

PAS-PALVELUIDEN RAJAPINNAT

v2.0.2

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



Lisenssi

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

Tämän määrittymisen 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 lisenssinsaaajaan 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
1.1	PAS-palvelut	4
1.2	Rajapinnat PAS-palveluissa	4
1.3	Käyttötapauksia	4
1.3.1	Aineistojen siirtäminen PAS-palveluihin	5
1.3.2	Aineistojen noutaminen PAS-palveluista	5
1.4	Päivittyvä määrittely	6
2	AINEISTOJEN SIIRTÄMINEN PAS-PALVELUIHIN	7
2.1	Autentikointi ja valvonta	7
2.2	Käyttäjän hakemistot	7
2.3	Siirtäminen PAS-palveluihin	7
2.4	Vastaanotossa hylätyn aineiston käsittely	8
3	AINEISTOJEN HAKU JA HALLINTA	9
3.1	Autentikointi ja valvonta	9
3.2	Viestien rakenne	9
3.2.1	Kyselyviestit	9
3.2.2	Palautusviestit	9
3.3	Hakuviestit	10
3.3.1	Hakutoiminto	12
3.3.1.1	Hakutoiminnon ominaisuudet ja syntaksi	12
3.3.1.2	Avain/arvo-parit	12
3.3.1.3	Tietopaketin tyyppin rajaaminen	13
3.4	Säilytyspakettien hallinta	13
3.4.1	Säilytyspakettiin kohdistuvat komennot	13
3.4.2	Säilytyspaketin muodostus jakelupaketiksi	13
3.5	Jakelupaketit	14
3.5.1	Jakelupaketin seuranta ja kohdistuvat komennot	15
3.5.2	Muodostettujen jakelupakettien noutaminen	15
3.6	Julkiset avaimet	17
LIITE A.	SIIRTOPAKETIN TARKASTUSRAPORTTI	18
A.1.	Objektit	18
A.2.	Tapahtumat	19
A.3.	Agentit	22
LÄHTEET		23

1 JOHDANTO

Tässä dokumentissa esitellään PAS-palveluiden rajapinnat, joiden avulla kansallisia PAS-palveluja hyödyntävät organisaatiot voivat mm. siirtää aineistoa säilytykseen PAS-palveluihin tai hakea niistä aineistoa. Rajapinnat on tarkoitettu käytettäväksi erityisesti tietojärjestelmien välisessä integraatiossa. Rajapintojen suunnittelussa on tavoiteltu mahdollisimman yksinkertaista, mutta skaalautuvaa ja myöhemmin uusilla toiminnallisuuksilla täydennettävissä olevaa mallia.

1.1 PAS-palvelut

PAS-palveluilla tarkoitetaan kulttuuriperintöaineistojen ja tutkimusaineistojen pitkäaikaissäilyttämiseen tuotettuja palveluita yhdessä.

Kulttuuriperintöaineistojen pitkäaikaissäilyttämiseen tuotettu 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 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.

1.2 Rajapinnat PAS-palveluissa

Siirrettävä aineisto paketoidaan Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen [META] sekä Säilytys- ja siirtokelpoiset tiedostomuodot [TDSTO] -määrytyksien mukaisesti ja siirretään PAS-palveluun SFTP-yhteydellä [IETF_SFTP] tässä dokumentissa määritellyn rajapinnan avulla. Aineiston siirrosta syntyy aina HTML- ja PREMIS-muotoiset [PREMIS] tarkastusraportit, jotka niin ikään ovat saatavilla SFTP-yhteyden kautta.

Aineiston hakeminen, muodostus jakelupaketeiksi ja noutaminen tapahtuu hyödyntäen HTTP-protokollaa [RFC_7230]. Tässä dokumentissa määritellään tähän viestintään tarvittavat REST-kysely- ja HTTP-vastausviestit [RFC_7231]. Rajapinnan vastausviestien rungon rakenteena käytetään RFC 7159 -määrytyksen [RFC_7159] mukaista JSON-muotoa lukuun ottamatta varsinaista säilytyksestä haettavaa aineistoa. Jakelupaketit toimitetaan ZIP- tai TAR-pakettina, ja jakelupaketin metatiedot XML-muodossa.

Virhetilanteissa tai muutoin poikkeavassa viestittelyssä käytetään SFTP-protokollan osalta koodistoa IETF-muistiossa [IETF_SFTP] esitetyllä tavalla ja REST-rajapinnan osalta HTTP-statuskoodistoa RFC 7231 -määrytyksessä [RFC_7231] esitetyllä tavalla, mutta kuitenkin yleisesti vakiintuneet tavat ja käytännöt huomioiden, mikäli tässä määrytyksessä ei muuta määritetä.

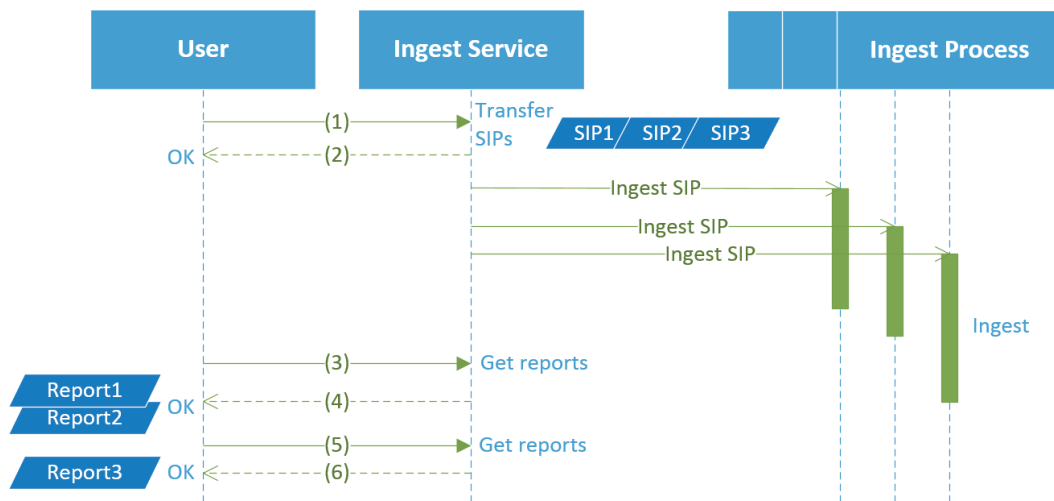
Tässä dokumentissa käyttäjällä tarkoitetaan käyttäjätunnuskohtaista PAS-palvelun rajapintojen hyödyntäjää, jolla tehtyjen sopimusten mukaisesti on tarvittavat oikeudet aineiston käsittelyyn.

1.3 Käyttötapauksia

Tässä luvussa esitellään kaksi yleistä käyttötapauksia rajapinnoille: Aineistojen siirtäminen PAS-palveluihin sekä Aineistojen noutaminen PAS-palveluista. Käyttötapauksien tarkoituksena on antaa yleiskuva rajapinnoista. Esitetyt käyttötapaukset ovat yksinkertaisia esimerkkejä, eikä niiden tarkoituksena ole kuvata rajapinnan toimintoja tyhjentävästi.

1.3.1 Aineistojen siirtäminen PAS-palveluihin

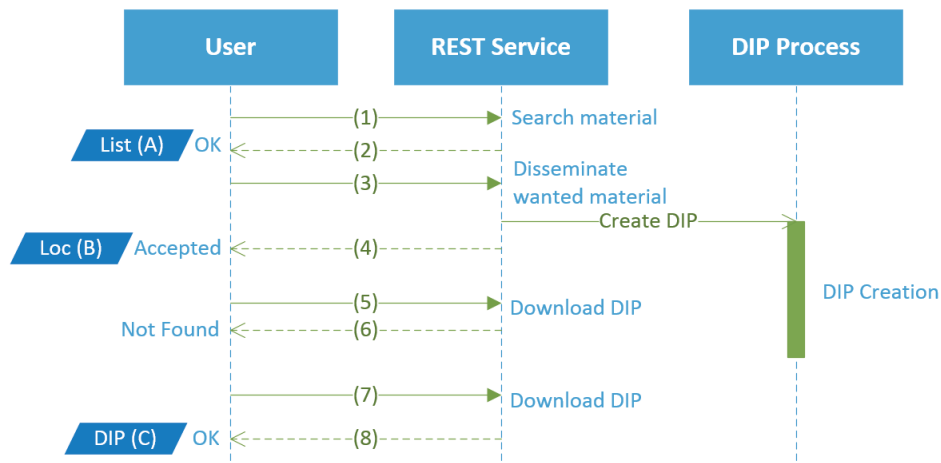
Aineistojen siirtäminen PAS-palveluihin on esitetty kuvassa 1. Aluksi käyttäjä avaa SFTP-yhteyden PAS-palveluun kirjautumalla PAS-palvelimelle ja siirtää paketoitumääritysten [META, TDSTO] mukaisesti luotuja siirtopaketteja palvelimelle. Kuvassa 1 siirretään kolme siirtopakettia. SFTP-protokolla itsessään kuittaa siirrot (2). Siirtopaketin tarkastusprosessi kestää jonkin aikaa. Kun aineisto on PAS-palvelussa siirretty säilytykseen, tai jos aineisto hylätään, saa käyttäjä kotihakemistoonsa tarkastusraportin siirrosta. Käyttäjä voi tavallisia SFTP-operaatioita käyttäen hakea, noutaa ja poistaa saamiaan tarkastusraportteja. Kuvan 1 kyselyssä (3) noudetaan raportteja hetkellä, jossa kaksi siirtopakettia on jo käsitelty, kun taas yksi on käsittelemättä. Palautuksena käyttäjä saa käsiteltyjen pakettien raportit (4). Viimeisenkin tarkastusprosessin päättyessä voidaan noutaa viimeinenkin raportti (5, 6). SFTP-rajapinta ja aineiston siirto on määritelty tarkemmin luvussa 2, ja sitä käytetään SFTP-protokollamäärityksen [IETF_SFTP] mukaisella tavalla.



Kuva 1: Aineiston siirtäminen PAS-palveluun ja tarkastusraportin hakeminen.

1.3.2 Aineistojen noutaminen PAS-palveluista

Aineistojen noutaminen PAS-palveluista on tiivistetysti esitetty kuvassa 2. Käyttäjä etsii säilytykseen siirtämäänsä aineistoa (1) ja saa vastausviestissä (2) aineistoluettelon (A) luvussa 3.3 määritellyllä tavalla. Tämän jälkeen käyttäjä valitsee luettelosta haluamansa aineiston (tai aineiston osan) ja tekee siitä jakelupaketin luontipyyntöä (3). Käyttäjä saa palautteen, joka kertoo, että jakelupaketin muodostusprosessi on aloitettu, mutta se ei ole vielä valmis noudettavaksi (4). Palautteen mukana saadaan jakelupaketin tuleva sijainti (B). Jakelupaketin valmistuminen kestää jonkin aikaa. Jos käyttäjä yrittää noutaa jakelupakettia ennen kuin se on valmis (5), PAS-palvelu palauttaa Not Found -viestin (6). Kun jakelupaketti on valmistunut, käyttäjä voi lähettää noutopyynnön onnistuneesti (7). Palautusviestissä (8) toimitetaan OK-viesti ja sen mukana jakelupaketti (C). Aineiston haku sekä jakelupaketin luonti ja nouto on määritelty tarkemmin luvussa 3, ja siinä käytetään REST-rajapintaa [RFC_7231] HTTP-protokollan [RFC_7230] yli.



Kuva 2: Aineiston hakeminen ja jakelupaketin noutaminen PAS-palvelusta.

1.4 Päivittyvä määrittäminen

Tämän määrittäksen ajanmukaisuus arvioidaan vuosittain, ja sitä kehitetään eteenpäin hyödyntävien organisaation tarpeiden mukaisesti. Olemassa olevia toiminnallisuuksia voidaan esimerkiksi rikastaa kattavammilla vastausviesteillä. Lisäksi erilaisia uusia toiminnallisuuksia voidaan toteuttaa. Näitä voivat olla esimerkiksi tarkastusraporttien haku- ja noutotoiminto REST-rajapinnan kautta sekä poistotoiminnot säilytyksessä olevalle aineistolle.

Määrittäksen kolmitasoinen versionumerointi kuvaa, minkä tasoisia muutoksia määrittäykseen on tehty. Jos ensimmäinen numero päivittyy, se tarkoittaa merkittäviä muutoksia myös hyödyntävien organisaatioiden järjestelmiin, jotka on integroitu PAS-palveluun. Keskimmäisen numeron kasvu kuvaa tilanteita, joissa muutoksilla voi olla vaikutuksia organisaatioiden järjestelmiin. Kolmannen numeron muutos kuvaa tilanteita, joissa määrittäykseen on tehty vähäisempiä korjauksia tai täsmennyksiä. Tällaisia voivat olla vaikkapa määrittäyksissä olevien esimerkkien täydentäminen.

Tässä dokumentissa esitetty REST-rajapinta ei ole yhteensopiva vastaavissa määrittäversioissa 1.0.X esitetyn rajapinnan kanssa.

2 AINEISTOJEN SIIRTÄMINEN PAS-PALVELUIHIN

Aineistojen PAS-palveluihin siirtämiseen sekä siirtopakettien tarkastusraporttien noutamiseen käytetään SFTP-rajapintaa määräyksensä [IETF_SFTP] mukaisesti, jolloin aineistojen siirto tapahtuu suojatussa yhteydessä. SFTP-rajapinnan avulla voi lisäksi noutaa luotuja jakelupaketteja.

2.1 Autentikointi ja valvonta

SFTP-yhteydessä autentikointiin käytetään julkisen ja yksityisen avainten pareja käyttöönoton yhteydessä erikseen saatavan ohjeen mukaisesti. Käyttäjä ei voi käsitellä toisen käyttäjän siirtämää aineistoa, eikä voi siirtää aineistoa toisen käyttäjän nimiin. Jokaisesta SFTP-yhteydestä tallentuu lokiin vähintään seuraavat tiedot automaattisesti:

- Kuka teki? (käyttäjätunnus)
- Mitä teki? (komennot, lähetetyt ja vastaanotetut tavumäärät)
- Milloin teki? (aikaleimat)
- Mistä teki? (IP-osoite, josta yhteys otettiin)

2.2 Käyttäjän hakemistot

Käyttäjällä on palvelimella omassa kotihakemistossaan seuraavat hakemistot:

- `/accepted` – Hakemisto, jonne säilytykseen hyväksytyjen siirtopakettien tarkastusraportit viedään.
- `/rejected` – Hakemisto, jonne siirrosta hylätyt paketit sekä niiden tarkastusraportit viedään.
- `/transfer` – Hakemisto, jonne käyttäjä vie siirtopaketit.
- `/disseminated` – Hakemisto, johon luvussa 3.4.2 muodostetut jakelupaketit syntyvät.

Tarkastusraportit säilyvät käyttäjän kotihakemistossa vähintään 20 päivää, sekä jakelupaketit ja hylätyt siirtopaketit vähintään kymmenen (10) päivää, ellei käyttäjä itse poista niitä aiemmin. Jakelupaketeista kerrotaan tarkemmin luvussa 3.5.

2.3 Siirtäminen PAS-palveluihin

Käyttäjä siirtää paketoitumääritysten mukaisesti luodun uuden aineiston SFTP-protokollan avulla oman kotihakemistonsa `/transfer`-hakemistoon.

Kun siirtopaketti otetaan käsittelyyn, se automaattisesti siirretään `/transfer`-hakemistosta tarkastukseen. Aineisto siirtyy säilytykseen, kun se on hyväksytty, ja käyttäjä saa tarkastusraportin siirrosta automaattisesti `/accepted`-hakemistoon. Jos aineisto hylätään, koko aineisto ja tarkastusraportti siirretään `/rejected`-hakemistoon. Luvussa 2.4 käsitellään tarkemmin hylätyyn aineiston käsittelyä. Tarkastusraportin rakenne on kuvattu liitteessä A ja se sisältää mm. seuraavat asiat:

- Siirtopaketin siirtäjä (käyttäjätunnus)
- Aikaleima, jolloin paketti otettiin tarkastuskäsittelyyn
- Siirtopaketin rakenne
- Tarkastustoimenpiteet, niiden tulokset, kohteet ja aikaleimat
- Aikaleima säilytyksen aloittamisesta ja säilytysvastuun siirtymisestä

Käyttäjän saataville ilmestyy raporttiin liittyen kaksi tiedostoa: PREMIS-muotoinen tarkka tarkastusraportti ja HTML-muotoinen selaimella luettava tiivistelmä. Näiden raporttiedostojen hakemistorakenne on hyväksytyssä tapauksessa muotoa:

```
/accepted/<date>/<transfer>/<transfer>-<uid>-ingest-report.xml
```

```
/accepted/<date>/<transfer>/<transfer>-<uid>-ingest-report.html
```

Näissä tiedostopolun osat määritellään seuraavasti:

- `<date>` – Päivämäärä (vuosi-kuukausi-päivä), jolloin raportti siirrettiin käyttäjän saataville.
- `<transfer>` – Siirtopaketin ZIP- tai TAR-tiedoston nimi.
- `<uuid>` – Eri siirrot yksilöivä tunniste.
- `xml` – PREMIS-muotoisen raportin tiedostopäätte.
- `html` – Tiivistelmäraportin tiedostopäätte.

Hylätyn siirtopaketin tapausta käsitellään tarkemmin luvussa 2.4. Hakemistojako päivämäärän mukaan on olemassa sen vuoksi, että raportteja ei tiedostolistauksen kannalta kertyisi kohtuutonta määrää samaan hakemistoon.

2.4 Vastaanotossa hylätyn aineiston käsittely

Vastaanotossa hylätty siirtopaketti palautuu käyttäjälle `/rejected`-hakemistoon. Siirtopaketin hakemistorakenteen muoto on:

```
/rejected/<date>/<transfer>/<transfer>-<uuid>/
```

ja vastaavat tarkastusraportit ovat tällöin:

```
/rejected/<date>/<transfer>/<transfer>-<uuid>-ingest-report.xml
```

```
/rejected/<date>/<transfer>/<transfer>-<uuid>-ingest-report.html
```

Näissä määreet määritellään seuraavasti:

- `<date>` – Päivämäärä (vuosi-kuukausi-päivä), jolloin raportti siirrettiin käyttäjän saataville.
- `<transfer>` – Siirtopaketin ZIP- tai TAR-tiedoston nimi.
- `<uuid>` – Eri siirrot yksilöivä tunniste.
- `xml` – PREMIS-muotoisen raportin tiedostopäätte.
- `html` – Tiivistelmäraportin tiedostopäätte.

Aineiston voi korjata `/rejected`-hakemistossa ehjäksi niin, että virheettömiä tiedostoja ei tarvitse toimittaa SFTP-rajapinnan yli palvelimelle uudelleen. Virheelliset tiedostot voidaan poistaa ja korjatut toimittaa tilalle. Kun paketti on käyttäjän mielestä korjattu ja uudelleen allekirjoitettu, tämä voi siirtää sen `/rejected`-hakemistosta suoraan `/transfer`-hakemistoon (ks. SFTP:n `rename`-komento), jolloin siirtoprosessi käynnistyy muokatulle materiaalille luvussa 2.3 esitetyllä tavalla.

3 AINEISTOJEN HAKU JA HALLINTA

Aineistojen hakemiseen, hallintaan ja noutamiseen käytetään REST-rajapintaa HTTP-protokollan yli. Tässä luvussa esitellyissä REST-viesteissä käytetään seuraavia määreitä:

- `<base>` – Merkkijono: `https://pas.csc.fi/api/2.0`
- `<contract>` – Sopimustunniste, joka rajaa käsittelyn ainoastaan tietyn sopimuksen alaiseen aineistoon.
- Eri komennoissa on olemassa myös muita määreitä, jotka on erikseen esitelty ko. viestin määrittelyssä yhteydessä.

3.1 Autentikointi ja valvonta

Autentikoinnissa käytetään TLS/SSL-suojauksia ja HTTP Basic Access -autentikointia. Autentikoitunut käyttäjä voi käyttää REST-viesteissä ainoastaan hänelle valtuutettuja sopimustunnisteita. PAS-palvelut tallentavat käyttäjän toimenpiteistä lokia niin, että jokaisesta REST-kyselyviestistä tallentuu lokiin vähintään seuraavat asiat automaattisesti:

- Kuka teki? (käyttäjätunnus, jolla viesti lähetettiin)
- Mitä teki? (REST-kyselyviestin sisältö)
- Milloin teki? (kyselyn aikaleima)
- Missä teki? (IP-osoite, josta viesti lähetettiin)
- Mitä vastattiin? (HTTP-statuskoodi ja vastauksen koko tavuina)

3.2 Viestien rakenne

3.2.1 Kyselyviestit

Kyselyviestit ovat muotoa:

```
{GET, POST} <base>/<contract>/<term>...
```

Näistä `<term>` kuvaa olemassa olevaa resurssia tai työkalua, ja se on jokin seuraavista:

- `search` – Hakutyökalu. Toiminnot rajapinnassa on kuvattu luvussa 3.3.
- `preserved` – Säilytyspakettien resurssi. Toiminnot rajapinnassa on kuvattu luvussa 3.4.
- `disseminated` – Jakelupakettien resurssi. Toiminnot rajapinnassa on kuvattu luvussa 3.5.

Myöhemmin rinnalle voidaan lisätä myös muita termejä, esimerkiksi `statistics` (tilastot).

Kyselyviestien muoto on tässä määrittelyssä esitetty ilman vapaaehtoisia HTTP-otsakekenttiä, joita voi lisätä mihin tahansa viestiin RFC 7231 -määrittelyksen mukaisesti. Esimerkiksi viestiin:

```
GET https://pas.csc.fi/api/2.0/csc
```

voidaan sisällyttää kenttiä seuraavasti:

```
GET https://pas.csc.fi/api/2.0/csc HTTP/1.1
Date: Mon, 13 Nov 2017 08:12:31 GMT
Host: pas.csc.fi
...
```

3.2.2 Palautusviestit

Palautusviestien rungossa käytetään JSON-muotoa [RFC_7159] käyttäen JSend-rakennetta [JSEND]. Onnistuneessa tilanteessa palautusviestiksi annetaan seuraava rakenne:

```
{
  "status": "success",
  "data": { ... }
}
```

Palautusviesteissä `status` kertoo kyselyn onnistumisesta tai epäonnistumisesta, ja `data` tulosjoukon. Epäonnistuneessa tilanteessa vastausviestiksi annetaan seuraava rakenne:

```
{
  "status": "fail",
  "data": { "message": "..."}
}
```

Virhetilanteessa `message` kertoo virheilmoituksen. Jos kuitenkin kyselyn parametri on ollut virheellinen, palautetaan virheilmoitus parametrin nimellä:

```
{
  "status": "fail",
  "data": { "<parameter>": "..."}
}
```

Esimerkki: Jos kyselyssä on ollut parametri `limit`, ja sille on annettu vääränlainen arvo, voisi ilmoitus olla seuraava:

```
{
  "status": "fail",
  "data": { "limit": "Value can only be an integer in range 1-1000"}
}
```

Jos käyttäjä yrittää hallita sellaista säilytyksessä olevaa aineistoa, johon hänellä ei ole oikeutta päästä käsiksi, palautetaan RFC 7231 -määrittämissä esitetyn "403 Forbidden"-koodin sijasta koodi: "404 Not Found".

Palautettava säilytyksestä haettu aineisto on ZIP- tai TAR-jakelupaketti, tai XML-muotoinen metadatatiedosto.

3.3 Hakuviestit

PAS-järjestelmässä tietty aineisto tallennetaan yhteen tai useampaan säilytyspakettiin. Jokainen säilytyspaketti sisältää paketin sisällön kuvaavan METS-dokumentin, joka indeksoidaan erilliseen tietokantaan. Myös olemassa olevien jakelupakettien METS-dokumentit indeksoidaan. Hakutoiminnolla voi etsiä säilytys- ja jakelupaketteja tekemällä hakuja tähän tietokantaan, ja saada tuloksena luettelon niistä tietopaketeista, joissa annettu hakuehto toteutuu. Jos käyttäjä tietää tarvitsemiensa säilytys- tai jakelupakettien tunnisteet ennestään, hän voi käyttää tunnisteita suoraan, ja silloin tämän hakutoiminnon voi ohittaa. Säilytyksessä olevan aineiston päivittäminen (esim. migraatio tai Aineistojen ja niiden metatietojen paketointi pitkäaikais säilytykseen –määrittämissä [META] mukainen päivityspaketti) ei muuta jo olemassa olevia säilytyspaketteja, vaan päivitettyille versioille luodaan uudet säilytyspaketit. Näin hakutoiminnolla ja sen palauttamilla tunnisteilla on mahdollista erottaa myös aineistojen eri versiot toisistaan.

Viestin muoto:

```
GET <base>/<contract>/search?<parameters>
```

Viestin määreet:

- `<parameters>` – Nolla, yksi tai useampi alla esitetystä parametreista &-merkeillä erotettuna.

```
q=<query>
```

Hakuehto, jota käytetään tulosjoukon haussa. Hakuehtoa esitellään erikseen luvussa 3.3.1. Oletuksena haetaan kaikesta aineistosta. On huomattava, että muodostettu hakuehto tulee URL-koodata REST-rajapintaa varten.

```
limit=<limit>
```

Parametri mahdollistaa tuloksien sivutuksen `page`-parametrin kanssa, jolloin `<limit>` kertoo, montako alkia (enintään) palautetaan sivua kohti, ja `page` kertoo sivunumeron. Oletuksena käytetään arvoa 20. Kokonaisluvut väliltä 1-1000 ovat sallittuja.

page=<page>

Parametri mahdollistaa tuloksien sivutuksen `limit`-parametrin kanssa, jolloin `limit` kertoo, montako alkia (enintään) palautetaan sivua kohti, ja `<page>` kertoo sivunumeron. Oletuksena käytetään arvoa 1, joka vastaa ensimmäistä sivua.

Palautusviesti normaalitilanteissa:

- HTTP 200 OK – Onnistunut viesti. Palautuksena annetaan luettelo löydetyistä tietopaketeista JSON-muodossa. Saaduilla URL-osoitteilla pääsee hallinnoimaan näitä paketteja.
- HTTP 400 Bad Request – Komennossa oli syntaksivirhe, esimerkiksi virheellinen päiväys parametrissa.
- HTTP 404 Not Found – Haettua resurssia ei löytynyt (esim. sopimustunniste väärin).
- HTTP 405 Method Not Allowed – Viestissä väärä metodi (muu kuin GET).

Onnistuneen kyselyn palautusviestin muoto on seuraava:

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2345
Allow: GET
Date: Mon, 13 Nov 2017 10:42:37 GMT
```

```
{
  "status": "success",
  "data": {
    "results": [
      {
        "location": "<base>/<contract>/preserved/<aip_id-1>",
        "createdate": "<aip_create_date-1>",
        "lastmoddate": "<dip_lastmoddate-1>",
        "match": <search_match-obj-1>,
        "id": "<aip_id-1>",
        "pkg_type": "AIP"
      },
      {
        "location": "<base>/<contract>/disseminated/<dip_id-1>",
        "createdate": "<dip_create_date-1>",
        "match": <search_match-obj-2>,
        "id": "<dip_id-1>",
        "pkg_type": "DIP"
      },
      ...
    ],
    "links": {
      "self": "<current-url>",
      "next": "<next-url>",
      "previous": "<previous-url>"
    }
  }
}
```

Hakutoiminnon palautusviestien rakenteeseen on etsitty vaikutteita mm. JSON API –rajapinnan [JSON_API] sekä Invenio-repositoriosovelluksen [INVENIO] rajapinnan rakenteesta. Vastauksessa `results` kertoo tulosjoukon, ja `links` on sivutustoiminto. Tulosjoukossa on nolla, yksi tai useampi säilytys- tai jakelupaketin tulosalkio. Tulosalkio sisältää seuraavat tiedot:

- `location` – Osoite, jolla voi hallinnoida säilytys- tai jakelupakettia. Ks. hallinnointi luvuista 3.4 ja 3.5.
- `createdate` – Tietopaketin sisältämän METS-dokumentin CREATEDATE-attribuutin arvo.
- `lastmoddate` – Tietopaketin sisältämän METS-dokumentin LASTMODDATE-attribuutin arvo. Annetaan, jos se on olemassa.
- `match` – Hakuehdon osuma, joka antaa tietoa, missä kohtaa METS-tietokantaa hakuehto toteutui.

- `id` – Tietopakettin tunniste.
- `pkg_type` – Termi AIP tarkoittaa säilytyspakettia ja DIP jakelupakettia.

Sivutustoiminto sisältää seuraavat tiedot:

- `self` – Nykyinen osoite.
- `next` – Seuraava sivu. Annetaan, jos seuraava sivu on olemassa.
- `previous` – Edellinen sivu. Annetaan, jos edellinen sivu on olemassa.

Tulosalkiossa tietojen `createdate` ja `lastmoddate` tarkoituksena on kuvata aineistoversio, johon tietopaketti kuuluu. Tulosalkioon voidaan myöhemmin tarvittaessa lisätä muitakin tietoja.

3.3.1 Hakutoiminto

Säilytyspakettien ja olemassa olevien jakelupakettien METS-dokumentit on kokonaan indeksoitu METS-tietokantaan, ja niiden tietoja voidaan hakea tässä luvussa esitetyllä tavalla. Hakuehdossa käytetään hyväksi Apache Lucene-syntaksia [LUCENE], joka mahdollistaa kirjon erilaisia toimintoja. Hakuehdossa käytetään aina avain:arvo-pareja, jossa avain vastaa METS-dokumentin tiettyä elementtiä tai attribuuttia ja arvo tämän avaimen sisältämää arvoa.

Hakutietokanta ei ole reaaliaikainen. Säilytys- ja jakelupakettien pitää ensin indeksoitua, jotta niitä voi hakutoiminnolla etsiä. Jos säilytyspaketin säilytysvastuu on vasta äskettäin siirtynyt, tai jakelupaketti on vasta äskettäin valmistunut, se ei välttämättä ole vielä indeksoitunut tietokantaan, eikä pakettia tällöin hakutoiminnolla löydy. Samoin, kun jakelupaketti poistetaan, sen tiedot siivotaan hakutietokannasta viiveellä. Tällä viiveellä ei kuitenkaan ole vaikutusta muihin rajapinnan toimintoihin.

3.3.1.1 Hakutoiminnon ominaisuudet ja syntaksi

Toiminnot ovat hakutoiminnossa käytettävissä Apache Lucene -syntaksilla [LUCENE] luqum-sovelluskirjaston [LUQUM] mukaisesti, ja se sisältää mm. seuraavia toiminnallisuuksia¹:

- loogiset operaattorit AND, OR ja NOT, joita voidaan myös sulkumerkein ryhmitellä
- sumea haku (fuzzy search)
- läheisyshaku (proximity search)
- jokerimerkit ? ja * (wildcard)
- haku tietyltä alueelta (range search)
- termin merkityksellisyyden korostaminen (boosting a term)

3.3.1.2 Avain/arvo-parit

Hakuehdossa avaimen nimi vastaa METS-dokumentin tiettyä elementtiä tai attribuuttia. Arvo taas jotakin, mitä tämän elementin tai attribuutin tulee sisältää. Avaimena voi käyttää myös XML-polkua, ja niissä elementtien/attribuuttien nimet kuvataan alaviivalla erotettuina ilman nimiavaruuksia. Koko polkua ei kuitenkaan ole pakko antaa, vaan pelkkä loppuosa riittää. Arvoja käsitellään merkkikokoriippumattomasti. Esimerkkejä:

- `q='mets_OBJID:id-123'` – Hae tiedot paketeista, joiden METS-dokumentissa on sellainen `<mets>`-elementti, jonka `@OBJID`-attribuutti sisältää arvon `id-123`. Esimerkissä avaimena toimii `mets_OBJID`.
- `q='mets_dmdSec_mdWrap_xmlData_subject:xxx'` – Hae tiedot paketeista, joiden METS-dokumentissa on sellainen `<mets>/<dmdSec>/<mdWrap>/<xmlData>/<subject>`-polku, jossa `<subject>`-elementti sisältää arvon `xxx`. Nimiavaruus saa XML-rakenteessa vaihtua kesken polun. Käytännössä tässä haetaan kuvailevan metatiedon lohkoa `<subject>`-juurielementtiä. Tuetuissa kuvailevan metatiedon muodoissa tämä on mahdollista vain Dublin Core -muodossa.

¹ Katso ohjeet ja esimerkit hakuehdon syntaksin käyttämiseksi osoitteesta:
https://lucene.apache.org/core/3_6_0/queryparsersyntax.html

- `q='subject:xxx'` – Hae tiedot paketeista, joiden METS-dokumentissa mikä tahansa `<subject>`-elementti tai `@subject-attribuutti` sisältää arvon `xxx`. Tuetuista kuvailevan metatiedon muodoista tämä huomioi Dublin Coren lisäksi myös esimerkiksi EAD- ja MODS-muotojen `<subject>`-elementit. Koska täyttä polkua ei anneta, saa haettava avain olla METS-dokumentissa missä tahansa.

3.3.1.3 Tietopaketin tyyppin rajaaminen

Hakuehtoon tulee lisätä AND-operaattorilla `pkg_type:AIP` tai `pkg_type:DIP` sen mukaan, halutaanko haun kohdistuvan vain säilytyspaketteihin (AIP) vai jakelupaketteihin (DIP). Sulkumerkkisäännöt on huomioitava. Oletuksena haetaan molemmista resursseista.

3.4 Säilytyspakettien hallinta

Toistaiseksi rajapinnan avulla voidaan hallita vain yhtä säilytyspakettia kerrallaan ja kokonaisuutena. Mikäli tulevaisuudessa nähdään riittävää tarvetta toiminnoille, jotka vaativat useamman säilytyspaketin tai niiden osien hallintaa, voidaan tämän rinnalle lisätä tarvittavat toiminnot.

3.4.1 Säilytyspakettiin kohdistuvat komennot

Tällä toiminnolla saa luettelon niistä komennoista, joita voidaan kohdistaa valitulle säilytyspaketille. Tällä hetkellä ainoa komento on jakelupaketin muodostus.

Viestin muoto:

```
GET <base>/<contract>/preserved/<aip-id>
```

Viestin määreet:

- `<aip-id>` – Säilytyspaketin tunniste, jonka perusteella kysely rajataan koskemaan säilytyksessä olevaa aineistoa. Tunnisteita voi hakea luvun 3.3 mukaisella toiminnolla.

Palautusviesti normaalitilanteissa:

- HTTP 200 OK – Onnistunut viesti. Palautuksena annetaan säilytyksessä olevalle aineistolle kohdistettavat komennot JSON-muodossa.
- HTTP 404 Not Found – Haettua aineistoa ei löytynyt.
- HTTP 405 Method Not Allowed – Viestissä väärä metodi (muu kuin GET).

Onnistuneen kyselyn palautusviesti on seuraava:

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2345
Allow: GET
Date: Mon, 13 Nov 2017 10:42:37 GMT
```

```
{
  "status": "success",
  "data": {
    "disseminate": "<base>/<contract>/preserved/<aip-id>/disseminate"
  }
}
```

Palautettu URL-muoto on esitelty luvussa 3.4.2.

3.4.2 Säilytyspaketin muodostus jakelupaketiksi

Tällä toiminnolla voi muodostaa jakelupaketin valitusta säilytyspaketista. Jakelupaketeista on kerrottu luvussa 3.5.

Viestin muoto:

```
POST <base>/<contract>/preserved/<aip-id>/disseminate?<parameters>
```

Viestin määreet:

- `<aip-id>` – Säilytyspaketin tunniste, jonka perusteella säilytyksessä olevasta aineistosta muodostetaan uusi jakelupaketti. Tunnisteita voi hakea luvun 3.3 mukaisella toiminnolla.
- `<parameters>` – Nolla, yksi tai useampi alla esitetyistä parametreista &-merkeillä erotettuna.

`catalog=<catalog>`

Jakelupaketti muodostetaan skeemakatalogin version `<catalog>` mukaisesti. Parametrina annetaan versionumerosta vain kaksi ensimmäistä numeroa. Esim. `catalog=1.6` palauttaa tulokset skeemakatalogin 1.6.X mukaisesti, missä X on uusin olemassa oleva 1.6-versiosarjan numero. Jakelupakettia ei voi muodostaa säilytyspakettia vanhempaan katalogiversioon. Oletuksena käytetään uusinta versiota.

`format=<format>`

Jakelupaketti käännetään tiedostomuotoon `<format>`. Sallitut arvot ovat `zip` ja `tar`. Oletuksena käytetään ZIP-pakkausta.

Palautusviesti normaalitilanteissa:

- HTTP 202 Accepted – Onnistunut viesti. Aineisto löytyi ja jakelupaketin muodostus alkoi, mutta muodostus vielä kestää. Palautusviestin yhteydessä annetaan viittaus jakelupakettiin. Luotu jakelupaketti muodostuu REST-rajapinnan lisäksi saataville SFTP-yhteyden kautta käyttäjän `/disseminated`-hakemistoon (ks. luku 2.2).
- HTTP 400 Bad Request – Määrityksen versio ei kelpaa.
- HTTP 404 Not Found – Haettua aineistoa ei löytynyt.
- HTTP 405 Method Not Allowed – Viestissä väärä metodi (muu kuin POST).

Onnistuneen kyselyn palautusviestin muoto on seuraava:

```
HTTP/1.1 202 Accepted
Content-Type: application/json
Location: <base>/<contract>/disseminated/<dip-id>
Allow: POST
Date: Mon, 13 Nov 2017 10:42:37 GMT
```

```
{
  "status": "success",
  "data": {
    "disseminated": "<base>/<contract>/disseminated/<dip-id>"
  }
}
```

Palautusviestissä `<dip-id>` on jakelupaketin tunniste. Palautettu URL-muoto on esitelty luvussa 0.

3.5 Jakelupaketit

Kun PAS-järjestelmästä halutaan noutaa aineistoa, pitää halutusta säilytyksessä olevasta aineistosta ensin muodostaa jakelupaketti (ks. luku 3.4.2), jonka voi noutaa. Jakelupaketti muodostetaan Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen -määrityksen [META] mukaiseksi. Paketille annetaan uusi tunniste, ja sille kirjoitetaan uusi METS-dokumentti. Lopuksi vielä paketin eheys varmistetaan, ja sen jälkeen se asetetaan saataville. Noudon voi tehdä joko REST- tai SFTP-rajapintaa käyttäen. Aineiston nouto ei poista aineistoa säilytyksestä.

3.5.1 Jakelupaketin seuranta ja niihin kohdistuvat komennot

Tällä toiminnolla voi seurata jakelupaketin muodostumista. Samalla toiminnolla saa myös luettelon niistä komennoista, jotka voidaan kohdistaa jakelupaketille.

Viestin muoto:

```
GET <base>/<contract>/disseminated/<dip-id>
```

Viestin määreet:

- <dip-id> – Jakelupaketin tunniste, jonka käyttäjä sai jakelupaketin muodostusviestissä luvun 3.4.2 kyselyssä. Muodostettuja jakelupaketteja voi hakea luvun 3.3 mukaisesti.

Palautusviesti normaalitilanteissa:

- HTTP 200 OK – Onnistunut viesti. Palautuksena annetaan jakelupaketille kohdistettavat komennot JSON-muodossa.
- HTTP 404 Not Found – Haettua aineistoa ei löytynyt.
- HTTP 405 Method Not Allowed – Viestissä väärä metodi (muu kuin GET).

Onnistuneen kyselyn palautusviesti on seuraava:

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2345
Allow: GET
Date: Mon, 13 Nov 2017 10:42:37 GMT
```

```
{
  "status": "success",
  "data": {
    "complete": <complete>,
    "actions": <actions>
  }
}
```

Palautusviestissä <complete> on jokin seuraavista: "false" (jakelupaketin muodostuminen on käynnissä) tai "true" (jakelupaketti on valmis). Viestissä <actions> on JSON-rakenne toiminnoista, joita jakelupaketeille voi tehdä. Suppeimmillaan, kun jakelupaketin muodostus on vielä käynnissä, rakenne on tyhjä {}. Laajimmillaan, kun jakelupaketti on valmis, rakenne on seuraava:

```
{
  "download": "<base>/<contract>/disseminated/<dip-id>/download",
  "metadata": "<base>/<contract>/disseminated/<dip-id>/metadata",
  "history": "<base>/<contract>/disseminated/<dip-id>/history"
}
```

Ilmoitetut toiminnot ovat käytettävissä heti. Palautetut URL-muodot on esitelty luvussa 3.5.2.

3.5.2 Muodostettujen jakelupakettien noutaminen

Tällä toiminnolla voidaan noutaa muodostettuja jakelupaketteja tai niiden metatietoja. Jakelupaketit voidaan noutaa tässä esitetyn tavan lisäksi myös SFTP-rajapinnan kautta käyttäjän kotihakemistossa olevasta /disseminated-hakemistosta. Muodostetut jakelupaketit säilyvät kymmenen (10) päivää, ellei käyttäjä itse poista niitä aiemmin kotihakemistostaan.

Viestin muoto:

```
GET <base>/<contract>/disseminated/<dip-id>/download
GET <base>/<contract>/disseminated/<dip-id>/metadata
GET <base>/<contract>/disseminated/<dip-id>/history
```

Viestin määreet:

- `<dip-id>` – Jakelupaketin tunniste, jonka käyttäjä sai jakelupaketin muodostusviestissä luvun 3.4.2 kyselyssä. Muodostettuja jakelupaketteja voi hakea luvun 3.3 mukaisesti.
- `download` – Koko jakelupaketti noudetaan. Jakelupaketti on lisäksi noudettavissa SFTP-yhteyden kautta käyttäjän `/disseminated`-hakemistosta (ks. luku 2.2).
- `metadata` – Vain METS-dokumentti noudetaan.
- `history` – Paketin koko tapahtumahistoria noudetaan PREMIS-muodossa.

Palautusviesti normaalitilanteissa:

- HTTP 200 OK – Onnistunut viesti. Jakelupaketti, sen METS-dokumentti tai sen historiaraportti annetaan palautusviestin mukana. Jakelupaketti toimitetaan ZIP- tai TAR-pakettina ja sen metatiedot XML-dokumenttina. JSON-muotoista viestiä ei palauteta.
- HTTP 404 Not Found – Jakelupakettia ei löytynyt (jakelupakettia ei ole luotu tai sen luonti on kesken).
- HTTP 405 Method Not Allowed – Viestissä väärä metodi (muu kuin GET).

Onnistuneen kyselyn palautusviestin muoto on seuraava:

- Jakelupaketti ZIP-muodossa:

```
HTTP/1.1 200 OK
Content-Type: application/zip
Content-Length: 23456
Allow: GET
Date: Mon, 13 Nov 2017 10:42:37 GMT
```

```
<jakelupaketti.zip>
```

- Jakelupaketti TAR-muodossa:

```
HTTP/1.1 200 OK
Content-Type: application/x-tar
Content-Length: 23456
Allow: GET
Date: Mon, 13 Nov 2017 10:42:37 GMT
```

```
<jakelupaketti.tar>
```

- Jakelupaketin METS-dokumentti:

```
HTTP/1.1 200 OK
Content-Type: text/xml
Content-Length: 2345
Allow: GET
Date: Mon, 13 Nov 2017 10:42:37 GMT
```

```
<mets...
```

- Jakelupaketin säilytyshistoria:

```
HTTP/1.1 200 OK
Content-Type: text/xml
Content-Length: 2345
Allow: GET
Date: Mon, 13 Nov 2017 10:42:37 GMT
```

```
<premis...
```


3.6 Julkiset avaimet

Toiminnolla voi noutaa jakelupaketin digitaalisessa allekirjoituksessa käytetyn avainparin julkisen osan. Avainta voi käyttää sen verifiointiin, että jakelupaketti on tullut PAS-palvelusta, ja että se on eheä.

Viestin muoto:

```
GET <base>/public_key/dip
```

Palautusviesti normaalitilanteissa:

- HTTP 200 OK – Onnistunut viesti. Jakelupaketin digitaalisen allekirjoituksen julkinen osa annetaan palautusviestin mukana.
- HTTP 405 Method Not Allowed – Viestissä väärä metodi (muu kuin GET).

LIITE A. SIIRTOPAKETIN TARKASTUSRAPORTTI

Tarkastusraportti sisältää joukon PREMIS XML -muodon mukaisia objekteja, tapahtumia ja agentteja (tapahtuman suorittajia). Tässä liitteessä jokainen tällainen objekti, tapahtuma ja agentti on tiivistetysti avattu. Lueteltuja objekteja, tapahtumia ja agentteja on tarkastusraportissa kutakin yksi, ellei taulukon kohdalla ole muuta mainittu. Tapahtuman nimi ja lopputulos sekä tapahtumaan liittyvien objektien tiedostonimet näkyvät erillisessä HTML-muotoisessa ihmisluettavassa raportissa.

A.1. Objektit

Vastaanotossa käytetyistä ja luoduista kokonaisuuksista muodostetaan PREMIS XML -muotoiset esitykset, joita kutsutaan PREMIS-objekteiksi. Eri objekteihin kohdistuu vastaanotossa erilaisia tapahtumia (ks. luku A.2), joita suorittavat agentit (ks. luku A.3). Näitä kokonaisuuksia ovat siirtopaketti, digitaalinen allekirjoitus, METS-dokumentti, digitaalinen objekti sekä luotu säilytyspaketti. Tässä kappaleessa on kuvattu näitä kutakin kokonaisuutta esittelevien PREMIS-objektien sisällöt.

Siirtopaketti:

Elementti	Määritys
<object>	Siirtopaketti tarkastusraportissa
@xsi:type	"representation" (PREMIS-objektin tyyppi)
-<objectIdentifier>	Siirtopaketin tunniste tarkastusraportissa
--<objectIdentifierType>	"preservation-sip-id" (tunnisteen tyyppi)
--<objectIdentifierValue>	Tunnisteen arvo
-<originalName>	ZIP- tai TAR-tiedoston nimi, jossa siirtopaketti alkujaan sijaisi
-<environment>	
--<dependency>	Riippuvuussuhde siirtopakettiin
---<dependencyIdentifier>	Siirtopaketin tunniste METS-dokumentissa
----<dependencyIdentifierType>	"mets:OBJID"
----<dependencyIdentifierValue>	Siirtopaketin tunnisteen arvo

Siirtopaketin digitaalinen allekirjoitus:

Elementti	Määritys
<object>	Siirtopaketin digitaalinen allekirjoitustiedosto tarkastusraportissa
@xsi:type	"representation" (PREMIS-objektin tyyppi)
-<objectIdentifier>	Digitaalisen allekirjoitustiedoston tunniste tarkastusraportissa
--<objectIdentifierType>	"preservation-signature-id" (tunnisteen tyyppi)
--<objectIdentifierValue>	Tunnisteen arvo
-<originalName>	Digitaalisen allekirjoituksen tiedostonimi
-<relationship>	Suhde siirtopakettiin
--<relationshipType>	"structural" (rakenteellinen suhde)
--<relationshipSubType>	"is included in" (sisältyy)
--<relatedObjectIdentification>	Siirtopaketin tunniste tarkastusraportissa
---<relatedObjectIdentifierType>	"preservation-sip-id" (siirtopaketin tunnisteen tyyppi)
---<relatedObjectIdentifierValue>	Siirtopaketin tunnisteen arvo

METS-dokumentti:

Elementti	Määritys
<object>	METS-dokumentti tarkastusraportissa
@xsi:type	"representation" (PREMIS-objektin tyyppi)
-<objectIdentifier>	METS-dokumentin tunniste tarkastusraportissa
--<objectIdentifierType>	"preservation-mets-id" (tunnisteen tyyppi)

--<objectIdentifierValue>	Tunnisteen arvo
-<originalName>	METS-dokumentin tiedostonimi
-<relationship>	Suhde siirtopakettiin
--<relationshipType>	"structural" (rakenteellinen suhde)
--<relationshipSubType>	"is included in" (sisältyy)
--<relatedObjectIdentification>	Siirtopaketin tunniste tarkastusraportissa
---<relatedObjectIdentifierType>	"preservation-sip-id" (siirtopaketin tunnisteen tyyppi)
---<relatedObjectIdentifierValue>	Siirtopaketin tunnisteen arvo

Digitaalinen objekti: Jokaiselle digitaaliselle objektille tulee raporttiin oma PREMIS-objektinsa.

Elementti	Määrittäminen
<object>	Digitaalinen objekti tarkastusraportissa
@xsi:type	"representation" (PREMIS-objektin tyyppi)
-<objectIdentifier>	Digitaalisen objektin tunniste tarkastusraportissa
--<objectIdentifierType>	"preservation-object-id" (tunnisteen tyyppi)
--<objectIdentifierValue>	Tunnisteen arvo
-<originalName>	Digitaalisen objektin tiedostonimi
-<environment>	
--<dependency>	Riippuvuussuhde digitaaliseen objektiin
---<dependencyIdentifier>	Digitaalisen objektin tunniste METS-dokumentissa
----<dependencyIdentifierType>	Digitaalisen objektin tunnisteen tyyppi
----<dependencyIdentifierValue>	Digitaalisen objektin tunnisteen arvo
-<relationship>	Suhde siirtopakettiin
--<relationshipType>	"structural" (rakenteellinen suhde)
--<relationshipSubType>	"is included in" (sisältyy)
--<relatedObjectIdentification>	Siirtopaketin tunniste tarkastusraportissa
---<relatedObjectIdentifierType>	"preservation-sip-id" (siirtopaketin tunnisteen tyyppi)
---<relatedObjectIdentifierValue>	Siirtopaketin tunnisteen arvo

Säilytyspaketti:

Elementti	Määrittäminen
<object>	Säilytyspaketti tarkastusraportissa
@xsi:type	"representation" (PREMIS-objektin tyyppi)
-<objectIdentifier>	Säilytyspaketin tunniste tarkastusraportissa
--<objectIdentifierType>	"preservation-aip-id" (tunnisteen tyyppi)
--<objectIdentifierValue>	Tunnisteen arvo
-<originalName>	Säilytyspaketin tiedostonimi
-<relationship>	Suhde siirtopakettiin
--<relationshipType>	"derivation" (johdannainen)
--<relationshipSubType>	"has source" (on olemassa lähde)
--<relatedObjectIdentification>	Siirtopaketin tunniste tarkastusraportissa
---<relatedObjectIdentifierType>	"preservation-sip-id" (siirtopaketin tunnisteen tyyppi)
---<relatedObjectIdentifierValue>	Siirtopaketin tunnisteen arvo

A.2. Tapahtumat

Vastaanotossa suoritetuista toimenpiteistä muodostetaan PREMIS XML -muotoiset esitykset, joita kutsutaan PREMIS-tapahtumiksi. Eri vastaanoton toimenpiteet kohdistuvat eri objekteihin (ks. luku A.1), ja niitä suorittavat agentit (ks. luku A.3). Näitä toimenpiteitä on useita, ja tässä kappaleessa on kuvattu kutakin toimenpidettä esittelevien PREMIS-tapahtumien sisällöt. Seuraavana on esitetty taulukkona

tapahtumarunko, joka on sama kaikille tapahtumille. Tämän jälkeen on luettelona esitetty vastaanoton tapahtumat ja tarkennuksia tapahtumarungon muutamille elementeille.

Tapahtumarunko:

Elementti	Määrittäminen
<event>	
--<eventIdentifier>	Tapahtuman tunnistus
--<eventIdentifierType>	Tunnisteen tyyppi
--<eventIdentifierValue>	Tunnisteen arvo
--<eventType>	Tapahtuman tyyppi, ks. tarkennus tapahtumista
--<eventDateTime>	Tapahtuman aikaleima
--<eventDetail>	Tapahtuman nimi, ks. tarkennus tapahtumista
--<eventOutcomeInformation>	
--<eventOutcome>	Tapahtuman lopputulos: success tai failure
--<eventOutcomeDetail>	
---<eventOutcomeDetailNote>	Tekstimuotoinen lisätieto tapahtuman lopputuloksesta. Käytössä tapahtumissa, jos ei toisin ole mainittu
---<eventOutcomeDetailExtension>	XML-muotoinen lisätieto tapahtuman lopputuloksesta. Käytössä tapahtumissa vain, jos erikseen mainittu
--<linkingAgentIdentifier>	Viittaus agenttiin tarkastusraportissa, ks. tarkennus tapahtumista
--<linkingAgentIdentifierType>	Agentin tunnisteen tyyppi
--<linkingAgentIdentifierValue>	Agentin tunnisteen arvo
--<linkingObjectIdentifier>	Viittaus objektiin tarkastusraportissa, ks. tarkennus tapahtumista
--<linkingObjectIdentifierType>	Objektin tunnisteen tyyppi
--<linkingObjectIdentifierValue>	Objektin tunnisteen arvo

Siirtopaketin siirto:

Elementti	Määrittäminen
--<eventType>	"transfer"
--<eventDetail>	"Transfer of submission information package"
--<linkingAgentIdentifier>	Viittaus aineiston siirtäneeseen käyttäjään
--<linkingObjectIdentifier>	Viittaus siirtopakettiin

Siirtopaketin purku:

Elementti	Määrittäminen
--<eventType>	"decompression"
--<eventDetail>	"Decompression of submission information package"
--<linkingAgentIdentifier>	Viittaus purkusovellukseen
--<linkingObjectIdentifier>	Viittaus siirtopakettiin

Virustarkastus:

Elementti	Määrittäminen
--<eventType>	"virus check"
--<eventDetail>	"Virus check of transferred files"
--<linkingAgentIdentifier>	Viittaus tarkastussovellukseen
--<linkingObjectIdentifier>	Viittaus siirtopakettiin

Digitaalisen allekirjoituksen tarkastus:

Elementti	Määrittäminen
--<eventType>	"digital signature validation"
--<eventDetail>	"Submission information package digital signature validation"

-<linkingAgentIdentifier>	Viittaus tarkastussovellukseen
-<linkingObjectIdentifier>	Viittaus digitaaliseen allekirjoitukseen

METS-skeemavalidointi:

Elementti	Määrittäminen
-<eventType>	"validation"
-<eventDetail>	"METS schema validation"
-<linkingAgentIdentifier>	Viittaus validointisovellukseen
-<linkingObjectIdentifier>	Viittaus METS-dokumenttiin

METS-lisävalidointi:

Elementti	Määrittäminen
-<eventType>	"validation"
-<eventDetail>	"Additional METS validation of required features"
---<eventOutcomeDetailNote>	Ei käytetä
---<eventOutcomeDetailExtension>	Käytetään
-<linkingAgentIdentifier>	Viittaus validointisovellukseen
-<linkingObjectIdentifier>	Viittaus METS-dokumenttiin

Digitaalisten objektien eheystarkastus: Tapahtuma sisältää yhdessä nipussa kaikkien siirtopaketissa olleiden digitaalisten objektien eheystarkastuksen.

Elementti	Määrittäminen
-<eventType>	"fixity check"
-<eventDetail>	"Fixity check of digital objects in submission information package"
-<linkingAgentIdentifier>	Viittaus tarkastussovellukseen
-<linkingObjectIdentifier>	Viittaus siirtopakettiin

Digitaalisten objektien validointi: Jokaiselle digitaaliselle objektille tulee yksi tai useampi tapahtuma.

Elementti	Määrittäminen
-<eventType>	"validation"
-<eventDetail>	"Digital object validation"
---<eventOutcomeDetailNote>	Esiintyy, jos elementtiä <eventOutcomeDetailExtension> ei käytetä.
---<eventOutcomeDetailExtension>	Esiintyy, jos elementtiä <eventOutcomeDetailNote> ei käytetä.
-<linkingAgentIdentifier>	Viittaus validointisovellukseen
-<linkingObjectIdentifier>	Viittaus digitaaliseen objektiin

Siirtopaketin validoinnin yhteenveto:

Elementti	Määrittäminen
-<eventType>	"validation"
-<eventDetail>	"Validation compilation of submission information package"
-<linkingAgentIdentifier>	Viittaus validointisovellukseen
-<linkingObjectIdentifier>	Viittaus siirtopakettiin

Säilytyspaketin luominen:

Elementti	Määrittäminen
-<eventType>	"creation"
-<eventDetail>	"Creation of archival information package"
-<linkingAgentIdentifier>	Viittaus luontisovellukseen
-<linkingObjectIdentifier>	Viittaus luotuun säilytyspakettiin

Säilytysvastuun siirtyminen:

Elementti	Määrittäminen
-<eventType>	“preservation responsibility change”
-<eventDetail>	“Preservation responsibility change to the digital preservation service”
-<linkingAgentIdentifier>	Viittaus siirtosovellukseen
-<linkingObjectIdentifier>	Viittaus säilytyspakettiin

A.3. Agentit

Vastaanotossa suoritettujen toimenpiteiden (ks. luku A.2) suorittajista muodostetaan PREMIS XML-muotoiset esitykset, joita kutsutaan PREMIS-agenteiksi. Suorittajana on joko käyttäjä tai ohjelma. Tässä kappaleessa on esitelty näitä suorittajia kuvaavien PREMIS-agenttien sisällöt.

Aineiston siirtänyt käyttäjä:

Elementti	Määrittäminen
<agent>	
-<agentIdentifier>	Agentin tunnistus
--<agentIdentifierType>	Tunnisteen tyyppi
--<agentIdentifierValue>	Tunnisteen arvo
-<agentName>	Aineiston siirtäjän käyttäjätunnus
-<agentType>	“organization” (agentin tyyppi)

PAS-ohjelmistokomponentti: Jokaiselle raportissa viitatuille ohjelmistokomponentille tulee oma agenttinsa.

Elementti	Määrittäminen
<agent>	
-<agentIdentifier>	Agentin tunnistus
--<agentIdentifierType>	Tunnisteen tyyppi
--<agentIdentifierValue>	Tunnisteen arvo
-<agentName>	Ohjelmakoodin nimi
-<agentType>	“software” (agentin tyyppi)

LÄHTEET

- [IETF_SFTP] Internet Engineering Task Force Internet-Draft: SSH File Transfer Protocol, J. Galbraith, O. Saarenmaa, July 2006. <http://tools.ietf.org/html/draft-ietf-secsh-filexfer-13>
- [INVENIO] Invenio. <http://invenio-software.org/>
- [JSEND] OmniTI Labs: JSend. <https://labs.omniti.com/labs/jsend>
- [JSON_API] JSON API. <http://jsonapi.org/>
- [LUCENE] Apache Lucene - Query Parser Syntax. https://lucene.apache.org/core/3_6_0/queryparsersyntax.html
- [LUQUM] Luqum's Documentation. <http://luqum.readthedocs.io/>
- [META] Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen. <http://www.kdk.fi/fi/pitkaaikaissailytys/145-kdkn-hallinnolliset-ja-rakenteelliset-metatiedot-ja-aineiston-paketointi>
- [PREMIS] Data Dictionary for Preservation Metadata: PREMIS version 2.2, July 2012. <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>
- [RFC_7159] Internet Engineering Task Force Request for Comments: The JavaScript Object Notation (JSON) Data Interchange Format. <http://tools.ietf.org/html/rfc7159>
- [RFC_7230] Internet Engineering Task Force Request for Comments: Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing. <http://tools.ietf.org/html/rfc7230>
- [RFC_7231] Internet Engineering Task Force Request for Comments: Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content. <http://tools.ietf.org/html/rfc7231>
- [TDSTO] Säilytys- ja siirtokelpoiset tiedostomuodot. <http://kdk.fi/fi/pitkaaikaissailytys/141-kdkn-saeilytys-ja-siirtokelpoiset-tiedostomuodot>