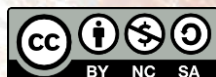


LOOGISEN TASON SÄILYTTÄMINEN – VAATIMUSMÄÄRITTELY

v1.0 (30.3.2021)

Tämä määrittely on osa opetus- ja kulttuuriministeriön
Avoimen tieteen ja digitaalisen kulttuuriperinnön kokonaisuutta



Lisenssi

Creative Commons Suomi CC-BY-NC-SA 4.0 (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.fi>)

Tämän dokumentin käyttäjälle annetaan oikeus jakaa eli kopioida, levittää, näyttää ja esittää teosta sekä valmistaa muutettuja teoksia seuraavilla ehdoilla:

- *Opetus- ja kulttuuriministeriö nimetään teoksen tekijäksi (ei kuitenkaan siten, että ilmoitus viittaisi lisenssinantajan tukevan lisenssinsaajaa tai teoksen käyttötapaa).*
- *Epäkaupallinen osapuoli eli käyttäjä ei saa käyttää teosta kaupallisesti.*
- *Mikäli käyttäjä tekee muutoksia tai käyttää teosta omien teostensa pohjana, tulee johdannaisteos jakaa samalla tai samankaltaisella lisenssillä.*

SISÄLTÖ

1	JOHDANTO	4
2	YLEISET VAATIMUKSET	7
3	PAS-PALVELUIDEN MÄÄRITTELYT	9
3.1	Tiedostomuotojen määrittelyjen ylläpitäminen.....	9
3.2	Teknisten metatietojen määrittelyn ylläpitäminen.....	10
3.3	Metatieto- ja paketoitumäärittelyn tuki loogiselle säilyttämiselle	10
3.4	Aineiston etsiminen, hallinnointi ja nouto	11
4	TIEDOSTOJEN KÄSITTELY	12
4.1	Muiden kuin säilytys- ja siirtokelpoisten tiedostomuotojen käsittely	12
4.2	Tiedostomuotojen validointi	12
5	SÄILYTYSSUUNNITELMAT	14
5.1	Tiedostomuotokohtaiset ominaisuudet	14
5.2	Aineiston piirteet.....	15
5.2.1	Piirteiden tunnistaminen	15
5.2.2	Piirteiden priorisointi säilytyksen suunnittelussa.....	15
5.3	Säilytyssuunnitelma	16
6	MIGRAATIOT	18
6.1	Migraatio tarpeen tunnistaminen ja ennakoiminen	18
6.2	Migraatiopolkujen määrittäminen tiedostomuodoille.....	19
6.3	Migraatioiden suunnittelu.....	20
6.4	Migraatioiden toteuttaminen ja laadunvarmistus	21
6.5	Migraatioiden päättäminen ja hyväksyntä.....	22
6.6	Vanhojen versioiden käsittely	22
6.7	Normalisointi	23
7	EMULOINTI	24
	LÄHTEET	25

1 JOHDANTO

Looginen säilyttäminen on prosessi, jossa varmistetaan että tiedostot, joista aineisto koostuu, säilyvät teknisesti ymmärrettävinä ja luettavina riippumatta teknologioissa ja kohdeyhteisössä tapahtuvista muutoksista. Prosessiin kuuluu muun muassa huolellinen säilytyksen suunnittelu, tiedostomuotomigraatiot ja riittävästä teknisistä metatiedoista huolehtiminen.

Tässä dokumentissa esitetään PAS-palveluita hyödyntävien organisaatioiden tämän hetken käsitys tarpeista loogisen säilyttämisen palveluille osana kansallisia PAS-palveluita. Dokumentti on tuotettu osana PAS-yhteistyöryhmän työskentelyä. Dokumentin tavoitteena on tunnistaa loogiseen säilyttämiseen liittyvät vaatimukset, kuvata niihin liittyvät tehtävät ja velvollisuudet niin PAS-palveluille kuin hyödyntäville organisaatiolle. Vaatimukset esitetään juoksevasti numeroituina taulukkomuodossa selittävän tekstin joukossa.

PAS-palveluilla tarkoitetaan kulttuuriperintöaineistojen ja tutkimusaineistojen pitkäaikaissäilyttämiseen tuotettuja palveluita yhdessä. Tässä dokumentissa hyödyntävällä organisaatiolla tarkoitetaan PAS-palvelua digitaalisten aineistojen pitkäaikaissäilyttämiseen käyttävää organisaatiota, laitosta tai muuta toimijaa.

Kulttuuriperintöaineistojen pitkäaikaissäilyttämiseen tuotettu PAS-palvelu (Kulttuuriperintö-PAS-palvelu) takaa kirjastojen, arkistojen ja museoiden keskeisten kansallisten digitalisten tietovarantojen pitkäaikaissäilyttämisen. Digitaalisilla kulttuuriperintöaineistoilla tarkoitetaan sekä digitoituja että digitaaliseen muotoon tuotettuja kulttuuriperintöaineistoja: lakisääteisen säilyttämisen piiriin kuuluvia digitaalisia kulttuuriaineistoja, kansalliseen kulttuuriperintöön kuuluvaa digitaalista asiakirjallista aineistoa sekä aineellisen ja henkisen kulttuuriperinnön säilyttämisestä vastaavien, opetus- ja kulttuuriministeriön hallinnonalalla toimivien organisaatioiden muita pitkäaikaissäilytyksen piiriin kuuluvia digitaalisia tietovarantoja.

Tutkimusaineistojen pitkäaikaissäilyttämiseen tuotettu PAS-palvelu (Fairdata PAS-palvelu) varmistaa tutkimuksen digitaalisten aineistojen saatavuuden ja pitkäaikaisen säilyvyyden. Tämä PAS-palvelu tukee osaltaan pysyvää ja koordinoitua toimintamallia tutkimusaineistojen hallinnan tueksi. Pyrkimyksenä on, että tutkimuksen todennettavuus ja toistettavuus elinkaaren eri vaiheissa onnistuu ja tulosten hyödyntäminen on helppoa. Tällöin tutkimustuloksia voidaan käyttää yhä uudelleen, arvioida, hyödyntää päätöksenteossa ja turvata digitalisoitumisen myötä yhä nopeammin kasvavat tietomäärät tulevien tutkijasukupolvien käyttöön.

Pitkäaikaissäilyttäminen voidaan jakaa karkeasti kolmeen tasoon (ks. kuva 1), joiden rajat ovat usein häilyviä. **Bittitason säilyttämisessä** huolehditaan siitä, että tiedostot pysyvät täysin muuttumattomina ajasta ja paikasta riippumatta. Tämän toiminnallisuuden toteuttamiseen vaaditaan aktiivista tiedostojen eheyden seurantaa, kopioiden hallintaa, teknologioiden virkistämistä hallitusti sekä aineistojen maantieteellistä hajauttamista. Kansallisten PAS-palveluiden tarkoituksena on, että hyödyntävien organisaatioiden ei tarvitse huolehtia, vaan PAS-palvelut huolehtia tämän tason säilyttämisestä, että tiedostot pysyvät täysin muuttumattomina. Bittitason säilyttäminen on PAS-palveluiden normaalia aktiivista päivittäistä toimintaa.

Loogisen tason säilyttämisessä, johon tämä dokumentti keskittyy, tarkoituksena on mahdollistaa aineistojen hyödyntäminen myös tulevaisuudessa. Se että tiedosto on eheä ja bitit ovat säilyneet muuttumattomina ei takaa sitä, että tiedosto voitaisiin myöhemmin avata ja tulkita millään sen hetkellä ohjelmistolla. Loogisen tason säilyttämisessä pyritään pitämään huolta siitä, että tiedosto voidaan avata ja tulkita myös tulevaisuudessa. Tämä edellyttää mahdollisiin teknisiin ongelmiin varautumista ennakkoon. Ensisijaisesti se tarkoittaa säilytykseen otettavien tiedostomuotojen rajaamista sellaisiin, joita palvelun avulla voidaan myöhemmin loogisen säilyttämisen toimin migroida. Käytännön tasolla tämä näkyy hyväksyttävien tiedostomuotojen rajauksena. Keskeinen osa loogiseen säilyttämiseen valmistautumista on osana säilytyksen

suunnittelua määrittää aineistolajeittain sekä tiedostomuodoittain yhteistyössä palvelua hyödyntävien organisaatioiden kanssa ne keskeiset piirteet, joiden tulee säilyttämistoimien yhteydessä säilyä. Kansallisissa PAS-palveluissa tiedostomuotojen hallinta on huomioitu jo toiminnan alkuvaiheissa, mukaan lukien erityiset tekniset vaatimukset tietyille tiedostomuodoille ja niiden teknisille metatiedoille [TDSTO]. Näillä vaatimuksilla pyritään varmistamaan se, että loogisen tason säilyttäminen onnistuu pitkällä tähtäimellä.

Loogisessa säilyttämisessä seurataan teknologian kehitystä ja sen vaikutuksia aineiston käyttöön, säilytysmenetelmien ja mallien muuttumiseen sekä tiedostomuotojen ja työkalujen kehitykseen. Tätä seurantaa tekevät sekä PAS-palvelu että hyödyntävät organisaatiot. Jos säilytyksessä oleva aineisto on vaarassa vanhentua teknisesti, tai aineiston käyttö on muuttunut ratkaisevasti, pyritään aineisto ensisijaisesti migroimaan uuteen tiedostomuotoon. Säilytys suunnitelmien avulla suunnitellaan migraatiot niin, että aineiston keskeiset ja tärkeät piirteet eivät muutu ratkaisevasti. Riskienhallinta on oleellinen osa pitkäaikaissäilyttämisen kokonaisuutta ja siten myös loogisen tason säilyttämistä.

Toisin kuin bittitason säilyttämisessä, loogisen tason säilyttämisessä tarvitaan PAS-palveluiden ja hyödyntävän organisaation tiivistä yhteistyötä.



Kuva 1: Säilyttämisen tasot

Semanttisen tason säilyttäminen keskittyy erityisesti aineistojen ymmärrettävyyden säilyttämiseen. Se vaatii kohdeyhteisön ja aineistojen syvällistä tuntemusta ja ymmärtämistä. Vaikka aineistot voidaan säilyttää eheinä ja muutoin käyttökelpoisina, niiden hyödyntäminen tulevaisuudessa voi olla mahdotonta, jos aineistojen sisältöä ei tulevaisuudessa ymmärretä. Aineistoa kuvaileva metatieto, rakenteelliset metatiedot ja aineiston historiatiedot ovat tässä oleellisessa roolissa. Vastuu semanttisen tason säilyttämisestä kuuluu lähinnä hyödyntävälle organisaatiolle, koska ei voida olettaa, että PAS-palveluilla on riittävää ymmärrystä tulkita kaikkia sen säilytyksessä olevia, keskenään hyvinkin heterogeenisiä aineistoja tai tuntea niiden kohdeyhteisöjä. Semanttisen tason säilyttämistä ei käsitellä tässä dokumentissa.

PAS-palvelu tarjoaa neuvonta- ja tukipalveluita hyödyntäville organisaatiolle digitaaliseen pitkäaikaissäilytykseen liittyvissä asioissa. Tämä on PAS-palvelulle sopimusvelvoite CSC:n ja hyödyntävän organisaation välisessä PAS-palvelusopimuksessa. Neuvonta- ja tukipalvelut koskevat luonnollisesti myös kaikkia loogiseen säilyttämiseen liittyviä asioita. Tästä syystä neuvonta- ja tukipalveluita ei tässä dokumentissa erikseen käsitellä tiettyyn toimintoon liittyvinä vaatimuksina.

Pitkäaikaissäilytykseen liittyvät riskit huomioidaan PAS-palveluiden suunnittelussa ja päivittäisessä toiminnassa. Riskienhallinnan avulla voidaan tunnistaa toimintaan liittyviä haasteita, analysoida

pitkäaikaissäilytyksen tavoitteen toteutumisen riskejä ja varautua niihin ennakolta. Koska riskienhallinta koskee PAS-palveluita kokonaisuudessaan, riskienhallintaa ei tässä dokumentissa erikseen käsitellä, ellei tiettyyn toimintoon liity erityisiä riskejä.

Tämä dokumentti lähestyy loogista säilyttämistä ja säilytyksen suunnittelua seuraavalla rakenteella:

- Luvussa 2 kuvataan yleiset vaatimukset loogiselle säilyttämiselle.
- Luvussa 3 esitetään vaatimukset, jotka liittyvät PAS-palveluita hyödyntäviä organisaatioita ohjaaviin PAS-palveluiden määrittelyihin. Ne luovat pohjan loogiselle säilyttämiselle.
- Luvussa 4 kuvataan tiedostojen käsittely PAS-palveluissa. Huolellinen tiedostojen käsittely (esim. validointi) mahdollistaa laadukkaiden migraatioiden toteuttamisen tulevaisuudessa.
- Luvussa 5 esitetään säilytysuunnitelmiin liittyvät vaatimukset. Säilytysuunnitelmat luovat pohjan migraatioille.
- Luvussa 6 kuvataan migraatioihin liittyvät vaatimukset.
- Luvussa 7 esitetään vaatimukset, jotka liittyvät emulaatioon, jota voidaan käyttää, jos migraatio ei ole mahdollinen säilytystoimenpide.

Tässä dokumentissa käytetty sanasto löytyy osoitteesta <https://digitalpreservation.fi/specifications/sanasto>

2 YLEISET VAATIMUKSET

Loogiseen säilyttämiseen kohdistuu yleisen tason vaatimuksia, jotka määrittelevät toiminnan kehukset ja asettavat sille reunaehdoja. Loogisesta säilyttämisestä vastaavat PAS-palvelu ja hyödyntävät organisaatiot yhdessä.

Koska looginen säilyttäminen edellyttää teknistä tuntemusta, on ensisijaisesti PAS-palvelun vastuulla analysoida tiedostomuotoja ja niiden kestävyyttä, tunnistaa loogisen säilyttämisen kannalta keskeisiä teknisiä piirteitä ja tukea hyödyntäviä organisaatioita näiden piirteiden säilyvyyden ja merkityksen arvioinnissa, ja seurata työkalujen kehitystä. Lisäksi PAS-palvelu tarjoaa digitaalisten objektien validointia, migraatioiden toteuttamista ja mallisuunnitelmien tekoa osana palvelua hyödyntäville organisaatioille. Näillä tavoin PAS-palvelu tukee organisaatioita, jotka voivat aineiston loogisessa säilyttämisessä keskittyä itse informaatioisällön tuntemukseen ja sen käytettävyyden varmistamiseen omassa toimintaympäristössään.

Hyödyntävät organisaatiot vastaavat kohdeyhteisön seuraamisesta, PAS-palveluun tallennettujen kuvailevien metatietojen ajantasaisuudesta ja kattavuudesta sekä tallennettujen aineistojen säilytettävien piirteiden tunnistamisesta ja niiden dokumentoimisesta. Organisaatiot osallistuvat säilytyksen suunnitteluun, jotta aineisto ja kohdeyhteisöjen tarpeet huomioidaan tarkoituksenmukaisella tavalla.

Vaatus	Kuvaus
V1	Loogisessa säilyttämisessä varmistetaan digitaalisten aineistojen käytettävyys tuleville sukupolville.
V2	PAS-palvelun ja hyödyntävien organisaatioiden yhteistyöllä varmistetaan loogisen säilyttämisen suunnittelu ja toteutus käyttäjä- ja tarvelähtöisesti.
V3	PAS-palvelu antaa asiantuntija-apua hyödyntäville organisaatioille aineiston säilytyskysymyksissä laajasti.
V4	Loogisen säilyttämisen vaatimukset kohdistuvat sekä PAS-palveluun hyödyntäviin organisaatioihin ja kohdeyhteisöihin.
V5	Loogisen säilyttämisen riskit on huomioitu PAS-palvelun kokonaisuuden riskienhallinnassa.
V6	PAS-palvelun määritykset mahdollistavat loogisen säilyttämisen.
V7	PAS-palvelun rajapinnat tukevat loogista säilyttämistä.
V8	Hyödyntävän organisaation ja PAS-palvelun tuottamat metatiedot mahdollistavat tarkoituksenmukaisten säilytystoimenpiteiden kohdistaminen aineistoon.
V9	PAS-palvelu tarjoaa loogisen tason säilyttämistä yhdessä rajatuille, hyödyntävien organisaatioiden käyttötärpeisiin soveltuville tiedostomuodoille, ns. säilytys- ja siirtokelpoisille tiedostomuodoille.
V10	Looginen säilyttäminen on aktiivista toimintaa, joka vaatii kohdeyhteisön ja teknologian seuranta sekä aineistojen käytön toimintaympäristön muuttumisen tunnistamista.
V11	PAS-palvelu seuraa tiedostomuotojen ja muiden toimialaansa ja asiakkaisiinsa vaikuttavien teknisten määritelmien päivitystarvetta.
V12	PAS-palvelu kehittää palveluitaan teknisen ja kansainvälisen ympäristön kehitystä seuraten ja siihen osallistuen.
V13	Hyödyntävä organisaatio seuraa käyttämiensä metatietoformaattien ja muiden omalle vastuulleen kuuluvien teknisten määritelmien kuten tunnisteiden päivitystarvetta.
V14	Hyödyntävän organisaation vastuulla on kohdeyhteisöjen tarpeiden seuranta ja niiden huomioon ottaminen aineistojen loogisessa säilyttämisessä.

Vaatus	Kuvau
V15	Hyödyntävä organisaatio voi tarvittaessa hyödyntää PAS-palvelussa olevaa aineistoa ja/tai sen metatietoja tuotantojärjestelmiensä avulla.

3 PAS-PALVELUIDEN MÄÄRITTELYT

PAS-palveluita hyödyntäviä organisaatioita ohjaa kolme määrittelyä PAS-palveluiden käyttämiseksi:

- *Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen*, joka määrittelee PAS-palveluissa säilytettävien aineistojen metatietoformaattit sekä siirto- ja jakelupaketin rakenteen erityisesti niissä käytettävien METS-profiilien osalta [METADATA].
- *Säilytys- ja siirtokelpoiset tiedostomuodot*, joka määrittelee tiedostomuodot, ja niiltä vaadittavat tekniset metatiedot, joissa PAS-palvelut säilyttävät ja vastaanottavat aineistoja [TDSTO].
- *PAS-palveluiden rajapinnat*, joka määrittelee PAS-palveluiden rajapinnat, joiden avulla kansallisia PAS-palveluja hyödyntävät organisaatiot voivat muun muassa siirtää aineistoa säilytykseen PAS-palveluihin tai noutaa niistä aineistoa [API].

3.1 Tiedostomuotojen määrittelyjen ylläpitäminen

PAS-palveluissa on määritelty tiedostomuodot, ja niiltä vaadittavat tekniset metatiedot, joissa PAS-palvelut säilyttävät ja vastaanottavat aineistoja. Tällä määrittelyllä on kaksi erityistä tarkoitusta. Ensinnäkin, määrittelyyn on hyväksytty vain tiedostomuotoja, joiden pitkäaikaissäilyttäminen voidaan varmistaa kustannustehokkaasti ja joista PAS-palvelu voi siten ottaa säilytysvastuun myös loogisella tasolla. Tiedostomuotojen säilyvyyden arvioimiseksi on olemassa prosessi, jossa arvioidaan säännöllisin väliajoin kunkin tiedostomuodon tekninen vakaus, avoimuus ja käytön laajuus. Määrittely ohjaa hyödyntäviä organisaatioita käyttämään tiedostomuotoja, jotka eivät vain ole PAS-kelpoisia, mutta muutenkin yleisesti käytettyjä ja hyväksi arvioituja tiedostomuotoja. Lisäksi tiedostojen teknisten metatietojen avulla kuvataan tiedostomuotokohtaisia teknisiä ominaisuuksia, joita hyödynnetään säilytyksen suunnittelussa, kun arvioidaan aineiston piirteiden säilymistä.

Tiedostomuotojen ja teknisten metatietojen määrittelyjen ylläpitoon on olemassa vakiintunut prosessi PAS-toiminnan vuosikellon mukaisesti. Lisäksi tiedostomuotojen arvioinnille on olemassa arviointikriteeristö, jota on jo usean vuoden ajan hyödynnetty. Tämä kattaa perustoiminnallisuuden tiedostomuotojen määrittelylle, ja se on todettu toimivaksi malliksi.

Vaatus	Kuvaus
V16	PAS-palvelu ylläpitää säilytuskelpoisten tiedostomuotojen määrittelyä yhteistyössä hyödyntävien organisaatioiden kanssa.
V17	PAS-palvelu ylläpitää siirtokelpoisten tiedostomuotojen määrittelyä yhteistyössä hyödyntävien organisaatioiden kanssa.
V18	PAS-palvelu ylläpitää tiedostomuotojen soveltamisohjeita hyödyntäviltä organisaatioilta saadun palautteen perusteella.
V19	Tiedostomuotojen määrittelyjen päivittämiseen on olemassa yhteisesti sovittu prosessi.
V20	PAS-palvelu tuottaa määrittelyjen sisältöä hyödyntäviltä organisaatioilta saadun palautteen perusteella.
V21	Uusia tiedostomuotoja varten on olemassa määritelty arviointiprosessi (ks. [TDSTO]), jonka tiedostomuodon on läpäistävä riittävällä tasolla ennen kuin sen voi lisätä määrittelyyn.
V22	PAS-palvelu kykenee migroimaan säilytys- tai siirtokelpoisessa tiedostomuodossa olevia aineistoja ja niiden metatietoja.
V23	PAS-palvelu on määrittellyt tavan ja prosessin, jolla muita kuin säilytys- ja siirtokelpoisia tiedostomuotoja voidaan ottaa vastaan bittitasoisen säilyttämisen palvelun piiriin.

3.2 Teknisten metatietojen määrittelyn ylläpitäminen

Vuosikellon mukaisesti ylläpidettävät määrittelydokumentit määrittelevät pakollisia teknisiä metatietoja muun muassa elävän kuvan sekä äänen- ja kuvien tiedostomuodoille. Loogisen tason säilyttämisessä oleellisessa osassa on metatietojen kattavuus sekä valittujen metatietoformaattien ylläpito ja yleisyys. Pakollisten metatietojen päivitys tai metatietoformaatin vaihtaminen arvioidaan tarvelähtöisesti PAS-palvelun ja hyödyntävien organisaatioiden yhteistyönä. Testimigraatioiden avulla kehittää menetelmiä metatietojen rikastamiseen, mikä luo edellytyksiä metatietovaatimusten parantamiseksi.

PAS-palvelu voi taata aineiston ymmärrettävyyden säilymisen vain siinä laajuudessa, joka ilmenee aineiston metatiedoista. On mahdollista, että aineistot vaarantuvat, jos niistä ei luoda riittäviä metatietoja riittävän aikaisessa vaiheessa. Esimerkiksi CSV-tiedoston vieminen PAS-palveluun pelkkänä tekstinä ilman teknisen rakenteen kuvailua tai XML-tiedoston vieminen PAS-palveluun ilman skeemaa voi aiheuttaa sen, että näiden aineistojen rakenteen säilyttäminen osoittautuu mahdottomaksi. Kuitenkaan kaikille tiedostomuodoille ei tällä hetkellä ole vastaavaa teknistä metatietomuotoa. Tällaisia ovat esimerkiksi PDF- tai ODF-muodot. Loogisessa säilyttämisessä tarvittavien teknisten metatietojen tuottaminen takautuvasti olla vaikeaa.

Vaatus	Kuvaus
V24	PAS-palvelu ylläpitää teknisten metatietojen määrittelyä yhteistyössä hyödyntävien organisaatioiden kanssa.
V25	PAS-palvelu tuottaa määrittelyt hyödyntäviltä organisaatioilta saadun palautteen perusteella.
V26	Tekniset metatiedot ovat kansainvälisten standardien mukaisia.
V27	Tekniset metatiedot on liitettävä osaksi siirtopaketin metatietoja.
V28	Teknisiä metatietoja hyödynnetään arvioitaessa aineistojen piirteitä ja migraatioiden vaikutuksia niihin sekä migraatioihin varautumisessa.

3.3 Metatieto- ja paketointimäärittelyn tuki loogiselle säilyttämiselle

Loogisessa säilyttämisessä itse aineisto ja sen hyödyntämiseen liittyvät ominaisuudet ovat keskiössä. Aineiston siirtoon, säilyttämiseen ja jakeluun tarkoitetut tietopaketit ja niiden rakenne eivät suoraan liity loogiseen säilyttämiseen, mutta tukevat sitä tarjoamalla yhtenäisen rakenteen, jossa sekä tiedostot että muut hyödyntämiseen liittyvät metatiedot ovat. Aineiston tietopaketteja on kuitenkin pystyttävä hallitsemaan, jotta aineisto on löydettävissä ja siihen voi kohdistaa tarkoituksenmukaisia säilytystoimenpiteitä. Lisäksi metatietojen on oltava määrityksien mukaisessa muodossa, jotta niiden käsittely ja hyödyntäminen loogisessa säilyttämisessä on mahdollista.

PAS-palvelussa olevasta aineistosta ja/tai sen metatiedoista syntyy uusi versio päivityksen tai migraation yhteydessä. Hyödyntävä organisaatio voi päivittää aineistoa eri syistä, esimerkiksi kun metatietoja on rikastettu, aineistoa on karttunut lisää, tai hyödyntävä organisaatio on itse suorittanut migraation aineistolle tai pelkille aineistometatiedoille. Hyödyntävän organisaation pitää kyetä etsimään ja hallinnoimaan jokaista yksittäistä säilytyksessä olevaa aineistoversiota.

PAS-palvelu luo aineistoista uusia versioita vain migraatioiden yhteydessä, yhteistyössä hyödyntävän organisaation kanssa. PAS-palvelun toteuttama migraatio voi kohdistua metatietoihin, tiedostomuotoon tai johonkin säilytyspaketin tekniseen ominaisuuteen.

Aineistojen ja niiden metatietojen paketointi pitkäaikais säilytykseen-määrittelyssä on määritelty aineiston tietopakettien tunnistet, versiointi ja sallitut metatietoformaattit.

Vaatus	Kuvaus
V29	PAS-palvelu ylläpitää metatietojen ja paketoinnin määrittämistä yhteistyössä hyödyntävän organisaation kanssa.
V30	Aineistot ja aineistokokonaisuudet yksilöidään tunnisteilla paketoitumäärityksessä kuvatulla tavalla.
V31	Tunnisteet ja muut metatiedot mahdollistavat aineiston löydettävyyden PAS-palvelusta.
V32	Säilytyksessä olevat eri aineistoversiot, jotka syntyvät esimerkiksi aineiston päivittämisen tai migraation myötä, pitää olla identifioitavissa PAS-palvelussa.

3.4 Aineiston etsiminen, hallinnointi ja nouto

PAS-palvelun rajapintojen kautta hyödyntävät organisaatiot voivat siirtää palveluun aineistoja metatietoineen, etsiä metatietoja ja aineistoa palvelusta sekä palauttaa niitä itselleen. Loogisen säilyttämisen näkökulmasta hyödyntävän organisaation pitää kyetä hallinnoimaan säilytyksessä olevaa aineistoaan ja sen metatietoja.

PAS-palvelu tarjoaa hyödyntävälle organisaatiolle mahdollisuuden saada tietoa siitä, mitä aineistoja ja metatietoja se säilyttää PAS-palvelussa, missä muodossa ne ovat, sekä säilytettävien aineistojen määrän ja laajuuden. Hyödyntävän organisaation pitää voida etsiä yksittäisiä metatietoja, aineistoja tai aineistokokonaisuuksia niiden metatietojen perusteella.

Looginen säilyttäminen vaatii aktiivista seuranta ja tietoa siitä, mitä toimenpiteitä säilytettävään aineistoon voi ja kannattaa kohdistaa, mitä varten hyödyntävä organisaatio saa PAS-palvelusta raportteja säilytyksessä olevasta aineistosta.

Hyödyntävä organisaatio siirtää PAS-palvelussa luotuja migroituja aineistoja metatietoineen omiin tuotantjärjestelmiinsä. PAS-toiminnan tavoitteena on varmistaa säilytettävien aineistojen käytettävyys, ja migraatio tehdään yleensä vain silloin, kun alkuperäisen aineiston käytettävyys on vaarassa ja se pitää korvata ajantasaisemmalla versiolla.

Migraatioiden yhteydessä tapahtuvasta raportoinnista ja kommunikaatiosta kerrotaan tarkemmin luvussa 5.

Vaatus	Kuvaus
V33	Hyödyntävä organisaatio voi päivittää tai migroida säilytyksessä olevaa aineistoa ja metatietoja.
V34	Hyödyntävä organisaatio voi etsiä PAS-palvelussa säilytettävää aineistoa metatietojen perusteella.
V35	PAS-palvelu tarjoaa raportointia säilytyksessä olevasta aineistosta eri hakukriteereillä.
V36	Hyödyntävä organisaatio voi noutaa siirtämänsä aineiston ja/tai metatiedot takaisin PAS-palvelusta rajapinnat määrittelyssä kuvatulla tavalla.
V37	Hyödyntävä organisaatio voi saada päivitetty metatiedot takaisin PAS-palvelusta.
V38	Hyödyntävä organisaatio voi saada päivitetyn, esimerkiksi migroidun, aineiston takaisin PAS-palvelusta.
V39	PAS-palvelun rajapinnat mahdollistavat aineiston ja metatietojen etsimisen ja noudon hyödyntäville organisaatioille.

4 TIEDOSTOJEN KÄSITTELY

4.1 Muiden kuin säilytys- ja siirtokelpoisten tiedostomuotojen käsittely

Tiedostomuotomäärittelyn [TDSTO] liitteessä C on kuvattu kolme tapausta vaatimuksineen ja käsittelyohjeineen, joiden mukaisesti PAS-palveluun voidaan siirtää säilytettäväksi sellaisia tiedostomuotoja, jotka ovat muussa kuin säilytys- tai siirtokelpoisessa muodossa. Tällainen edellyttää kuitenkin hyödyntävän organisaation ja PAS-palvelun yhteistä arviointia ja hyväksyntää. Hyväksyntä on sopimuskohtaista, eli tietylle aineistolle hyväksytty tapa ei ole automaattisesti hyväksyttyä toisen sopimuksen piiriin kuuluvalla aineistolle, vaikka tiedostomuodot olisivatkin samoja. PAS-palvelu tarjoaa tällaiselle aineistolle vain bittitason säilytystä, ja aineiston ymmärrettävyyden säilyttäminen jää hyödyntävän organisaation vastuulle. Käyttöönotoissa on varmistuttava, että poikkeuskäsittelylle on aito tarve, ja että asiaa ei voida muulla tavoin ratkaista.

Vaatus	Kuvaus
V40	Muiden kuin säilytys- ja siirtokelpoisten tiedostomuotojen käsittelyyn ja vastaanottoon on olemassa toimintatapa.
V41	Hyödyntävän organisaation on aina sovittava PAS-palvelun kanssa muiden kuin säilytys- ja siirtokelpoisten tiedostomuotojen siirtämisestä PAS-palveluun.
V42	Muut kuin säilytys- ja siirtokelpoiset tiedostomuodot on normalisoitava säilytyskelpoiseen muotoon, joka toimitetaan epäkelvon tiedostomuodon mukana samassa siirtopaketissa. Jos normalisointi ei ole mahdollista, tiedoston on oltava eheä ja tiedostomuodon dokumentoitu.
V43	PAS-palvelu takaa vain bittitason säilytyksen muille kuin säilytys- ja siirtokelpoisille tiedostomuodoille.
V44	PAS-palvelu ei tarjoa valmista migraatiopolkua tai säilytysuunnitelmaa aineistolle, joka on muussa kuin säilytys- ja siirtokelpoisessa tiedostomuodossa, mutta voi tukea hyödyntävää organisaatiota ymmärrettävyyden säilyttämisessä.
V45	Hyödyntävä organisaatio vastaa muiden kuin säilytys- ja siirtokelpoisten tiedostomuotojen käytettävyyden/hyödynnettävyyden varmistamisesta.

4.2 Tiedostomuotojen validointi

PAS-palvelu validoi kaikki säilytys- ja siirtokelpoiset tiedostot vastaanotossa teknisesti. Validoinnissa varmistetaan se, että tiedosto on virheetön ja vastaa ilmoitettua. Teknisiä virheitä sisältävien tiedostojen hyväksyminen säilytykseen voi vaarantaa niiden pitkäaikaissäilytyksen.

Aineistot voidaan kuitenkin tapauskohtaisesti tallentaa PAS-palveluun niiden virheellisyydestä huolimatta. PAS-palvelu hylkää ja palauttaa hyödyntävälle organisaatiolle käsiteltäväksi kaikki siirtopaketit jotka sisältävät tiedostomuotoja, jotka eivät ole säilytys- tai siirtokelpoisia ja joita ei ole erikseen sovittu otettavaksi säilytykseen PAS-palveluun.

Vaatus	Kuvaus
V46	PAS-palvelu validoi säilytys- ja siirtokelpoisessa muodossa olevan aineiston osana aineiston vastaanottoa.
V47	Vastaanotossa tapahtuva validoinnin tapa ja laajuus on määritelty ja dokumentoitu ("miksi se tehdään ja mitä siitä seuraa").
V48	Jokaiselle määrittelyissä olevalle tiedostomuodolle on oltava käytössä vähintään yksi validaattori.

Vaatus	Kuvaus
V49	Ainoastaan eheä validoitu aineisto voidaan ottaa säilytykseen PAS-palveluissa. Poikkeustapauksia varten on olemassa toimintamalli.
V50	PAS-palvelu kirjaa validoinnin tulokset vastaanottoraporttiin.
V51	PAS-palvelu tukee hyödyntäviä organisaatioita hylättyjen tiedostojen tai pakettien korjaamisessa erikseen sovittavalla tavalla.
V52	PAS-palvelu on määritellyt prosessin, jolla tietyt validoinnissa hylätyt aineistot voidaan ottaa säilytykseen poikkeuskäsittelynä, jos se on perusteltua.
V53	PAS-palvelu julkaisee ja ylläpitää listaa validoinnissa käytetyistä työkaluista ja niiden käyttötavasta.

5 SÄILYTYSSUUNNITELMAT

Säilytyssuunnitelma on menetelmä, jolla hyödyntävä organisaatio kertoo PAS-palvelulle mitä piirteitä aineisto sisältää, ja miten nämä piirteet tulisi säilyttää. Suunnitelmien avulla PAS-palvelu voi kohdentaa oikeanlaisia säilytystoimenpiteitä säilytettävälle aineistolle.

Säilytyssuunnitelma kuvaa tavoitteet aineiston pitkäaikaissäilyttämiseksi ja menetelmät tavoitteiden saavuttamiseksi. Säilytyssuunnitelma voi kohdistua aineistokokonaisuuteen, yksittäisiin tiedostomuotoihin, metatietoihin tai jopa yksittäiseen digitaaliseen objektiin.

Säilytyssuunnitelman avulla tulee kyetä määrittämään aineistotyypeittäin niitä keskeisiä piirteitä, joiden halutaan säilyvän migraatioissa ja jotka ovat koneellisesti migraation yhteydessä tarkistettavissa. Säilytyssuunnitelmaa määriteltäessä tulee huomioida aineiston käyttötarkoitus: esimerkiksi digitoitu taideteos voi vaatia erilaisia säilytystoimenpiteitä kuin kuviksi digitoitu tekstidokumentti, ja väitöskirjalla voi olla tiukemmat vaatimukset sisällön säilyttämisen suhteen kuin muilla PDF/A-aineistoilla. Suunnitelmassa määritellyjä ominaisuuksia käytetään migraatiotyökalun valinnan arvioinnissa ja testaamisessa, sillä yhtä tiedostomuotoa voidaan käyttää useisiin tarkoituksiin. Säilytyssuunnitelman pitää mahdollistaa tällaisen eron perustelu, ja sen perusteella hyödyntävät organisaatiot voivat päättää soveltaa eri migraatiotyökaluja samalle tiedostomuodolle, ja sama hyödyntävä organisaatio voi migroida eri aineistokokonaisuuksia eri välineillä.

Tässä luvussa käsitellään säilytyssuunnitelmia PAS-palvelussa säilytyksessä olevien tai sinne tulevaisuudessa siirrettävien aineistojen kannalta keskittyen erityisesti näiden aineistojen teknisiin ominaisuuksiin. Hyödyntävällä organisaatioilla voi olla, ja usein on, omia säilytyssuunnitelmia (aineistohallintasuunnitelmat, kokoelmapolitiikka, tms.) aineistoilleen. Niissä voidaan esimerkiksi arvioida, mitä aineistokokonaisuuksia pidetään erityisen tärkeinä ja onko tietty aineistokokonaisuus ylipäättään säilyttämisen arvoista. Organisaatioiden omat säilytyssuunnitelmat voivat myös ottaa kantaa myös digitaalisten aineistojen säilytettäviin ominaisuuksiin. Hyödyntävien organisaatioiden omien ja PAS-toimintaa varten laadittujen säilytyssuunnitelmien on oltava yhteismitallisia, niissä ei saa olla ristiriitaisia linjauksia koskien esimerkiksi jonkin aineistokokonaisuuden säilyttämistä.

5.1 Tiedostomuotokohtaiset ominaisuudet

Tiedostomuotokohtaisten ominaisuuksien kattava kartoittaminen on edellytys sille, jotta realistinen säilytyssuunnitelma on mahdollista tehdä. Tätä kartoittamista on PAS-palveluissa jo tehty. Joistakin tiedostomuodoista on luotu kansainvälisiä arkistokelpoisia profiilistandardeja, jotka rajoittavat vaikeasti säilytettävien ominaisuuksien käyttöä. Esimerkiksi PDF-tiedostomuodolle on olemassa erillinen arkistokelpoinen määrittäminen (PDF/A), ja EPUB3 -tiedostomuodolle on laadittu vastaava tekninen spesifikaatio (ISO/IEC TS 22424-1 ja 22424-2). Nämä standardit kieltävät pitkäaikaissäilytyksen kannalta haitallisten ominaisuuksien kuten salauksen käytön, ja edellyttävät sitä tukevien ominaisuuksien soveltamista.

Hyödyntävän organisaation on määriteltävä ne aineistoissaan tai aineistokokonaisuuksissaan olevat merkittävät ominaisuudet, jotka eivät saa muuttua migraatioissa. Nämä ominaisuudet voivat olla hyvinkin erilaisia riippuen aineiston luonteesta ja käyttötarkoituksesta. Esimerkiksi jos äänitiedosto sisältää musiikkia, äänenlaatu ei saisi huonontua migraatioissa, mutta jos äänitiedosto sisältää haastattelun, niin migraation vaikutus äänenlaatuun ei välttämättä ole yhtä oleellista.

Vaatus	Kuvaus
V54	PAS-palvelu kartoittaa säilytys- ja siirtokelpoisten tiedostomuotojen tekniset ominaisuudet, jotka on mahdollista/tarpeellista huomioida säilytyssuunnitelmien laadinnassa.

Vaatus	Kuvaus
V55	Tunnistettuja teknisiä ominaisuuksia käytetään koneluettavien säilytys suunnitelmien pohjana.

5.2 Aineiston piirteet

5.2.1 Piirteiden tunnistaminen

Aineistojen piirteillä tarkoitetaan aineistokohtaista kuvaavaa tietoa. Tiedostomuotokohtaiset ominaisuudet kertovat kyseisen tiedostomuodon teknisistä ominaisuuksista sisällöstä riippumatta, mutta aineiston piirteet yksilöivät ja kuvaavat aineistokokonaisuuksia sisällön näkökulmasta. Esimerkiksi karttakuvilla keskeiset piirteet ovat erilaisia kuin taidevalokuvilla, vaikka molempien tiedostomuoto olisi sama.

PAS-palvelu pystyy tunnistamaan tiedostomuotojen teknisiä ominaisuuksia, mutta aineistokohtaisten piirteiden määrittelyyn tarvitaan kokoelmien sisällön tuntemusta. Tähän tarvittava tietotaito on vain hyödyntävällä organisaatiolla. Jos esimerkiksi karttakuvissa viivojen paksuudella on semanttista merkitystä, hyödyntävän organisaation täytyy dokumentoida tämä piirre ja huolehtia siitä, että se dokumentoidaan karttakuvien säilytys suunnitelmaan.

Säilytys- ja siirtokelpoiset tiedostomuodot voidaan jakaa neljään pääryhmään: kuvamuodot, äänimuodot, elävän kuvan muodot, rakenteiset muodot ja tekstidokumenttimuodot. Osa säilytys- ja siirtokelpoisista tiedostomuodoista ei tosin sovi näihin ryhmiin. Kunkin pääryhmän tiedostumuodoilla on yhteisiä piirteitä, jotka kukin tiedostomuoto toteuttaa omalla tavallaan. Yhteisien piirteiden lisäksi voi olla myös tiedostomuotokohtaisia erityispiirteitä, jotka eivät ole migroitavissa. Aineistosta riippuen näitä tiedostumuotojen yhteisiä ja yksilöllisiä ominaisuuksia on hyödynnetty eri tavoin.

Vaatus	Kuvaus
V56	Hyödyntävät organisaatiot tunnistavat ja dokumentoivat aineistokohtaisesti säilytys suunnitelmiin vaikuttavat aineistojen piirteet.
V57	PAS-palvelu tukee hyödyntäviä organisaatioita tarjoamalla työkaluja ja tukea tiedostomuoto- ja aineistokohtaisten piirteiden määrittelyyn, tunnistamiseen ja dokumentointiin.
V58	PAS-palvelu tukee hyödyntäviä organisaatioita tarjoamalla tietoa tiedostomuotojen teknisistä ominaisuuksista ja niiden sovittamisesta aineistokohtaisten piirteiden tunnistamiseen.
V59	Hyödyntävät organisaatiot tukevat toisiaan ja PAS-palvelua tarjoamalla tietoa aineistojensa teknisistä ominaisuuksista ja säilytettävistä piirteistä.
V60	PAS-palvelu tukee hyödyntäviä organisaatioita laatimalla ohjeita ja malleja aineiston piirteiden viemisestä säilytys suunnitelmiin.

5.2.2 Piirteiden priorisointi säilytyksen suunnittelussa

Säilytettävien piirteiden priorisoinnissa on oleellista, millä perusteella ja ketä varten aineistoa säilytetään (kohdeyhteisö). Kohdeyhteisön muuttuminen voi vaikuttaa oleellisesti säilytyksen prioriteetteihin. Piirteiden priorisoinnissa on otettava huomioon piirteiden merkitys kohdeyhteisölle ja piirteiden säilyttämisen kustannukset. Piirteiden merkitys kuvaa sen tärkeyttä: jos oleellinen piirre häviää tai muuttuu, se voi haitata ratkaisevasti aineiston käyttöä ja siten sen nykyistä kohdeyhteisöä. Toiset piirteet taas voivat olla hyvinkin merkityksettömiä nykyisen kohdeyhteisön kannalta. Piirteiden säilyttämisellä on myös kustannuksia: jokin piirre voi olla yksinkertaista säilyttää, jolloin säilyttämisestä aiheutuvat kustannukset ovat pienet, mutta jonkin muun piirteiden säilyttäminen käyttökelpoisena kohdeyhteisölle voi olla niin työlästä tai säilytyskapasiteettia kuluttavaa, että siitä koituu merkittäviä kustannuksia. Suunnittelussa on otettava huomioon sekin, miten yleinen jokin piirre on.

Piirteiden priorisointi ei onnistu vielä aineiston syntyessä tai edes paketoinnissa PAS-palveluun. Piirteiden painon määrittämiseen ei ole välttämättä ole riittävästi tietoa, ja kustannuksia voidaan arvioida vasta sitten, kun tiedetään migraation kohdetiedostomuoto, migraation laatuvaatimukset ja käytettävissä olevat migraatiotyökalut. Lisäksi säilytettävien piirteiden priorisointi voi muuttua kohdeyhteisön tai sen tarpeiden muuttuessa.

Jokin piirre voidaan jättää säilyttämättä, jos se arvioidaan nykyiselle kohdeyhteisölle suhteellisen merkityksettömäksi ja jos siihen vaadittava työ on arvioituun hyötyyn nähden kohtuutonta. Merkityksetön piirre voi olla esimerkiksi mahdollisuus muokata tiedoston sisältöä myöhemmin, jos ollaan varmoja siitä, että tiedoston sisältöä ei ole tarkoitus muokata koskaan. Ongelmana näissä ratkaisuissa on se, että nykyhetkessä merkityksetön piirre voi osoittautua tärkeäksi tulevaisuuden kohdeyhteisölle. Siksi voidaan myös etsiä kompromissia, jossa piirre säilyy käytön kannalta heikommassa muodossa kuin alun perin, mutta säilyttämisen kustannukset pysyvät kohtuullisina. Yksi esimerkki tällaisesta on alun perin interaktiivisen ominaisuuden säilyttäminen kuvana, jolloin varsinaisesta toiminnallisuudesta luovutaan.

Tulevia kohdeyhteisöjä ajatellen merkityksettömiksi arvioidutkin piirteet kannattaa säilyttää, jos se ei vaadi lisätoimia. On myös mahdollista, että jonkin tietyn piirteen säilyttäminen on kustannustehokkaampaa, kuin sen säilyttämättä jättäminen. Jos migraatiossa käytettävä sovellus tukee valmiiksi tietyn piirteen säilymistä, sen poistaminen tai muuttaminen vaatisi lisätyötä.

Vaatus	Kuvaus
V61	Hyödyntävät organisaatiot priorisoivat säilytettävän aineiston piirteitä hyödyntäen ihmisluettavaa säilytysuunnitelmapohjaa.
V62	PAS-palvelu tukee hyödyntyviä organisaatioita tiedostomuotojen ja aineistojen piirteiden tunnistamisessa ja priorisoinnissa tarjoamalla työkaluja ja malleja piirteiden priorisointiin.
V63	PAS-palvelu ja hyödyntävä organisaatio sopivat yhdessä tiedostomuotojen ja aineiston piirteiden priorisoinnista, ottaen huomioon tekniset vaatimukset ja kustannukset.

5.3 Säilytysuunnitelma

Jokaiselle säilytettävälle digitaaliselle objektille on oltava yksi siihen sovellettavissa oleva säilytysuunnitelma. Yksittäinen suunnitelma voi kuitenkin kohdistua useaan teknisesti samankaltaisiin objekteihin, kuten kaikkiin TIFF-tiedostoihin. Tällöin niihin voidaan soveltaa säilytystoimenpiteitä samoilla työkaluilla ja parametreilla. Toisaalta hyödyntävällä organisaatiolla voi myös olla tarpeita saman tiedostomuodon erilaisille säilytysuunnitelmille, jolloin samalle tiedostomuodolle voi olla useita säilytysuunnitelmia jotka määrittelevät erilaisia migraatiopolkuja. Hyödyntävällä organisaatiolla voi olla esimerkiksi yksi migraatiopolku TIFF-kuville ja toinen karttakuville riippumatta niiden tiedostomuodosta. Suunnitelmien erot eivät koske välttämättä migraatiota sinänsä vaan esimerkiksi muunnettujen tiedostojen validointia.

Jos jollekin objektille on sekä tiedostomuoto- että aineistokohtainen säilytysuunnitelma, hyödyntävä organisaatio valitsee migraatiossa sovellettavan suunnitelman. Aineistokohtaiset vaatimukset ovat tiedostomuotokohtaisia spesifimmät ja soveltuvat paremmin kyseiselle aineistolle. Toimimalla parhaiten sopivan suunnitelman mukaisesti sekä arvioimalla säilytyksen tavoitteet ja menetelmät aika-ajoin uudelleen voidaan varmistaa PAS-palveluun tallennettujen aineistojen esityskelpoisuus ja ymmärrettävyys pitkälläkin aikavälillä.

Jokaisella PAS-palveluun tallennetulla digitaalisella objektilla on oltava olemassa sitä koskeva tiedostomuoto- tai aineistokohtainen säilytysuunnitelma. Nekin tiedostomuodot, jotka ovat nyt säilytyskelpoisia, vanhenevat joskus. Lisäksi kohdeyhteisön tai hyödyntävän organisaation käyttämien sovelluksien muutokset voivat edellyttää migraatiota jo ennen kuin alkuperäinen tiedostomuoto vanhenee.

Alustavien suunnitelmien mukaan PAS-palvelu tekee yhdessä hyödyntävien organisaatioiden kanssa myöhemmin määriteltävällä tavalla valmiita tiedostomuotoisia mallisuunnitelmia, joihin hyödyntävä organisaatio voi viitata siirtopaketeissaan PAS-palvelun tehtävänä on laatia ja ylläpitää näitä mallisuunnitelmia hyvin dokumentoituina. Hyödyntävän organisaation tehtäväksi jää säilytysuunnitelman valinta ja sen kohdentaminen aineisto- tai tiedostotyypeittäin. Vaihtoehtoisesti hyödyntävä organisaatio voi mukauttaa tiedostomuotoista mallisuunnitelmaa ja toimittaa sen PAS-palveluun erikseen sovittavalla tavalla. Myös tällaisten mukautettujen säilytysuunnitelmien, tai ainakin osan niistä, on hyvä olla saatavilla julkisesti valmiina malleina niin, että muutkin hyödyntävät organisaatiot voivat niitä hyödyntää. On toki kiinnitettävä erityistä huomiota siihen, että jos mukautettu säilytysuunnitelma julkaistaan muiden organisaatioiden käyttöön, se ei voi sisältää mitään ei-julkista tietoa. Lisäksi on huomioitava, että hyödyntävät organisaatiot voivat tarvita säilytysuunnitelman laatimiseksi työkaluja, joilla voidaan määritellä valituista tiedostomuodoista keskeisiä ominaisuuksia ja määrittää näille aineisto- tai organisaatiokohtainen kriittisyys.

PAR (Preservation Action Registries) on malli, jonka avulla säilytystoimenpiteitä tukevaa tietoa voidaan välittää eri järjestelmien välillä. Suunnittelussa on mukana pitkäaikais säilytyksen merkittäviä kansainvälisiä toimijoita. PAS-palvelu seuraa PAR-järjestelmän kehitystä ja mahdollisuuksien mukaan osallistuu siihen. PAR ei välttämättä tuo valmiita malleja säilytysuunnitelmille, mutta jos hankkeen tuotokset tulevat laajalti käyttöön, sillä voi olla merkittäviä vaikutuksia säilytyksen suunnittelussa samoin kuin muissa säilytykseen liittyvissä toiminnoissa.

Vaatus	Kuvaus
V64	PAS-palvelu ylläpitää ihmislueattavia ja täydennettäviä säilytysuunnitelmapohjia.
V65	PAS-palvelu ylläpitää ja julkaisee tiedostomuotoisia mallisuunnitelmia ja työkalusovelluksia, jotka tukevat säilytysuunnitelmien laatimista.
V66	Hyödyntävät organisaatiot laativat tiedosto- ja aineistokohtaisia säilytysuunnitelmia mallisuunnitelmien pohjalta yhteistyössä PAS-palvelun kanssa ja liittävät ne osaksi säilytettävää aineistoa.
V67	Jokaisella säilytyksessä olevalla digitaalisella objektilla on oltava säilytysuunnitelma.
V68	Hyödyntävät organisaatiot liittävät aineistokohtaiset säilytysuunnitelmat säilytettävään aineistoon.
V69	Hyödyntävät organisaatiot ylläpitävät aineistokohtaisia säilytysuunnitelmia yhteistyössä PAS-palvelun kanssa, ottaen huomioon kohdeyhteisöjen ja niiden tarpeiden muutokset.
V70	Hyödyntävät organisaatiot valitsevat ja laativat säilytysuunnitelmia aineiston piirteiden priorisoinnin pohjalta ja kohdentavat suunnitelmat aineistotyypeittäin.
V71	PAS-palvelu laatii säilytysuunnitelmista koneluettavat versiot, joita hyödynnetään migraatioiden suunnittelussa.
V72	PAS-palvelu ylläpitää listaa säilytystoimenpiteistä.
V73	PAS-palvelu ylläpitää järjestelmää, joka osaa yhdistää koneluettavia säilytysuunnitelmia, säilytettävän aineiston teknisiä ominaisuuksia sekä mahdollisia aineistoon kohdistuvia säilytystoimenpiteitä.
V74	PAS-palvelu hyödyntää kansainvälisiä säilytyksen suunnitteluun liittyviä työkaluja ja järjestelmiä sekä osallistuu niiden kehitykseen.

6 MIGRAATIOT

Migraatio on tapahtuma, jossa säilytyspaketin tietosisältöä tai säilytystietoa muutetaan. Esimerkiksi siirtokelpoisen PDF-tiedoston muuntaminen säilytyskelpoiseen PDF/A tiedostomuotoon.

Migraatiossa säilytyspaketin tietosisältöä tai säilytystietoa muutetaan. Tiedostomuotojen migraatiot ovat keskeinen loogisen säilyttämisen toimenpide. Sen ansiosta kohdeyhteisöt voivat käyttää aineistoja kulloinkin ajantasaisilla ohjelmistoilla ja välineillä. Migraatio voi kohdistua säilytyksessä olevaan tiedostoon ja/tai sen metatietoihin. Migraatiossa tuotetaan aina sen kohteena olevasta tiedostosta uusi versio uudessa tiedostomuodossa. Pelkkä kuvailevien metatietojen päivittäminen/korjaaminen ei ole migraatio, vaan päivitystapahtuma, jossa edeltävä tietue korvataan uudella. Tiedoston migraation jälkeen vanha versio voidaan joko poistaa tai tallentaa uuteen säilytyspakettiin uuden tiedostomuodon rinnalle.

Tällä hetkellä akuuttia tarvetta migraatiolle ei ole näköpiirissä, koska PAS-palveluiden kehittämisessä on alusta alkaen huolellisesti määritelty PAS-palveluun sallitut tiedostomuodot. Poikkeus tästä on normalisointi, jossa hyödyntävä organisaatio muuntaa objektin säilytyskelpoiseen tiedostomuotoon. Normalisointi tapahtuu valmisteluvaiheessa PAS-palvelun ulkopuolella ennen aineistojen siirtoa säilytykseen. Normalisointia käsitellään tarkemmin luvussa 6.7.

6.1 Migraatio tarpeen tunnistaminen ja ennakoiminen

Migraatiotarpeen tunnistamisen ja ennakoimisen merkitys kasvaa tulevaisuudessa. Siihen tarvitaan muun muassa formaattikirjastoja (esim. PRONOM), joiden pohjalta voidaan analysoida tiedostomuodon migraatiotarve. Kansainvälinen teknologian seuranta on avainroolissa. Migraatiotarpeen tunnistamisessa on myös huomioitava, että tiedostomuodon versio voi vanhentua nopeammin kuin itse tiedostomuoto, ja hyödyntävän organisaation voi olla tarve toteuttaa migraatio muista syistä, kuten kohdeyhteisön muutoksen vuoksi tai levytilan säästämiseksi, jo ennen kuin migroitava tiedostomuoto vanhenisi teknisesti.

Migraatiotarpeen voi tunnistaa sekä PAS-palvelu (tiedostomuotojen ja migraatiotyökalujen seuranta ja työkalujen mahdollinen kehittäminen) että hyödyntävä organisaatio (aineiston käytön ja kohdeyhteisön sekä organisaation käytettävissä olevien sovellusten seuranta). Tarve migraatiolle ei aina synny vain teknisistä lähtökohdista. Hyödyntävän organisaation on aktiivisesti seurattava omaa toimintaympäristöään ja sen muutoksia ja informoitava näistä PAS-palvelua. On mahdollista, että hyödyntävän organisaation kohdeyhteisö muuttuu tai siirtyy käyttämään toista tiedostomuotoa, jolloin alkuperäisen tiedostomuodon käyttäminen on (uudelle) kohdeyhteisölle vaikeaa tai jopa mahdotonta, vaikka alkuperäinen tiedostomuoto olisi toiselle kohdeyhteisölle edelleen täysin relevantti. Tämä voi johtua esimerkiksi ohjelmistojen vaihtumisesta tai niiden käyttöoikeuksien poistumisesta. Migraatiotarpeen vahvistaminen ja sitä kautta migraation suunnittelun käynnistäminen on aina hyödyntävän organisaation vastuulla. PAS-palvelu ei tee migraatiota ilman hyödyntävän organisaation hyväksyntää.

Vaatus	Kuvaus
V75	PAS-palvelu seuraa tiedostomuotojen ja metatietoformaattien elinkaarta.
V76	PAS-palvelu seuraa myös säilytysuunnitelmissa esitettyjen aineistojen kriittisten ominaisuuksien vanhenemista.
V77	PAS-palvelu seuraa migraatiotyökalujen kehitystä.
V78	PAS-palvelu tunnistaa tiedostomuodon tai metatietoformaatin vanhenemisen ja tiedottaa siitä hyödyntäville organisaatiolle.
V79	PAS-palvelu seuraa säilytysmenetelmien ja mallien kehitystä ja tunnistaa niiden vaikutukset aineiston säilyttämiselle.

Vaatus	Kuvaus
V80	Hyödyntävät organisaatiot seuraavat tiedostomuotojen ja metatietoformaattien elinkaarta omassa toimintaympäristössään kohdeyhteisön muutokset ja muuttuvat tarpeet huomioon ottaen.
V81	Hyödyntävä organisaatio ilmoittaa PAS-palvelulle tarpeesta aloittaa jonkin tiedostomuodon tai aineiston migraatio.

6.2 Migraatiopolkujen määrittäminen tiedostomuodoille

Tiedostomuotojen ryhmittely tietyn tyyppisiin muotoihin, joilla on yhteisiä piirteitä, ja näiden ryhmien ominaisuuksien käsittely johtaa ymmärrykseen migraatioiden kulusta. Jokaiselle aineistolle on lisäksi erilaisia yksilöllisiä piirteitä, jotka on migraatioissa huomioitava tiedostomuotokohtaisten piirteiden lisäksi.

Migraatiopolkujen määrittelyssä on arvioitava eri vaihtoehtoja seuraavia asioita koskien: migroitavan tiedoston kohdeformaatti (joskus monikossa kohdeformaatit), aineiston kohdeyhteisön tarpeet, migraation laatuvaatimukset (millaiset sisällön ja ulkoasun muutokset ovat sallittuja), käytettävissä olevat migraatiotyökalut, kustannukset (mm. ohjelmointityö) sekä muut sellaiset tekijät, joilla on merkitystä aineiston kohdeyhteisön näkökulmasta.

Tietulle lähdeformaatile voi olla useita mahdollisia migraatiopolkuja. Säilytyksen suunnittelussa on tärkeää arvioida eri käytettävissä olevat vaihtoehdot eri näkökulmista. Arviointi auttaa migraatiopolkujen priorisoinnissa, mutta se ei voi olla ratkaiseva tekijä. Joskus voi olla järkevää valita esimerkiksi toiseksi parhaan kokonaisarvion saanut migraatiopolku, jos siihen on selkeät perusteet, kuten parhaan arvion saaneen polun liian korkeat kustannukset. Valitun migraatiopolun perustelut on aina kirjattava, jotta näihin voidaan tulevaisuudessa palata tarvittaessa. Jos alkuperäinen versio aineistoista säilytetään, migraatio voidaan tehdä uudestaan soveltaen muita työkaluja, jos yksittäisessä dokumentissa havaitaan merkittäviä ongelmia. Alkuperäisen dokumentin säilyttäminen mahdollistaa myös sen, että kun migraatioissa tuotettu tiedostomuoto vanhenee, seuraavakin migraatio voidaan tehdä alkuperäisaineistosta, jos tarvittavat migraatiotyökalut ovat joko valmiiksi olemassa tai riittävän helposti kehitettävissä.

Migraatiotapahtuman tiedot (ajankohta, toimijat, migraation aiheuttamat muutokset) tulee tallentaa säilytyspaketteihin pitkäaikaissäilytyksen metatietona, jonka tulee olla hyödyntävien organisaatioiden käytettävissä. Erityisen tärkeää on pyrkiä tallentamaan sisällön ja ulkoasun muutokset mahdollisimman kattavasti.

Vaatus	Kuvaus
V82	PAS-palvelu määrittelee siirtokelpoisille tiedostomuodoille tiedostomuotokohtaisia migraatiopolkuja.
V83	PAS-palvelu määrittelee säilytyskelpoisille tiedostomuodoille tiedostomuotokohtaisia migraatiopolkuja.
V84	PAS-palvelu seuraa siirtokelpoisten tiedostomuotojen elinkaarta, ja tiedottaa hyödyntäville organisaatioille, kun näiden tiedostomuotojen elinkaari on päättymässä.
V85	PAS-palvelu seuraa säilytyskelpoisten tiedostomuotojen elinkaarta, ja tiedottaa hyödyntäville organisaatioille, kun näiden tiedostomuotojen elinkaari on päättymässä.
V86	PAS-palvelu suunnittelee migraation uudelle siirtokelpoiselle tiedostomuodolle yhdessä hyödyntävien organisaatioiden kanssa ennen tiedostomuodon lisäämistä tiedostomuotomäärittelyyn.
V87	PAS-palvelu aloittaa migraation suunnittelun säilytyskelpoisille tiedostomuodoille yhdessä hyödyntävien organisaatioiden kanssa, kun tarve migraatiolle on tunnistettu.

Vaatus	Kuvaus
V88	Hyödyntävä organisaatio valitsee tiedostomuoto- ja aineistokohtaisesti sopivat migraatiopolut osana migraation suunnittelua.

6.3 Migraatioiden suunnittelu

Edellisessä luvussa kuvatusi migraatio on keskeinen loogisen säilyttämisen toimenpide, joka tulee aina suunnitella huolellisesti. Suunnittelussa korostuvat vähintään yleinen, esimerkiksi tiettyyn tiedostomuotoon kohdistuva, yleinen suunnittelu ja tapauskohtainen, tiettyyn aineistokokonaisuuteen kohdistuva suunnittelu, joka yleensä koskee vain tiettyä hyödyntävää organisaatiota.

Migraation suunnittelun tarkoitus on varmistua mahdollisimman pitkälle migraation lopputuloksesta ottaen huomioon käytettävissä olevat työkalut, resurssit, kustannukset, aineiston ominaisuudet ja olemassa olevat säilytysuunnitelmat. Migraation hyväksyminen edellyttää tietoa migraation seurauksista sekä PAS-palveluille että hyödyntävälle organisaatiolle. Tässä vaiheessa (viimeistään) hyödyntävän organisaation on syytä varmistaa, että aineiston säilytysuunnitelma on edelleen paikkansapitävä, ja tarvittaessa päivittää se. Tässä vaiheessa myös laaditaan migraatiolle kustannusarvio, johon vaikuttavat muun muassa migroitavan aineiston määrä, tarvittava säilytyskapasiteetti ja henkilötyön tarve PAS-palvelussa ja hyödyntävässä organisaatiossa.

Tyypillisesti hyödyntävällä organisaatiolla on säilytyksessä sekä lukumäärältään että datamäärältään huomattavia määriä aineistoja, joten migraatiot ja niiden laadunvarmistus on automatisoitava mahdollisimman pitkälle. On tärkeää, että suunnitteluvaiheessa tehdään koemigraatio aineisto-otokselle, jonka hyödyntävä organisaatio on valinnut. Varsinainen migraatio voidaan käynnistää vasta hyödyntävän organisaation hyväksytyä tämän koemigraation tulokset.

Jos PAS-palvelu ja hyödyntävä organisaatio eivät pääse yksimielisyyteen migraation tavoitteista, hyödyntävä organisaatio päättää etenemistavasta. Migraatiota voidaan lykätä myöhemmäksi, tai sen toteuttamiseen voidaan hankkia lisäresursseja.

Vaatus	Kuvaus
V89	PAS-palvelu tarjoaa työkaluja ja muita resursseja migraatioiden toteuttamiseen.
V90	Hyödyntävä organisaatio valitsee migraatiototeutuksen säilytysuunnitelmien pohjalta yhteistyössä PAS-palvelun kanssa.
V91	PAS-palvelu suunnittelee migraatiototeutuksen automatisoinnin sekä tulosten analysoinnin ja raportoinnin yhdessä hyödyntävän organisaation kanssa.
V92	PAS-palvelu suunnittelee teknisten ja hallinnollisten metatietojen päivittämisen migraation yhteydessä.
V93	PAS-palvelu analysoi migraation vaikutuksen aineistoon säilytysuunnitelman pohjalta ja raportoi siitä hyödyntävälle organisaatiolle.
V94	Hyödyntävä organisaatio valitse aineisto-otoksen testausta varten.
V95	Hyödyntävä organisaatio arvioi migraation vaikutukset omaan toimintaympäristöönsä nähden.
V96	Hyödyntävä organisaatio varautuu päivittämään metatietoja migraation yhteydessä.
V97	Hyödyntävä organisaatio suunnittelee yhdessä PAS-palvelun kanssa, kuinka migroitua aineistoa ja metatietoa siirretään sen taustajärjestelmiin.
V98	Hyödyntävä organisaatio päättää mitä migroidulle aineistolle tehdään ja kuinka uusi aineisto sovitetaan aineiston elinkaaren hallintaan.
V99	PAS-palvelu tukee hyödyntävää organisaatiota migraation tulosten arvioinnissa.

Vaatus	Kuvaus
V100	PAS-palvelu suorittaa koemigraation
V101	Hyödyntävä organisaatio tarkastaa koemigraation tulokset ja hyväksyy/hylkää tulokset.
V102	Hyödyntävä organisaatio tekee päätöksen migraation etenemisestä.
V103	Jos PAS-palvelu ja hyödyntävä organisaatio eivät saavuta yksimielisyyttä migraation tavoitetasosta, hyödyntävä organisaatio päättää jatkosta.
V104	Migraation valmisteluvaiheessa PAS-palvelu dokumentoi alkuperäisen tiedostomuodon ja mahdollisten uusien tiedostomuotojen erot eli tekniset ominaisuudet, joiden migraatio ei ole mahdollista.

6.4 Migraatioiden toteuttaminen ja laadunvarmistus

Migraation suunnittelun ja testauksen jälkeen voidaan käynnistää itse migraatio, jonka PAS-palvelu toteuttaa säilytys suunnitelmien ja migraation suunnittelun perusteella. PAS-palvelun toteuttamien migraatioiden yhteydessä PAS-palvelu päivittää hallinnollisiin metatietoihin tarvittavat muutokset

Kuvailevien metatietojen päivittäminen on hyödyntävän organisaation vastuulla. Migraatiolla tuotetun uuden objektin metatiedot voidaan luoda kahdella tapaa. Yksinkertaisin malli on generoida ne PAS-palvelussa alkuperäisen objektin metatiedoista hyödyntävän organisaation toimittamien määritysten mukaisesti. Muutostarve koskee mm. objektin tunnusta sekä teknisiä metatietoja (tiedostomuoto, tiedoston koko).

Metatieto on mahdollista luoda myös hyödyntävän organisaation tuotantojärjestelmässä sen jälkeen, kun migroitu aineisto teknisine metatietoineen on siirretty sinne PAS-palvelusta. Uusi kuvaileva metatieto lähetetään PAS-palveluun siirtopakettina, joka päivittää PAS-palveluun luotua säilytyspakettia. Jälkimmäistä toimintamallia sovelletaan myös silloin, jos PAS-palveluun siirretyn aineiston kuvailevia metatietoja päivitetään hyödyntävän organisaation tuotantojärjestelmässä tai jos kuvailevan metatiedon formaatti muuttuu.

Vaikka migraatio olisi suunniteltu huolellisesti ja sekä PAS-palvelu että hyödyntävä organisaatio olisivat hyväksyneet koemigraation, voidaan varsinaisen migraatioprosessin aikana huomata, että tulokset eivät aina ole haluttuja. Tällöin migraatioprosessi on keskeytettävä mahdollisimman nopeasti. Jotta mahdolliset ongelmat voidaan havaita, on sekä PAS-palvelun että hyödyntävän organisaation seurattava migraatioprosessia aktiivisesti. Ongelman voi huomata joko PAS-palvelu tai hyödyntävä organisaatio, joten kumpi tahansa osapuoli voi pyytää migraatioprosessin keskeyttämistä. Tässä dokumentissa ei oteta kantaa siihen, millainen peruste riittää migraation keskeyttämiseen, koska syyt voivat olla hyvin moninaisia riippuen tiedostomuodosta ja aineistoryhmästä. Jos migraatioprosessi keskeytetään, se voidaan aloittaa alusta, mutta vasta sitten kun mahdolliset ongelmat on ratkaistu.

Hyödyntävä organisaatio voi tehdä migraation myös itse. Tällöin se noutaa aineistot PAS-palvelusta jakelupaketteina, tekee migraation haluamilleen aineistoille ja siirtää migroidut aineistot metatietoineen takaisin PAS-palveluun siirtopaketteina. Tämä tapa migroida on mahdollinen jo nykyisessä PAS-palvelussa. Se soveltuu tiedostomuodoille tai aineistoille, joita on säilytyksessä suhteellisen vähän ja joiden migraatioprosessi on hyvin monimutkainen tai vaatii työkaluja/osaamista joita PAS-palveluilla ei ole tarjota.

Vaatus	Kuvaus
V105	PAS-palvelu toteuttaa aineiston migraation migraatiosuunnittelun pohjalta.
V106	PAS-palvelu validoi uudet tiedostot.
V107	PAS-palvelu analysoi ja raportoi toteutetun migraation tulokset ja tilastot.
V108	PAS-palvelu päivittää tekniset ja hallinnolliset metatiedot, mukaan lukien kuvauksen migraatiosta.

Vaatus	Kuvaus
V109	PAS-palvelu luo uudet säilytyspaketit migroidusta ja päivitetystä aineistosta ja metatiedoista.
V110	Hyödyntävä organisaatio vastaa kuvailevien metatietojen päivittämisestä uudelle aineistoversiolle tai tarjoaa PAS-palvelulle välineet, joiden avulla nämä metatiedot voidaan luoda alkuperäisen aineiston metatiedoista.
V111	Migraatiota pitää pystyä keskeyttämään sekä PAS-palvelun että hyödyntävän organisaation toimesta.
V112	Hyödyntävä organisaation voi toteuttaa PAS-palvelun valmisteleman tai muun migraation omissa järjestelmissään.

6.5 Migraatioiden päättäminen ja hyväksyntä

Migraatio on aina hyödyntävän organisaation kanssa sovittava toimenpide/prosessi, joka kohdistuu ennalta sovittuun rajalliseen joukkoon digitaalisia objekteja. Migraation valmistuttua PAS-palvelu informoi tästä hyödyntävää organisaatiota, joka hyväksyy migraation tulokset. Hyödyntävä organisaatio voi perustelluista syistä hylätä migraation tulokset myös tässä vaiheessa, vaikka mitään ongelmia ei olisi aiemmin havaittu. Jos migraatioprosessin tulokset syystä tai toisesta hylätään, on hyödyntävän organisaation ja PAS-palvelun sovittava yhteistyössä jatkotoimien valmistelusta.

Riippumatta siitä hyväksyttiin vai hylättiin migraation lopputulos, on tärkeää, että lopputulos ja kokemukset dokumentoidaan julkisesti myös muille PAS-palvelua hyödyntäville organisaatioille mahdollisuuksien mukaan ja tarvittaessa myös kansainvälisesti (esim. PAR-järjestelmään).

Migraation jälkeen voidaan migraation tuloksena syntyneet aineistot siirtää organisaation taustajärjestelmään jakelupaketteina, jos tähän on tarvetta.

Vaatus	Kuvaus
V113	Migraatiosta tuotetaan loppuraportti, jonka keskeiset tulokset julkaistaan tarvittaessa myös kansainvälisesti.
V114	Hyödyntävä organisaatio hyväksyy tai hylkää migraation tulokset.
V115	Jos migraation tulokset hylätään, PAS-palvelu ja hyödyntävä organisaatio sopivat yhteistyössä jatkotoimenpiteistä.
V116	Hyödyntävä organisaatio voi siirtää migraation tuloksena syntyneet aineistot ja niiden metatiedot organisaation taustajärjestelmiin jakelupaketteina.

6.6 Vanhojen versioiden käsittely

Migraatiossa syntyy aina uusi versio migraation kohteena olevasta objektista, joka migraation määritelmän mukaan korvaa version josta migraatio tehtiin. Voi kuitenkin olla tarpeellista – tai jopa välttämätöntä – säilyttää myös vanha(t) versio(t) esimerkiksi aineiston autenttisuuden varmistamiseksi. Vanhojen versioiden säilyttäminen vie kuitenkin hyödyntävän organisaation säilytyskapasiteettia, joten migraation jälkeen hyödyntävän organisaation tulee tehdä päätös, poistetaanko vanha versio säilytyksestä vai jatketaanko sen säilyttämistä edelleen.

Joskus migraatio voidaan tehdä täysin häviöttömästi. Esimerkiksi joissakin kuvatiedostojen migraatioissa voidaan varmistua siitä, että kaikki kuvan pikselit ja muut ominaisuudet säilyvät täysin ennallaan, vaikka tiedostomuoto muuttuu. Tällöin vanhan tiedostomuodon säilyttäminen voi olla tarpeellista. Yleensä on kuitenkin jo ennalta selvää, että kaikki tiedoston ominaisuudet eivät säily. Tällöin tulee harkita huolellisesti kannattaako vanha versio säilyttää, sillä vaikka menetettävät piirteet eivät vielä migraatiota toteutettaessa tuntuisi tärkeiltä, ne voivat tulevaisuudessa osoittautua sellaisiksi kohdeyhenteisöjen ja niiden tarpeiden muuttuessa.

Riittävän pitkän ajan kuluttua monet PAS-palvelussa olevat tiedostot on migroitu jo useampaan kertaan, eli samasta objektista voi säilytyksessä olla useampia versioita, jos edellinen versio on migraatioiden yhteydessä aina päätetty säilyttää. PAS-palvelun toteutuksessa on oltava mahdollisuus poistaa joku tai jotkin väliversiot myöhemmin siten, että looginen ketju jäljelle jääneiden versioiden välillä säilyy.

Vaatus	Kuvaus
V117	Vanhoja migroituja aineistoversioita on mahdollista poistaa. Päätöksen poistosta tekee hyödyntävä organisaatio ottaen huomioon migraation laadun ja käytettävissään olevan säilytyskapasiteetin.
V118	Vanhat aineistoversiot on mahdollista linkittää uusin migraatioissa syntyneisiin aineistoversioihin, eikä ketju saa katketa, vaikka välistä poistettaisiin yksi tai useampia versioita.

6.7 Normalisointi

Normalisointi on tapahtuma, jossa objekti muunnetaan säilytyskelpoiseen tiedostomuotoon.

Mikäli säilytettävän aineiston tiedostomuoto ei ole säilytys- eikä siirtokelpoinen, tiedostot tulee muuntaa suoraan säilytyskelpoiseen tiedostomuotoon ennen aineiston siirtämistä PAS-palveluun. Tätä muunnosta kutsutaan normalisoinniksi. Normalisoinnista vastaa ensisijaisesti hyödyntävä organisaatio, jolle PAS-palvelut voivat tarjota muunnoksiin liittyvää tukea (esim. osaamista ja tietoa soveltuvista migraatiotyökaluista). Ennen normalisointia on suositeltavaa olla yhteydessä PAS-palveluun, jotta tiedostot muunnetaan mahdollisimman tarkoituksenmukaiseen tiedostomuotoon mahdollisimman tarkoituksenmukaisella prosessilla. Tiedostot pitää muuntaa suoraan säilytyskelpoiseen tiedostomuotoon, koska muutoin PAS-palvelussa jouduttaisiin tekemään vielä uusi muunnos siirtokelpoisesta säilytyskelpoiseen muotoon. Jokainen tiedostomuodon muunnos – migraatio tai normalisointi – on riski datan säilyvyydelle, joten muunnoksien määrä tulee minimoida.

Normalisoinnissa alkuperäinen tiedostomuoto voi olla periaatteessa mitä tahansa. Toisin kuin migraatioihin, normalisointiin eivät PAS-palvelu tai edes hyödyntävä organisaatio voi varautua samalla tavalla ja esimerkiksi suunnittelemalla valmiita migraatiopolkuja tai kehittämällä valmiita säilytysuunnitelmamalleja. Näitä tapauksia varten hyödyntävä organisaatio ja PAS-palvelut tekevät tiivistä yhteistyötä.

Koska normalisointeihin ei voi valmistautua yhtä hyvin kuin migraatioihin, normalisoinnin jälkeen voi olla tarvetta säilyttää myös alkuperäinen tiedosto, esimerkiksi aineiston autenttisuuden takaamiseksi. Alkuperäiset tiedostot voidaan siirtää PAS-palveluun bittitason säilytykseen normalisoitujen tiedostojen kera.

Jos riittävän laadukas normalisointi ei ole teknisesti mahdollista, edellyttäisi liikaa resursseja tai jos bittitason säilytys on hyväksyttävä ratkaisu, aineisto voidaan siirtää PAS-palveluun normalisoimattomana.

Vaatus	Kuvaus
V119	PAS-palvelu tarjoaa tukea hyödyntäville organisaatioille aineiston normalisoinnissa tarjoamalla tietämystä, työkaluja sekä tukea normalisoinnin tulosten analysoinnissa.
V120	Ellei muuta sovita, hyödyntävä organisaatio vastaa aineiston normalisoinnista ja normalisoidun aineiston siirrosta PAS-palveluun, tarvittaessa yhdessä alkuperäisen aineiston kanssa.
V121	Normalisointiprosessi on syytä dokumentoida ja julkaista, jotta muut hyödyntävät organisaatiot voivat tarvittaessa soveltaa samaa prosessia.

7 EMULOINTI

Tiettyjen erityisaineistojen kohdalla aineistojen hyödyntäminen voi vaatia emulointia, jossa jäljitellään, tyypillisesti vanhentuneen, laitteen tai järjestelmän toimintaa ja mahdollistetaan tämän avulla aineistojen käyttö nykyisissä ympäristöissä. Tällaisiin aineistoihin sisältyy esimerkiksi tietokonepelejä, -ohjelmistoja ja mediataidetta.

Emulaatiossa lähtökohdaksi otetaan vanhentuneen käyttöjärjestelmä- ja sovellusohjelmistoympäristön käyttäminen uudistuneessa laiteympäristössä siten että alkuperäinen käyttökokemus säilyy. Tällöin vanhentuneen tiedostomuodon esittämiseen hyödynnetään samaa sovellusohjelmistoa ja käyttöjärjestelmää kuin aineiston luontihetkellä on ollut käytössä.

KDK-hankkeen aikana sovittiin, että emulointi ei ole PAS-palvelun tukema säilytysmenetelmä, vaan hyödyntävä organisaatio huolehtii emuloinnista emulointia vaativille aineistoille. Organisaatio voi kuitenkin hyödyntää PAS-palvelua myös tällaisten aineistojen ja esim. niiden vaatimien sovellusohjelmistojen bittitason säilytykseen. Lisäksi PAS-palvelu tarjoaa yhteistyöverkoston, jonka avulla näitä aineistoja omistavat organisaatiot voivat vaihtaa kokemuksia ja auttaa toisiaan aineistojen säilyttämisessä.

Vaatus	Kuvaus
V122	PAS-palvelu huolehtii emuloitavan aineiston bittitason säilyttämisestä.
V123	Hyödyntävä organisaatio vastaa aineiston emuloinnista.
V124	PAS-palvelu selvittää vuosittain hyödyntävien organisaatioiden emulaatiotarpeita. Emulaatioon liittyvien kehittämis- ja palvelutehtävien (esim. emulaatiosovellusten kehittäminen ja/tai käytön tuki) lisäämisestä PAS-palvelun vastuisiin päätetään saatujen tietojen nojalla yhdessä Opetus- ja kulttuuriministeriön kanssa.

LÄHTEET

- [API] PAS-palveluiden rajapinnat. 2020. [urn:nbn:fi-fe2020100578097](https://nbn-resolving.org/urn:nbn:fi-fe2020100578097)
- [METADATA] Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen. 2020. [urn:nbn:fi-fe2020100578093](https://nbn-resolving.org/urn:nbn:fi-fe2020100578093)
- [TDSTO] Säilytys- ja siirtokelpoiset tiedostomuodot. 2020. [urn:nbn:fi-fe2020100578095](https://nbn-resolving.org/urn:nbn:fi-fe2020100578095)