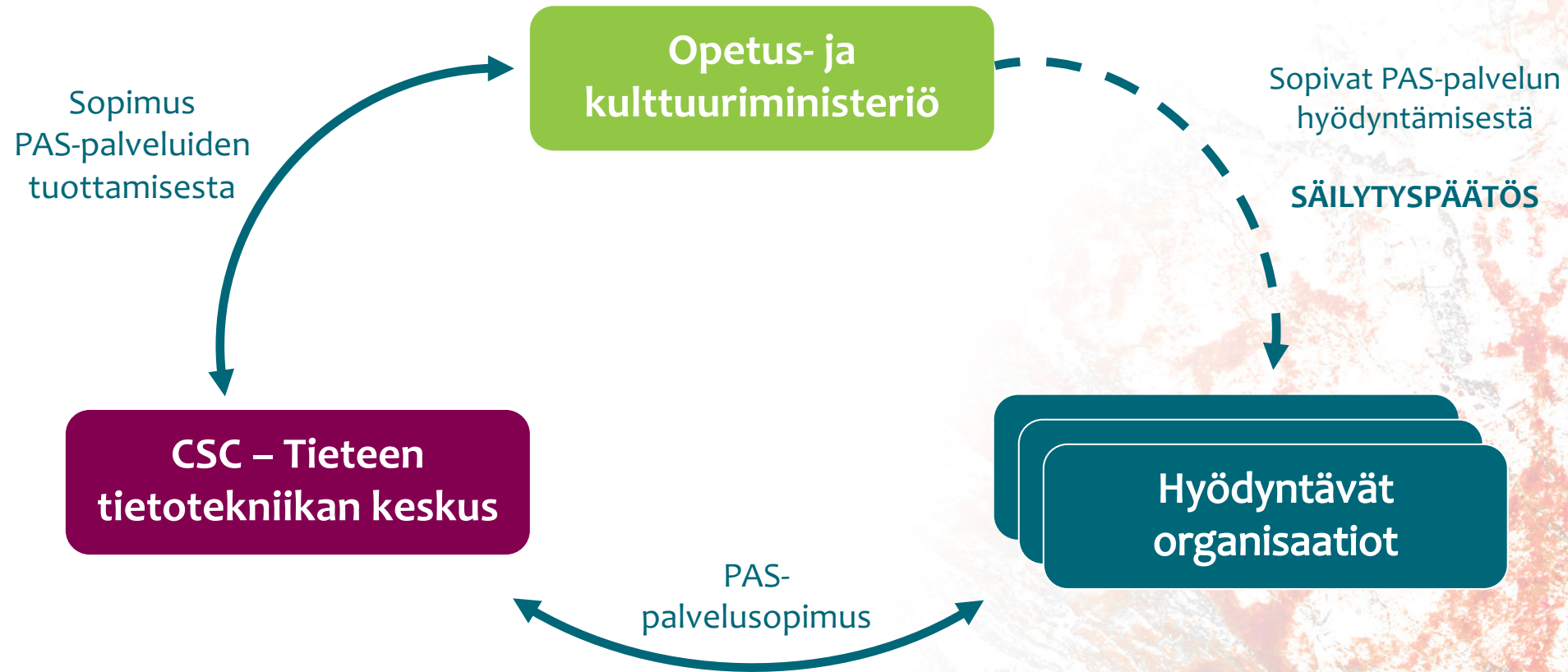
- Storage
- Integrity
- Control
- Metadata
- Content

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

Yhteistyö mahdollistaa aineistojen pitkäkestoisen hyödyntämisen



Pitkäaikaissäilytyksen sopimukset

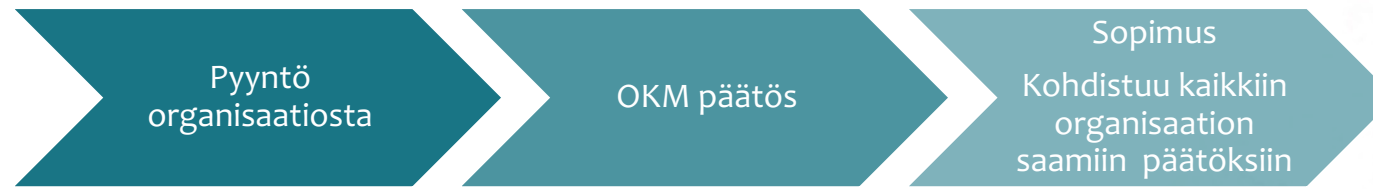


- Useita säilytyspäätöksiä ja näihin liittyviä käsittelytoimien kuvauksia

Fairdata PAS -sopimus



Kukin PAS-palvelua hyödyntävä organisaatio kirjoittaa vain yhden, organisaation kattavan PAS-sopimuksen. Tähän sopimukseen voidaan liittää useita, eri aineistoihin kohdistuvia OKM:n säilytyspäätöksiä sekä kullekin näistä eri käsittelytoimien kuvaus.



- 1) Organisaation virallisen edustajan (PAS-yhteyshenkilö) pyyntö postilistalle fairdata-okm@postit.csc.fi. Viestissä voi olla liitteenä täytetty pyyntölomake (saatavilla Fairdata-palveluiden verkkosivuilta) tai vastaavat tiedot sähköpostissa. **Organisaation oma arvio kansallisesta merkittävydestä riittää perusteeksi.**
 - o OKM:n hyväksynnän jälkeen CSC aloittaa sopimuksen laadinnan
- 2) OKM ilmoittaa organisaatiolle ja CSC:lle päätöksestään, ja että sopimus voidaan solmia. Säilytyspäätös julkaistaan.
- 3) CSC on yhteydessä organisaation PAS-sopimusyhteyshenkilöön, ja sopimus solmitaan.
- 4) Käyttöönotto voi alkaa
- 5) **Organisaation seuraava käyttöoikeuspäätös ei edellytä uutta sopimusta, vaan päätös liitetään organisaation PAS-sopimukseen**

<https://www.fairdata.fi/hyodyntaminen/kayttoonotto/>

CSC auttaa organisaatioita säilyttämisen kaikissa vaiheissa

- Tämä tuki on merkittävä osa organisaatioille tarjottavaa palvelua
- Kokemus osoittanut, että organisaatiot tarvitsevat usein tukea käyttöönotoissa
- Emme jätä organisaatioita yksin missään
PAS-palveluihin liittyvässä asiassa

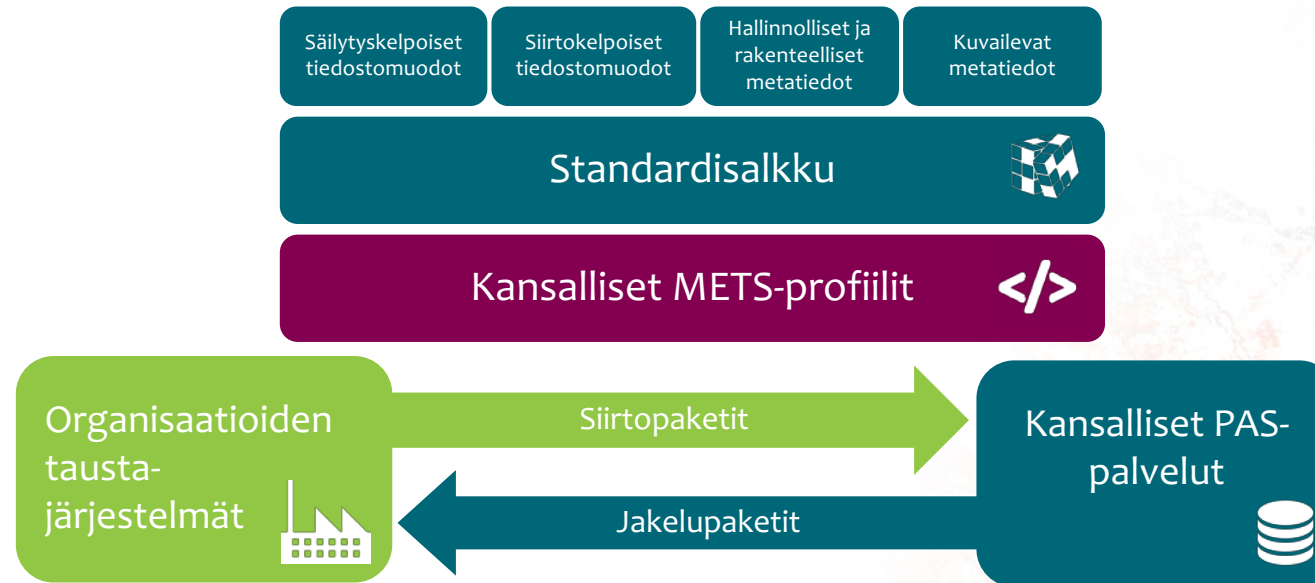
pas-support@csc.fi



Vaihtoehdot aineistojen paketointiin ja siirtämiseksi PAS-palveluun



Hyödyntäviä organisaatioita ohjaavat PAS-määritykset



- Näkyvin osa PAS-palvelusta hyödyntäville organisaatioille
 - Erityisesti jos organisaatio paketoit itse aineistonsa
- PAS-määritykset on tehty tiiviissä yhteistyössä hyödyntävien organisaatioiden kanssa
 - Ja näin myös jatketaan tulevaisuudessa

Paketoinnin vaihtoehdot

ORGANISAATIO PAKETOI ITSE

A. Paketointikomponentin hyödyntäminen

✓ Paljon aineistoa

- + Melko joustava
- + Ylläpidettävyyys määritysten muuttuessa
- + Vähentää tarvetta METS- ja PREMIS-tuntemukselle
- Jonkin verran omaa koodausosaamista
- Määritysten ymmärtäminen

B. Organisaation räätälöimä paketointi

✓ Paljon aineistoa; monimutkaisia tarpeita

- + Joustava ja skaalautuva
- (Runsaasti) omaa koodausosaamista
- Määritysten syvälinen ymmärtäminen
- Ylläpidettävä itse määritysten muuttuessa

PAKETOINTIPALVELUN HYÖDYNTÄMINEN

C. Väliaikainen tiedostojen säilytyspalvelu

✓ Aineistoa vähän; yksinkertainen rakenne

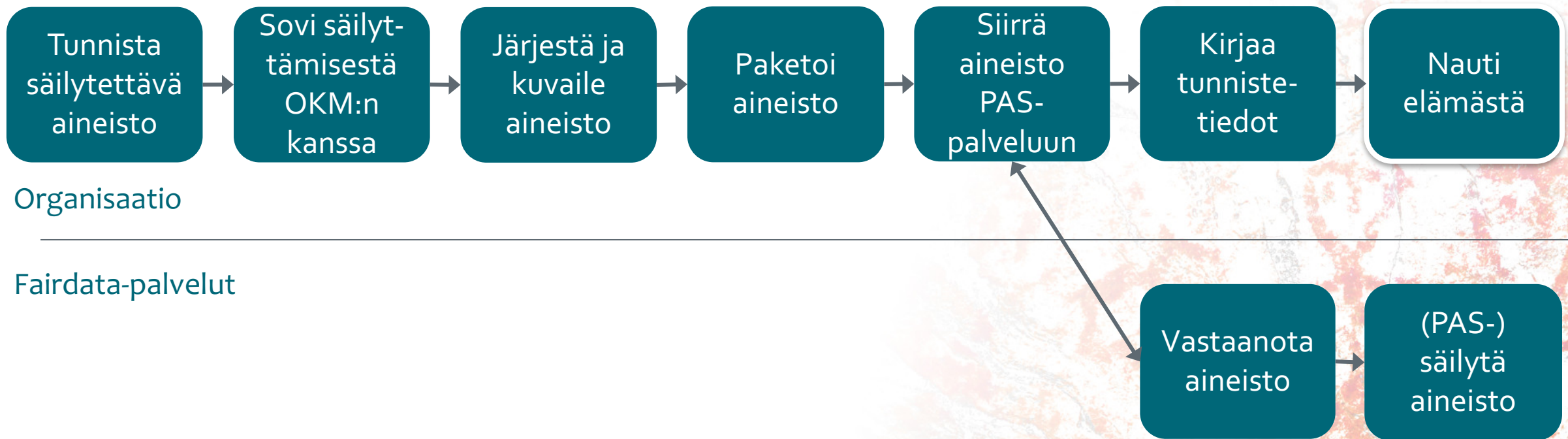
- + Valmiit moduulit
- + Ylläpidettävyyys määritysten muuttuessa
- + Ei vaadi esim. METS- ja PREMIS-tuntemusta
- Saattaa vaatia hieman omaa koodausosaamista (tilanteesta riippuen)
- Joustamaton

D. IDA-säilytyspalvelu

✓ Aineistot IDA:ssa

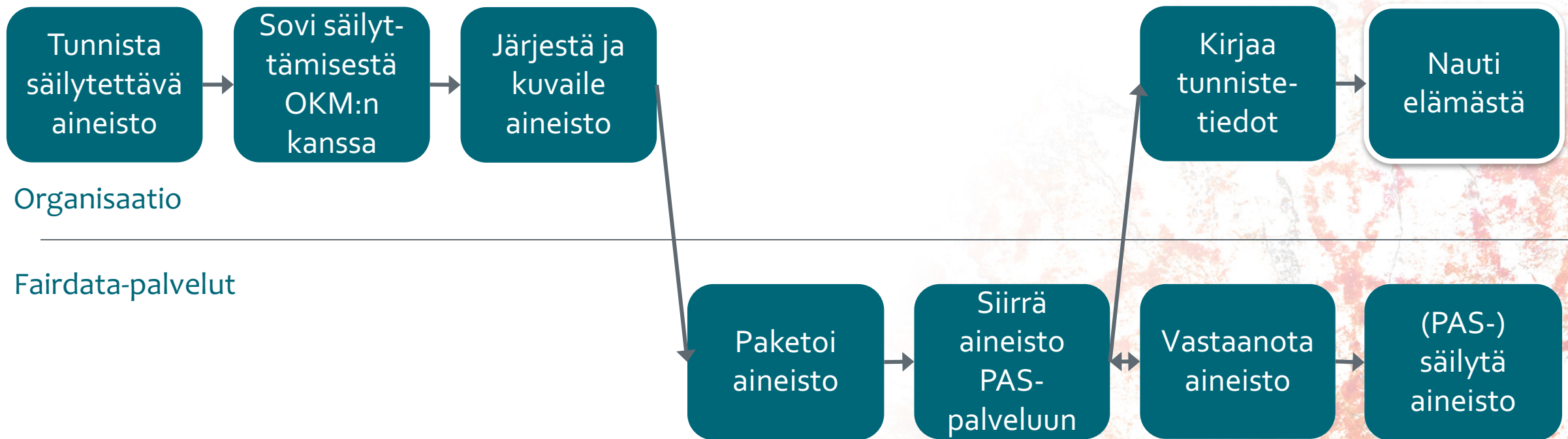
- + Ei omaa koodausosaamista
- + Ylläpidettävyyys määritysten muuttuessa
- + Ei vaadi esim. METS- ja PREMIS-tuntemusta
- Joustamaton

Esimerkkiprosessi tutkimusaineistojen siirtämisestä PAS-palveluun



- A. Paketointikomponentin hyödyntäminen
- B. Organisaation räätälöimä paketointi

Esimerkkiprosessi tutkimusaineistojen siirtämisestä PAS-palveluun

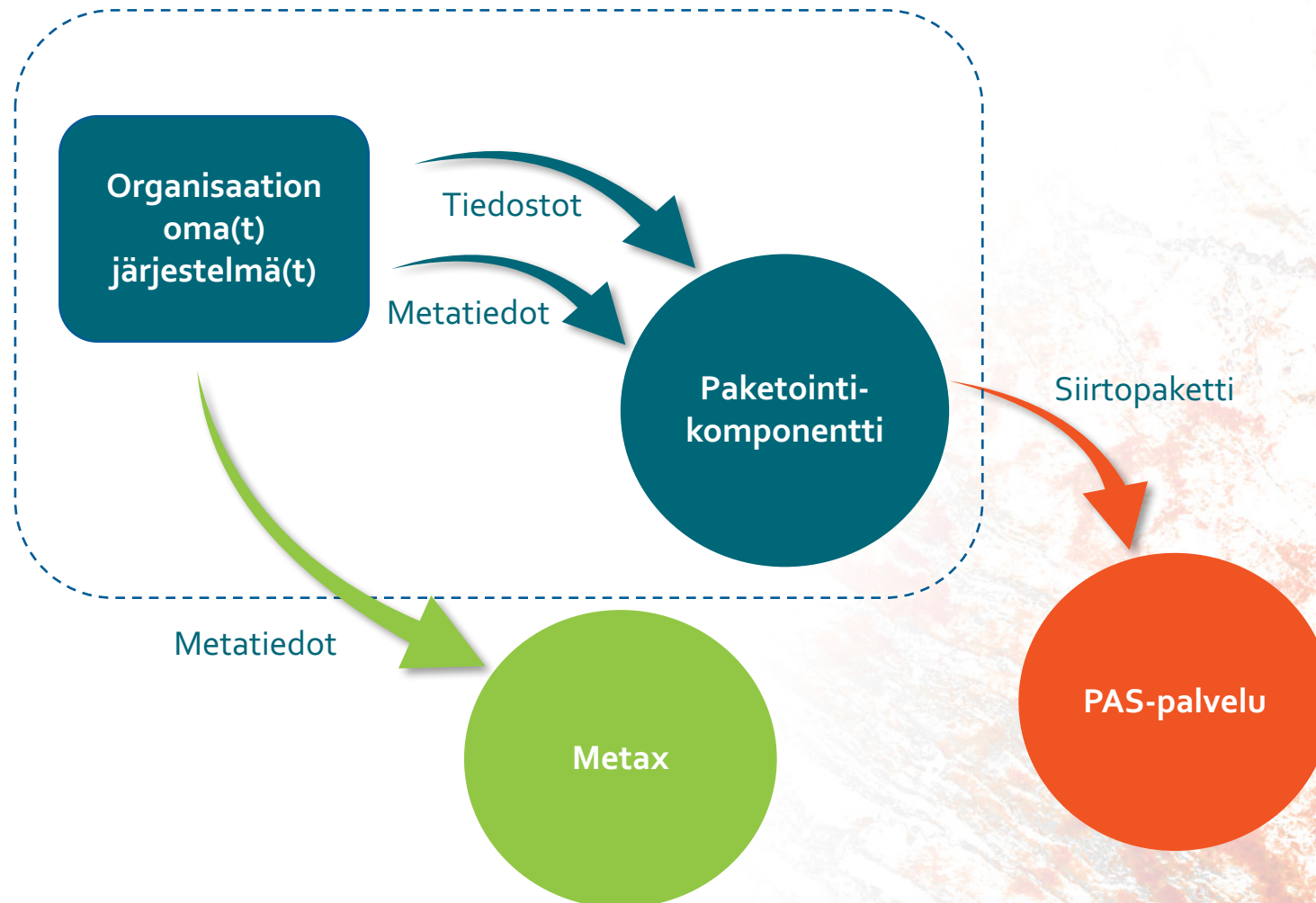


C. Väliaikainen tiedostojen säilytyspalvelu

D. IDA säilytyspalvelu

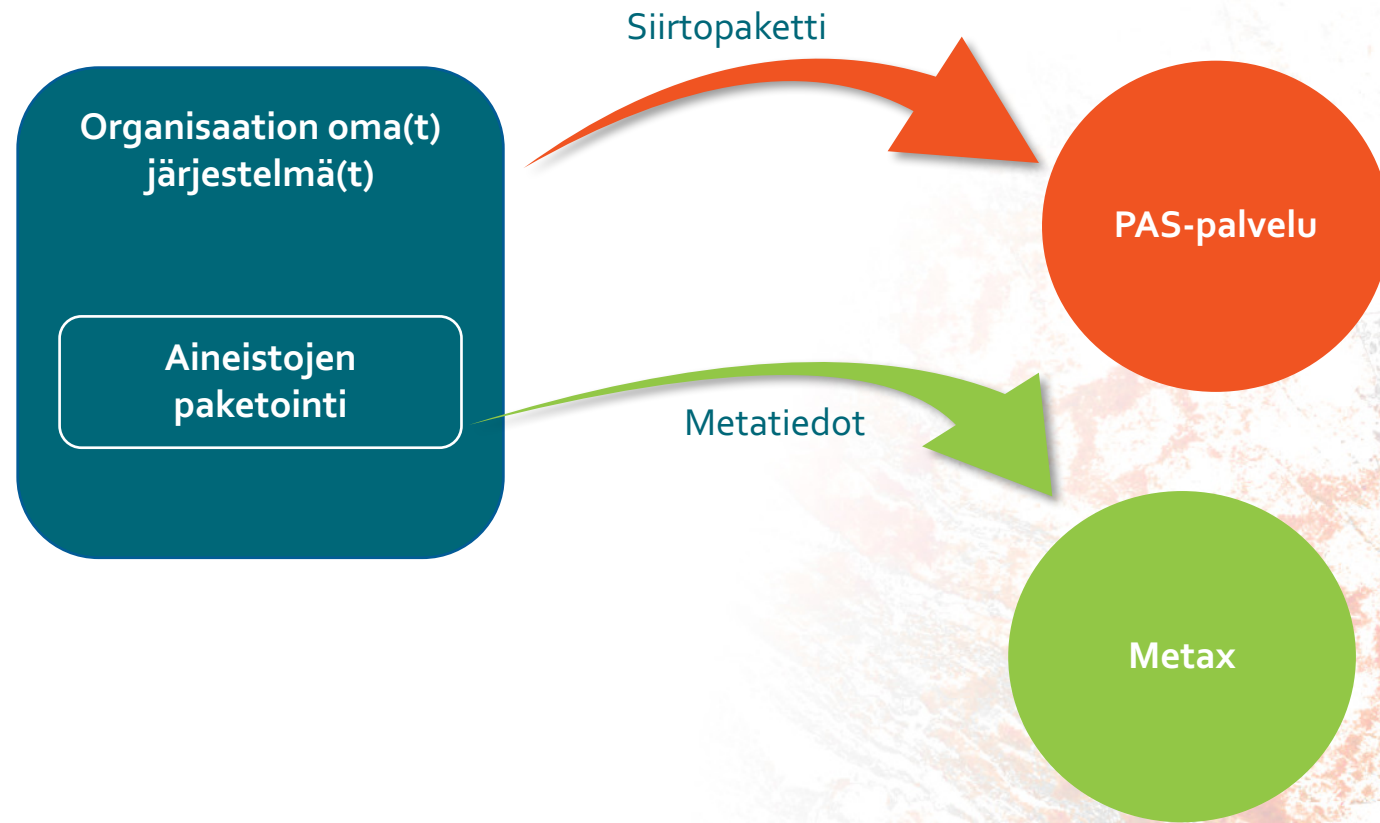
Aineistojen paketointi

A. Paketointikomponentin hyödyntäminen



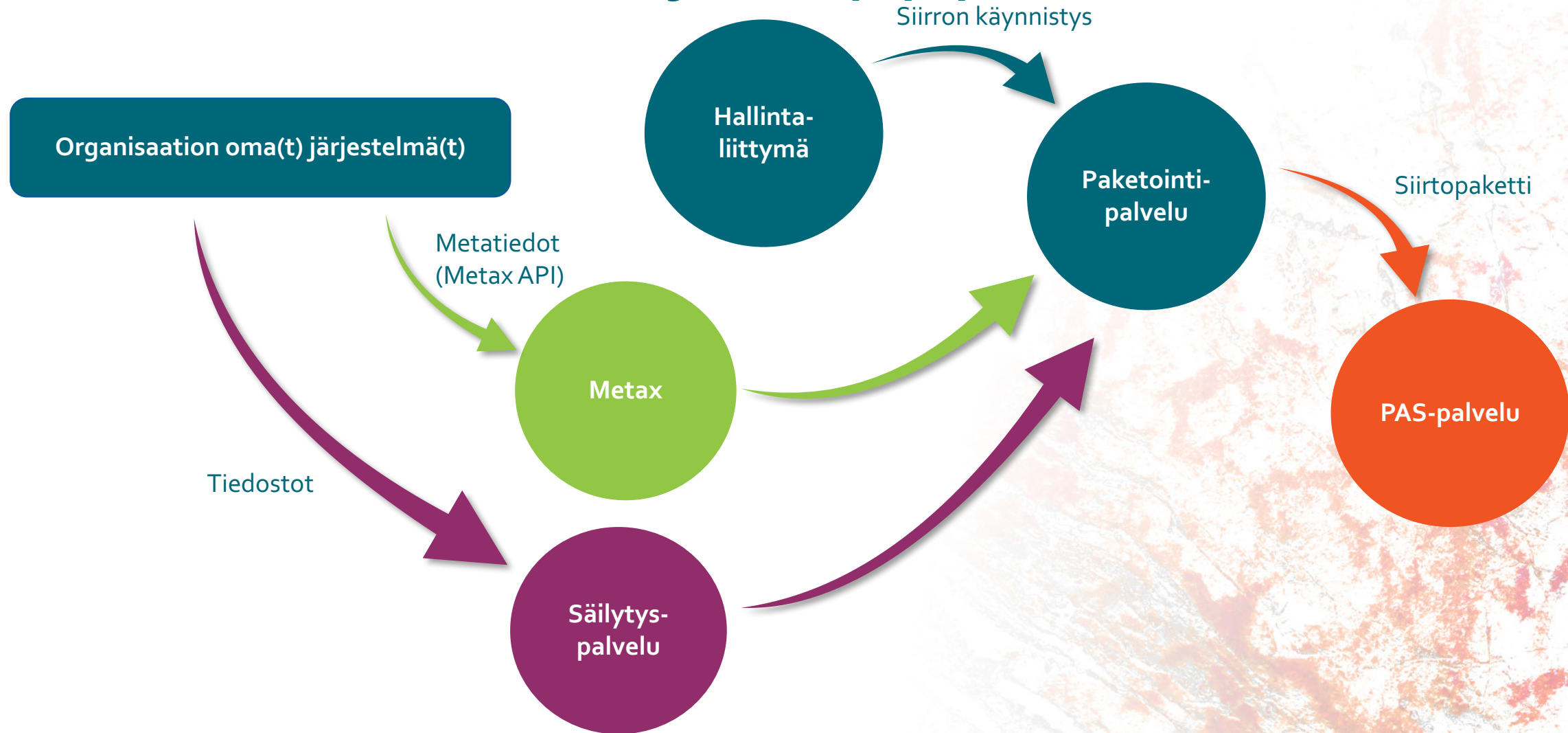
Aineistojen paketointi

B. Organisaation räätälöimä paketointi



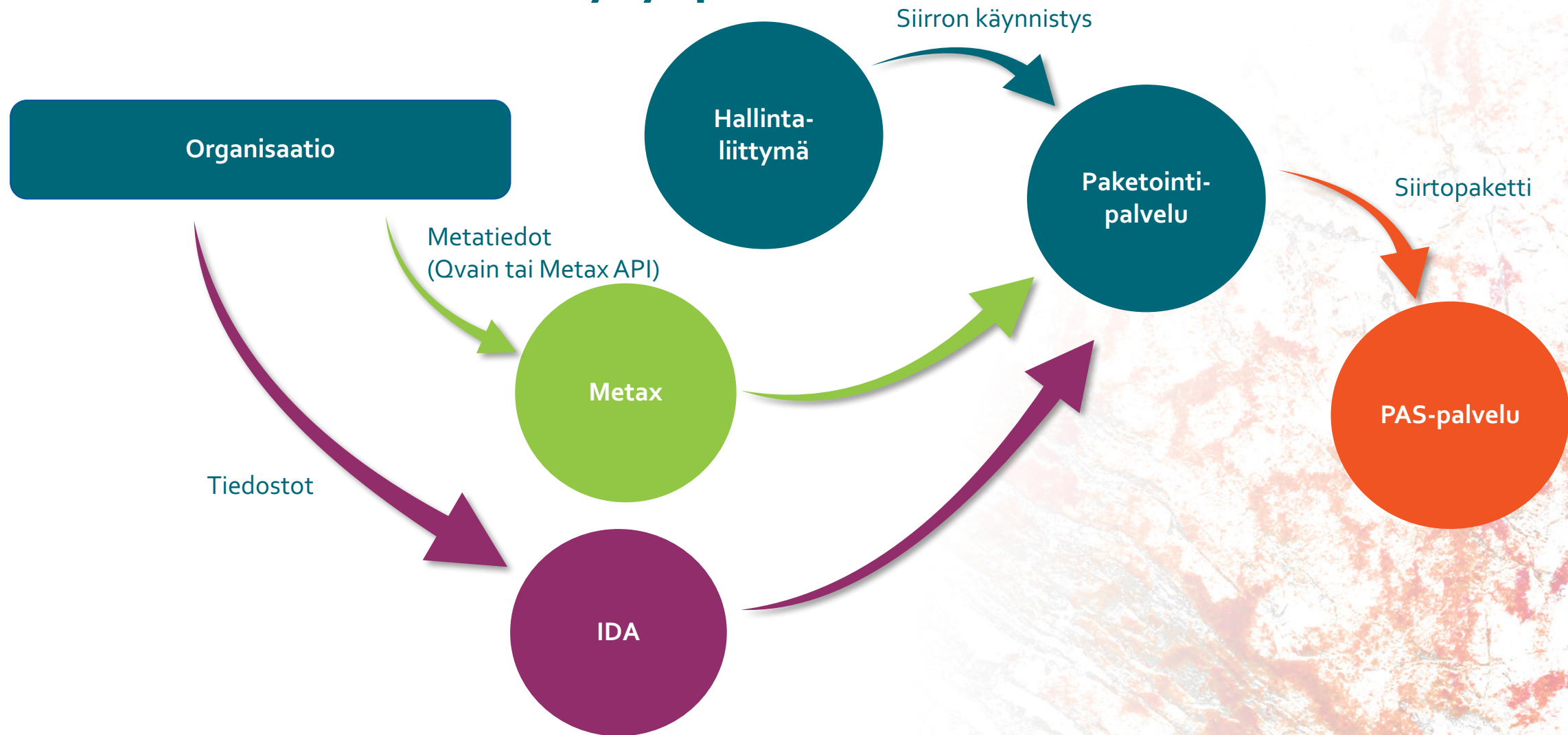
Aineistojen paketointi

C. Väliaikainen tiedostojen säilytyspalvelu



Aineistojen paketointi

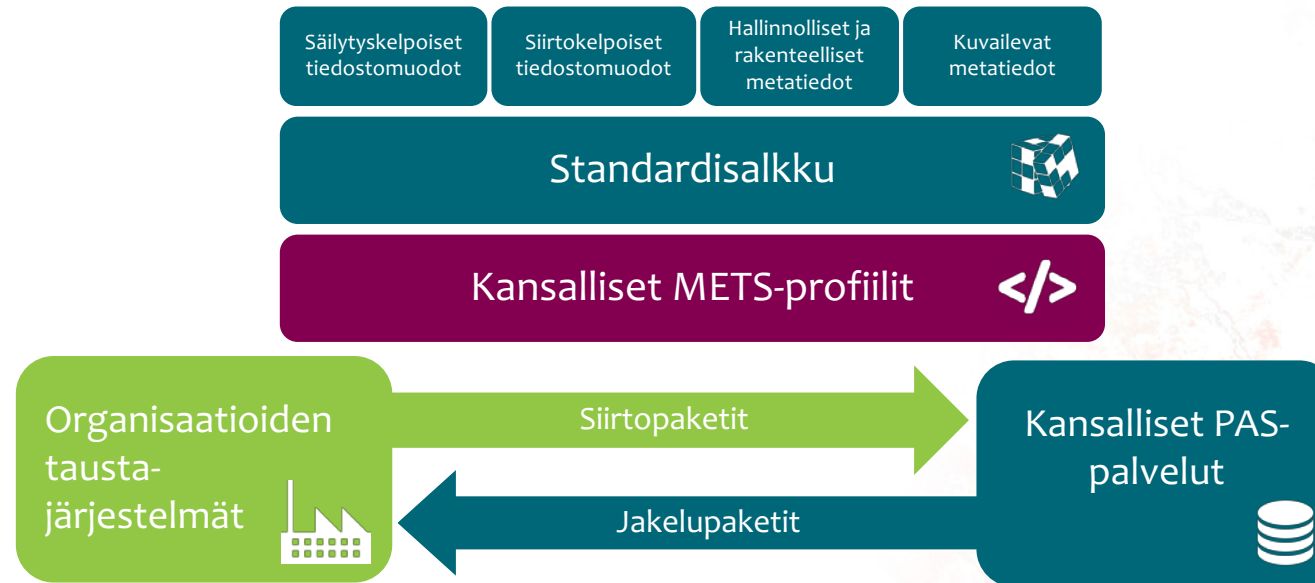
D. Aineistot IDA-säilytyspalvelussa



Pitkääaikaissäilytyksen määrittäykset



Hyödyntäviä organisaatioita ohjaavat PAS-määrittelykset



- Standardisalkku
- Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen
- Säilytys- ja siirtokelpoiset tiedostomuodot
- PAS-palveluiden rajapinnat

Määriykset

Standardisalkku

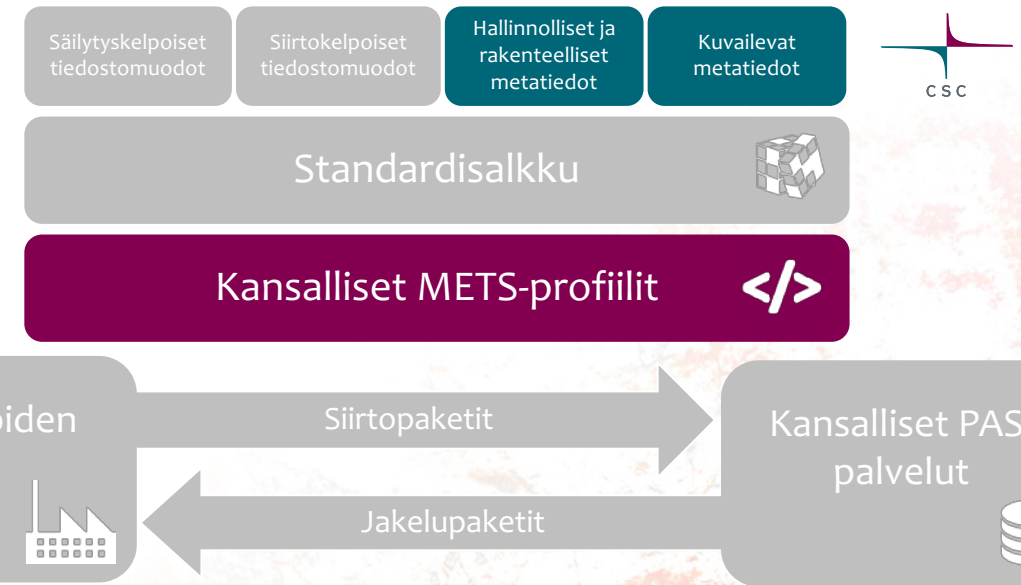
- Kuvaa yhteisesti käytetyt standardit ja määriykset perusteluineen
 - Metatiedot, rajapinnat, soveltamisohjeet, ...
- Mahdollistaa yhtenäisten palveluiden tarjoamisen ja aineistojen yhdisteltävyyden sekä metatietojen kehittämisen
 - Semanttinen yhteismitallisuus



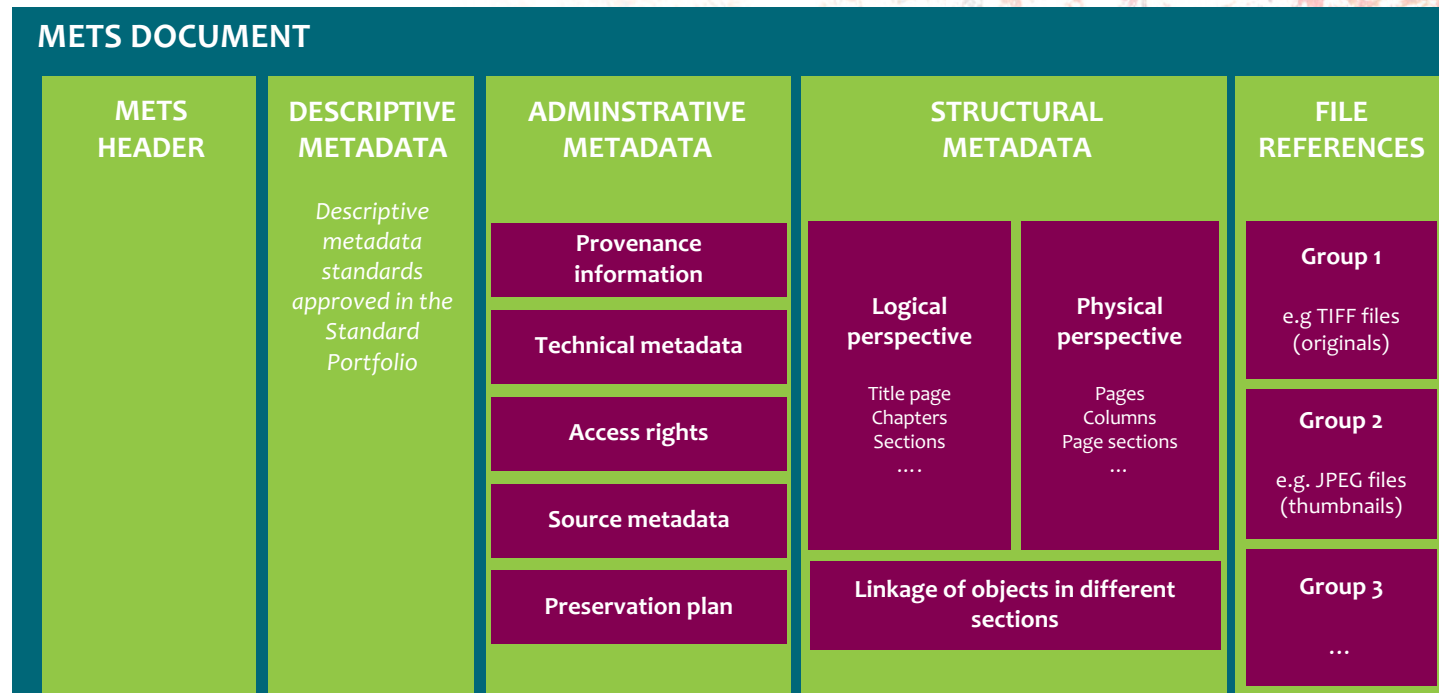
Määritykset

Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen

- Määrittelee
 - Pakolliset ja ehdollisesti pakolliset metatiedot
 - Kuvailevat, hallinnolliset ja rakenteelliset metatiedot
 - Siirtopaketin (SIP) ja jakelupaketin (DIP) teknisen rakenteen



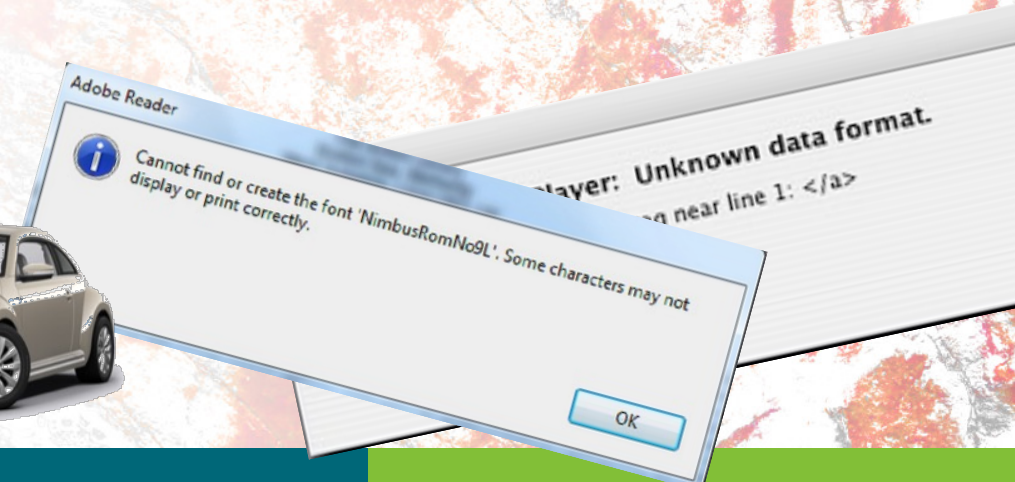
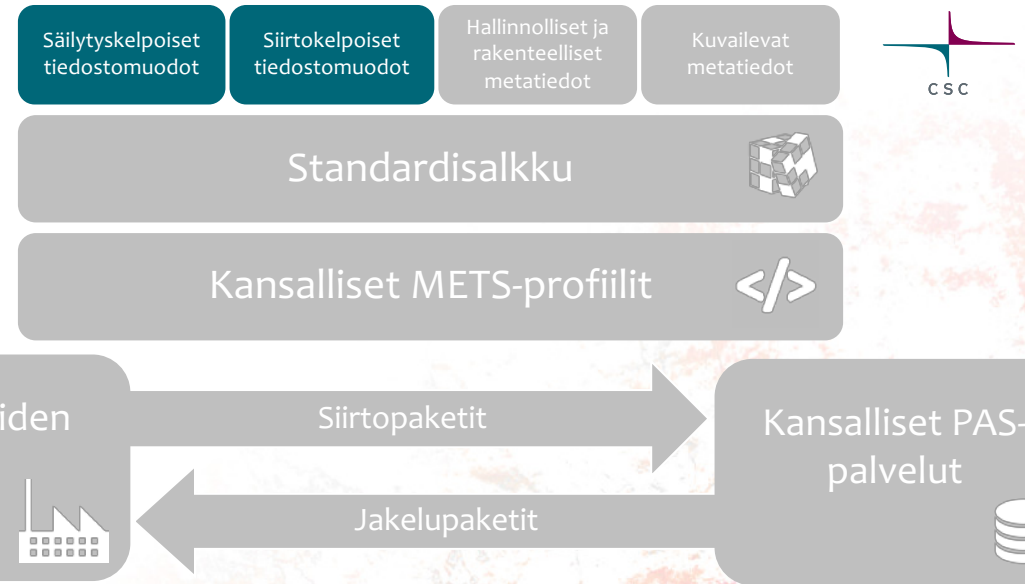
- Perustuu METS ja PREMIS XML-standardeihin
- Vaatimukset mahdollistavat aineistojen automaattisen vastaanoton
- Fairdata PAS-paketointipalvelu toteuttaa varsinaisen paketoinnin



Määritykset

Säilytys- ja siirtokelpoiset tiedostomuodot

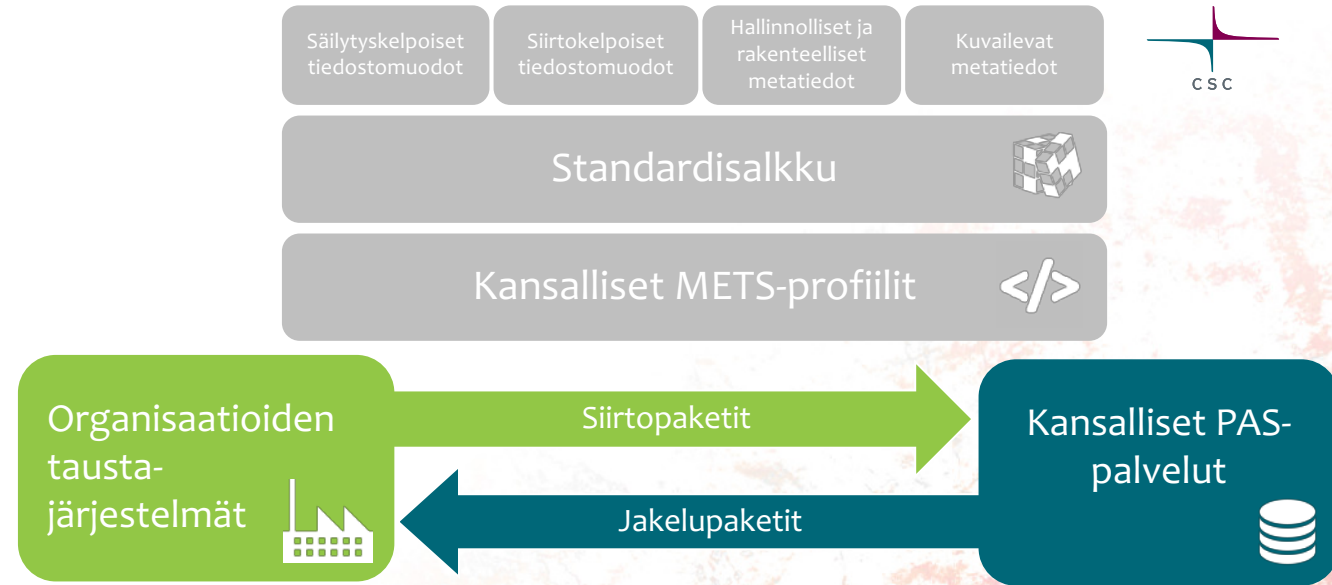
- Määrittelee
 - Yleiset vaatimukset tiedostomuodoille ja kriteeristön tiedostomuotojen arvioimiseen
 - Säilytyskelpoiset tiedostomuodot
 - Siirtokelpoiset tiedostomuodot
 - Kontrolloidun sanaston tiedostomuotojen ilmaisemiseen
 - Vaatimukset teknisille metatiedoille
- Säilytyskelpoiset tiedostomuodot
 - Tiedostomuodot, joiden arvioidaan olevan käyttökelpoisia vielä pitkään
- Siirtokelpoiset tiedostomuodot
 - Tiedostomuodot, joita arvioidaan olevan hyödyntävillä organisaatiolla runsaasti, mutta jotka eivät sovellu pitkäaikaiseen säilyttämiseen



Määritykset

PAS-palveluiden rajapinnat

- Määrittelee
 - Aineistojen siirtämisen PAS-palvelun vastaanottoon
 - Fairdata PAS-paketointipalvelu tekee siirron
 - Aineistojen haun ja hallinnan
 - Siirtopaketin tarkastusraportin
- Perustuu SFTP- ja REST-rajapintoihin



Määritykset päivitetään vuosittain yhteistyössä



Palautteen kerääminen hyödyntäviltä organisaatioilta

- Epäviralliset palautteet ja virallisempi palautekysely



Päätökset muutoksista

- Ovatko muutokset välttämättömiä juuri nyt?
- Perustelut



Määritysten päivittäminen yhteistyössä hyödyntävien organisaatioiden kanssa

- Hyväksyntä yhteistyöryhmässä



Uusien versioiden julkaiseminen

- XML-skeemat
- Muu dokumentaatio

Määritykset saatavilla

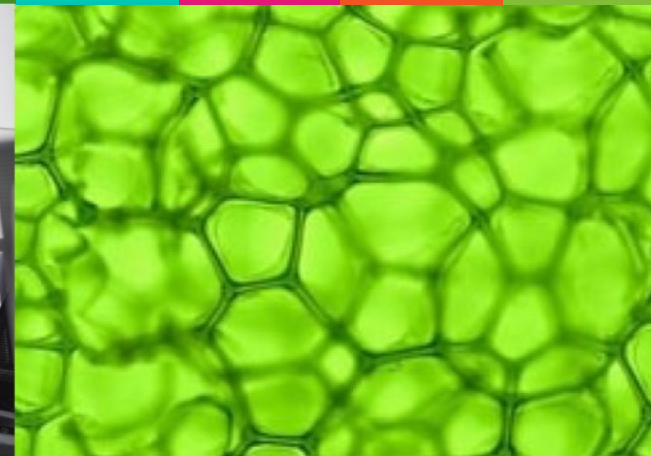
Myös englanniksi



- <http://digitalpreservation.fi/specifications>
 - Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen (v.1.7.1)
 - Säilytys- ja siirtokelpoiset tiedostomuodot (v.1.7.0)
 - PAS-palveluiden rajapinnat (v.2.1.0)
- Ohjelmistokoodia saatavilla @GitHub
 - <https://github.com/Digital-Preservation-Finland/>



Paketointikomponentti



Paketointikomponentti

- Työkalu kansallisten PAS-määritysten mukaisten siirtopakettien muodostamiseksi
- Saatavilla GitHubista:
 - <https://github.com/Digital-Preservation-Finland>
 - Lisenssi: LGPLv3
- **Paketointikomponenttikoulutus 26.11.2019**
 - Ilmoittautuminen käynnissä 20.11. asti:
<http://digitalpreservation.fi/>
- Ei tarvita, jos hyödyntävä organisaatio hyödyntää Fairdata PAS-paketointipalvelua



Pakentointikomponentti: Paketoinnin vaiheet



Tekninen metatieto
Perusmetatiedot (myös
tiedostomuotovalidointi)



**Tiedostomuotokohtainen
tekniinen metatieto**
Kuva, ääni, video,
rakenteellinen teksti

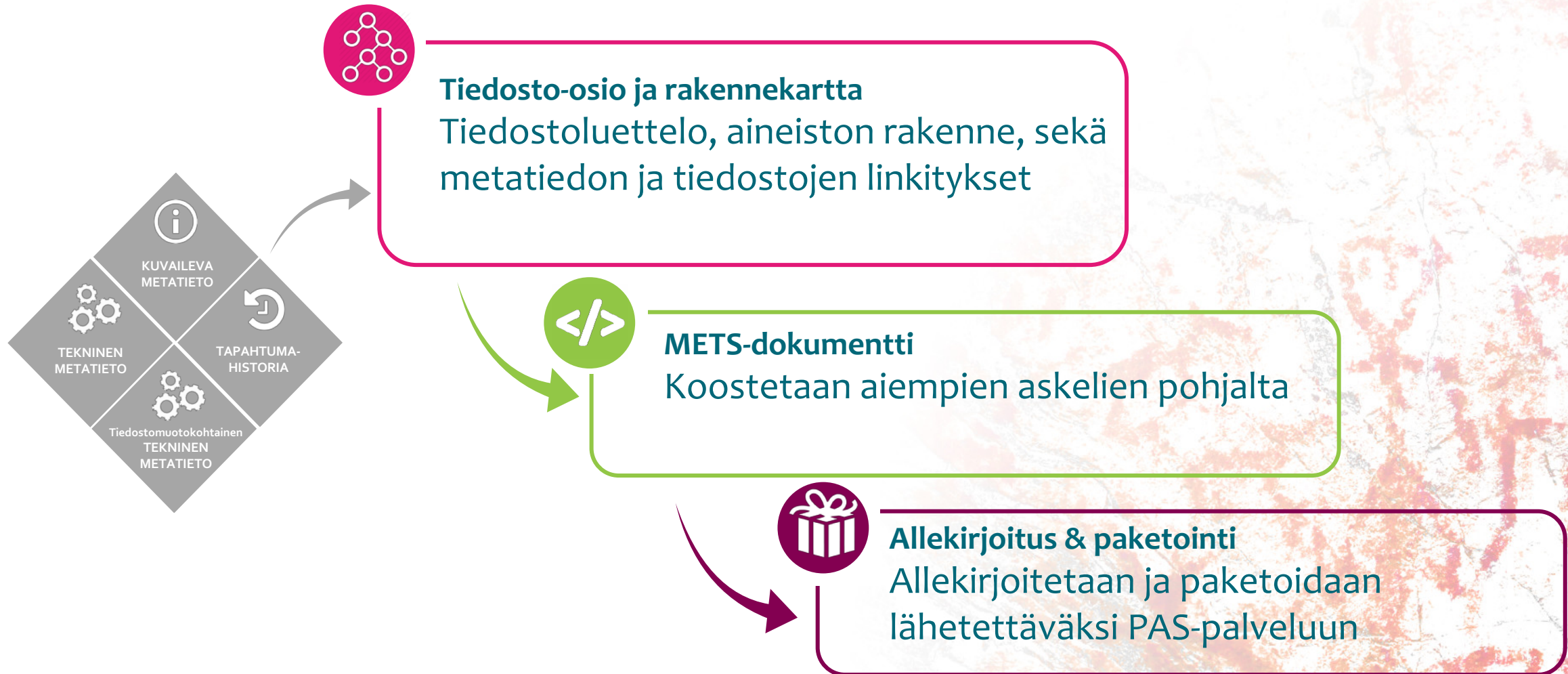


Kuvaileva metatieto
Valmiin XML-muotoisen
kuvailumetatiedon tuonti

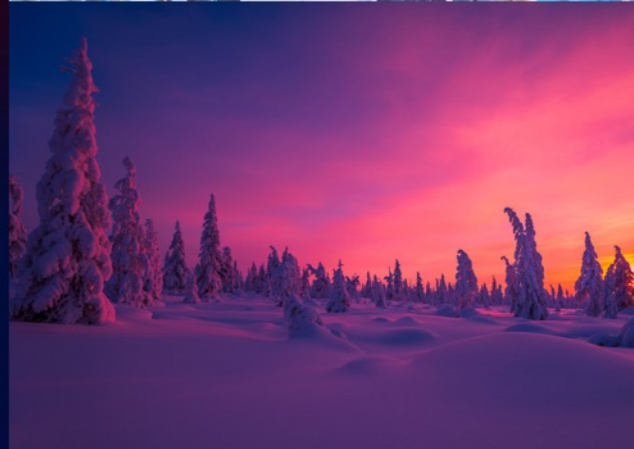


Tapahtumahistoria
Aineiston syntyhistoria ja muut
aineistoon kohdistuneet
tapahtumat

Pakentointikomponentti: Paketoinnin vaiheet



Hallintaliittymä



Hallintaliittymä

Aineiston siirto PAS-palveluun



- Aineiston siirtoa PAS-palveluun voi hallita hallintaliittymässä, kun
 - riittävät metatiedot on viety Metaxiin, ja
 - tiedostot on siirretty joko
 - IDA-säilytyspalveluun, tai
 - väliaikaiseen tiedostojen säilytyspalveluun
- Hyödyntävän organisaation eri käyttäjillä voi olla erilaisia rooleja
 - katseluoikeudet, ehdotusoikeudet, hyväksymisoikeudet
- Aineistoa ensin ehdotetaan pitkäaikaissäilytykseen, jonka jälkeen aineiston metadatan voi validoida ja ehdotuksen hyväksyä (siirtyy PAS-palvelun käsittelyyn)

Status ▲	Dataset	Owner	Contract	Created	Size	Action
Initialized	FDPAS test		CSC sopimus	Oct 8, 2019 15:28:16	4 Bytes	 Propose

Hallintaliittymä

Ehdottaminen pitkäaikaissäilytykseen

Dataset: FDPAS test

Set contract for the dataset

Selected contract: CSC sopimus

Select contract from the list

Contract Name	Organisation	Quota
CSC sopimus	CSC	977 MiB
Title of Contract	Mysterious organization	109 KiB
Title of Contract 2	Mysterious organization	109 KiB
Title of Contract 3	Mysterious organization	109 KiB
Title of Contract 4	Mysterious organization	109 KiB

Remove

Select

Close

Next

Reset

CONTRACT

PROPOSE

VALIDATE

CHECK

ACCEPT



Dataset: FDPAS test

Propose dataset for preservation

Clicking the Propose button will create and validate the dataset's technical metadata. Pressing Propose is an idempotent operation, and it can be repeated multiple times.

Dataset:
FDPAS test

Message:

Proposing test data for preservation

Previous

Propose

Close

Next

Reset

CONTRACT

PROPOSE

VALIDATE

CHECK

ACCEPT

Hallintaliittymä

Metadatan validointi

Dataset: FDPAS test

Validate dataset metadata

Dataset already validated

Click 'Validate' button to validate dataset

Dataset validation completed

Metadata validation report

Preservation state: Valid metadata

Details: Metadata passed validation

PreviousValidateCloseNext

Reset

CONTRACTPROPOSEVALIDATECHECKACCEPT



Dataset: FDPAS test

Check metadata

Please check that the following metadata exists in the Qvain service. Fairdata preservation service can't check the existence of required metadata at this time.

Dataset

- Contract ID - identifier for the contract between the organization and the Fairdata digital preservation service
- Digital provenance

Files

- Use category
- File format
- File format version
- Text encoding for text files

PreviousCloseNext

Reset

CONTRACTPROPOSEVALIDATECHECKACCEPT

Hallintaliittymä

Hyväksyminen, paketointi ja siirto pitkäaikaissäilytykseen



Dataset: FDPAS test

Accept dataset

After the dataset is accepted, it will be packaged and submitted to the preservation service. An ingest report detailing the result will be generated and can be viewed later.

Dataset:
FDPAS test

Quota:
Total capacity: 977 MiB
Free space: 937 MiB
Size of dataset: 4 Bytes

39 MiB

937 MiB

After accepting this dataset 937.43 MiB of space will remain free.

Previous

Accept

Close

Reset

CONTRACT

PROPOSE

VALIDATE

CHECK

ACCEPT



Status ▲	Dataset	Owner	Contract
In packaging service	FDPAS test		CSC sop



FDPAS test

Status

Preservation state:	In digital preservation
Preservation state modified:	Oct 8, 2019 15:50:12
System description:	Accepted to preservation
User description:	Metadata confirmed
Size: 4 Bytes	

Hallintaliittymä

PAS-palvelun siirtoraportti

- Raportista selviää PAS-palvelun suorittamat vaiheet vastaanotossa sekä niiden lopputulos
- Kun aineisto läpäisee nämä vaiheet PAS-palvelussa, se on siirtynyt pitkäaikaissäilytykseen
 - Vastuu säilytyksestä siirtynyt PAS-palvelulle

FDPAS test

Status

Preservation state: **In digital preservation**
Preservation state modified: Oct 8, 2019 15:50:12
System description: Accepted to preservation
User description: Metadata confirmed

Size: 4 Bytes

✓ Dataset

📄 Contract

📄 Ingest Report

📄 METS Report

📄 Preservation Plan

Ingest Report

Digital preservation service – Ingest report

Ingest report – Submission accepted

2019-10-08T15:41:02.239231

Digital preservation service has received submission information package **doi:10.23729/9477a00e-9e9a-4349-b6bd-c9a4fe4a787a** from user **tpas** and performed following actions. All actions were successful and preservation responsibility has been changed to the digital preservation service.

Digital signature *signature.sig*

Event name	Outcome
Submission information package digital signature validation	success

Archival information package *aineisto_28718-b09c6e37-e1ae-42d9-80ec-dba09b660981.tar-5c5ef23f-95e0-47bf-9f71-c8217e85f61d.tar*

Event name	Outcome
Creation of archival information package	success
Preservation responsibility change to the digital preservation service	success

Submission information package */home/tpas/transfer/aineisto_28718-b09c6e37-e1ae-42d9-80ec-dba09b660981.tar*

Event name	Outcome
Virus check of transferred files	success
Fixity check of digital objects in submission information package	success
Validation of service contract properties	success
Decompression of submission information package	success

Metadata object *mets.xml*

Event name	Outcome
METS schema validation	success
Additional METS validation of required features	success
Additional METS validation of required features	success

Paketointikomponenttikoulutus 26.11.2019

Ilmoittautuminen 20.11. mennessä:

<http://digitalpreservation.fi/>



<https://fairdata.fi/>

<http://digitalpreservation.fi/>

<https://github.com/Digital-Preservation-Finland/>

Twitter: @dpres_fi

pas-support@csc.fi