

KAVIn digitaalinen arkistointi

PAS REST-työpaja
Pekka Tähtinen, 12.6.2018



Kansallinen audiovisuaalinen instituutti

- Opetus- ja kulttuuriministeriön alainen valtion laitos
- Kansallinen audiovisuaalinen arkisto (KAVA vuosina 2008–2013, Suomen elokuva-arkisto vuosina 1957–2007)
- Mediakasvatus- ja kuvaohjelmakeskus (MEKU vuosina 2012–2013, Valtion elokuvatarkastamo vuosina 1946–2011)
- Yhdistyivät 1.1.2014 Kansalliseksi audiovisuaaliseksi instituutiksi.

Kansallinen audiovisuaalinen instituutti

- Lakisääteisiin tehtäviin kuuluvat:
 - elokuvien ja televisio- ja radio-ohjelmien säilyttäminen
 - niihin liittyvä tutkimus
 - kuvaohjelmien tarjoamisen valvonta
 - mediakasvatuksen edistäminen

KAVIn kokoelmat

- **Kotimainen elokuvakokoelma**

- 1 454 pitkää näytelmäelokuva
- 33 690 lyhytelokuva
- vuodesta 1984 talletukset lakisääteisiä

- **Ulkomainen elokuvakokoelma**

- 16 430 pitkää näytelmäelokuva
- 24 918 lyhytelokuva
- talletukset vapaaehtoisia

- **Videokokoelma**

- 117 397 tallennetta eri formaateissa
- ammattimaista tuotantoa
- Suomessa levitettyjen tallenteiden vapaakappaleet



KAVIn kokoelmat



- **Perinnekkokoelma**

- painopiste kotimaisessa elokuvahistoriassa: vuosien 1896–2000 *teknisiä laitteita, rekvisiittaa, pukuja jne.*
- kokoelmat karttavat lahjoituksin ja ostoin
- 18 000 esinettä, joista 50 % luetteloitu ja valokuvattu

- **Oheisaineistot**

- *Valokuvat:* n. 450 000 kotimaista, digitoitu n. 176 000 kpl;
yli 500 000 ulkomaista, digitoitu n. 64 000 kpl
- *Julisteet:* n. 4 600 kotimaista, jotka lähes kaikki digitoitu,
yli 35 000 ulkomaista
- *Käsikirjoitukset:* 6 450 kotimaista, 22 400 ulkomaista

KAVIn “digivarasto”

- Digivarasto käyttöönotettu 2010-2011 CSC:llä
- Suunniteltu elokuvien digitaaliseen arkistointiin: omat digitoinnit ja digitaaliset talletukset.
- 4.6.2018 tilanne: arkistoituna yhteensä 2364 Tt.

	<i>Digitaalisia materiaaleja</i>	<i>Nimikkeitä</i>	<i>Datamäärä</i>
Elokuvat	8 504	3 522	2255 Tt
Radio- ja tv-ohjelmat	15 531	14 364	168 Tt
Digitoidut videot	22 735	21 419	98 Tt
(RTVA ohjelmavirta	2 500 000	x	530 Tt)

Apple ProRes

- Elokuvien käyttökopioita,
 - Tv-mastereita,
- Digitoidut mastervideot

DCP

- Elokuvien esityskopiot

DPX

- Skannattu filmi,
- Digitoitujen elokuvien masterit

MPEG 2 part 2 / H.262

- Digitoidut VHS:t,
- Digitoidut mastervideot,
 - Tv-mastereita

MPEG4 Part 10 / H.264 / MPEG4 AVC

- RTVA ohjelmavirta,
 - Tv-masterita,
- Videotalletuksia

XDCAM HD

- Tv-mastereita

DPX

Digitoitujen elokuvien masterit

- DPX-kuvajono SMPTE-standardi vuodesta 1994
- Standardi skannatulle filmille, käytetään laajasti elokuvien jälkituotannossa
- Pakkaamaton, logaritminen
- PAS-säilytyskelpoinen formaatti

DPX

Digitoitujen elokuvien masterit

90 min = 24 ruutua / sekunti x 60 sekuntia x 90 minuuttia = **130 000 ruutua**

4:3 overscan 4K filmiruutu: 4300x3324 pikseliä, RGB 10 bit ~**55 Mt per ruutu**

“Master archive package”: 55 Mt x 130 000 ~ **7,1 Tt**

DPX

Digitoitujen elokuvien masterit

- METS-paketissa jokainen DPX-filmiruutu käsitellään omana objektinaan
 - > esim. *16 min* lyhytelokuvassa *23 166 kpl* digitaalista objektia
 - > yksi elokuva jakautuu useampaan METS-pakettiin (max 1.5 Tt / paketti)
- KAVIn toteutuksessa METS-pakettiin vietävä metadata minimissä
- Pakettien muodostus automatisoitu kokonaan
- Hyödynnetään PAS:in tarjoamaa paketoitikonponenttia (siptools)
- Paketointi, tarkistussummaus, xml:n muodostus ja lähetys verkon yli integroitu yhteen operaatioon.

**R-123_Elokuva/
/image/orig_R-123_Elokuva/
/R-123_Elokuva_osa-1/*.dpx
/R-123_Elokuva_osa-2/*.dpx
/R-123_Elokuva_osa-3/*.dpx**

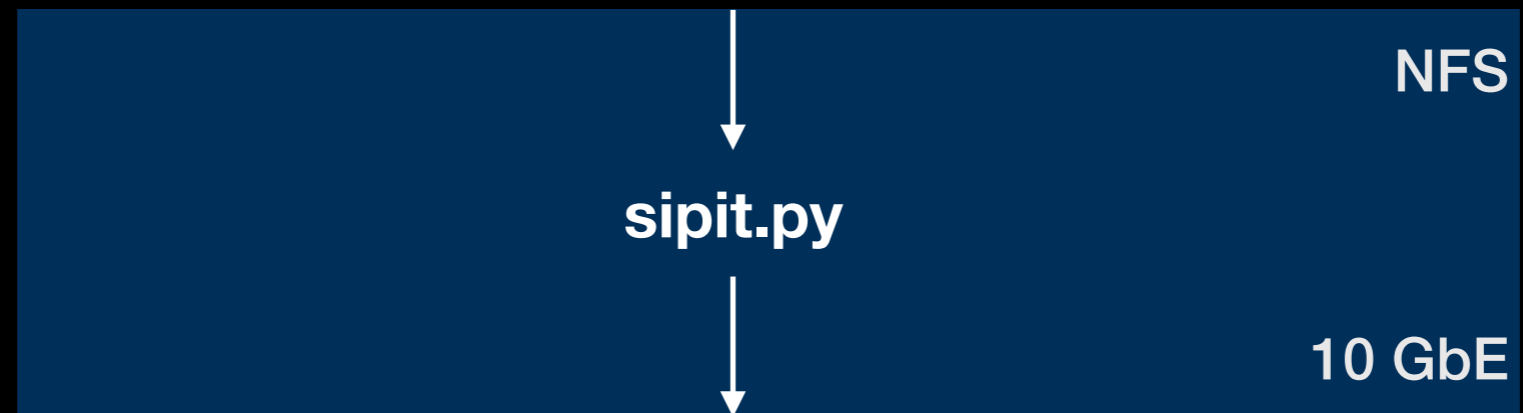


sipit.py

- Jaottelee sip_max_size-paramerin mukaan tiedostot yhteen tai useampaan listaan (alihakemistoittain)
- Yksi lista menee yhteen sip-pakettiin
- Jokaiselle listalle suoritetaan kaikki siptoolsin komennot järjestyksessä (poislukien compress, tar)
- Tarrauksen skripti hoitaa itse:
 - kaikki listan tiedostot suoraan projektihakemistosta
 - perään “appendataan” mets.xml ja signature.sig työhakemistosta.
 - Tarriin menevää rakennetta ei siis tarvitse muodostaa levyille, vaan ainoastaan suoraan tarriin.

KAVI

**R-123_Elokuva/
/image/orig_R-123_Elokuva/
/R-123_Elokuva_osa-1/*.dpx
/R-123_Elokuva_osa-2/*.dpx
/R-123_Elokuva_osa-3/*.dpx**



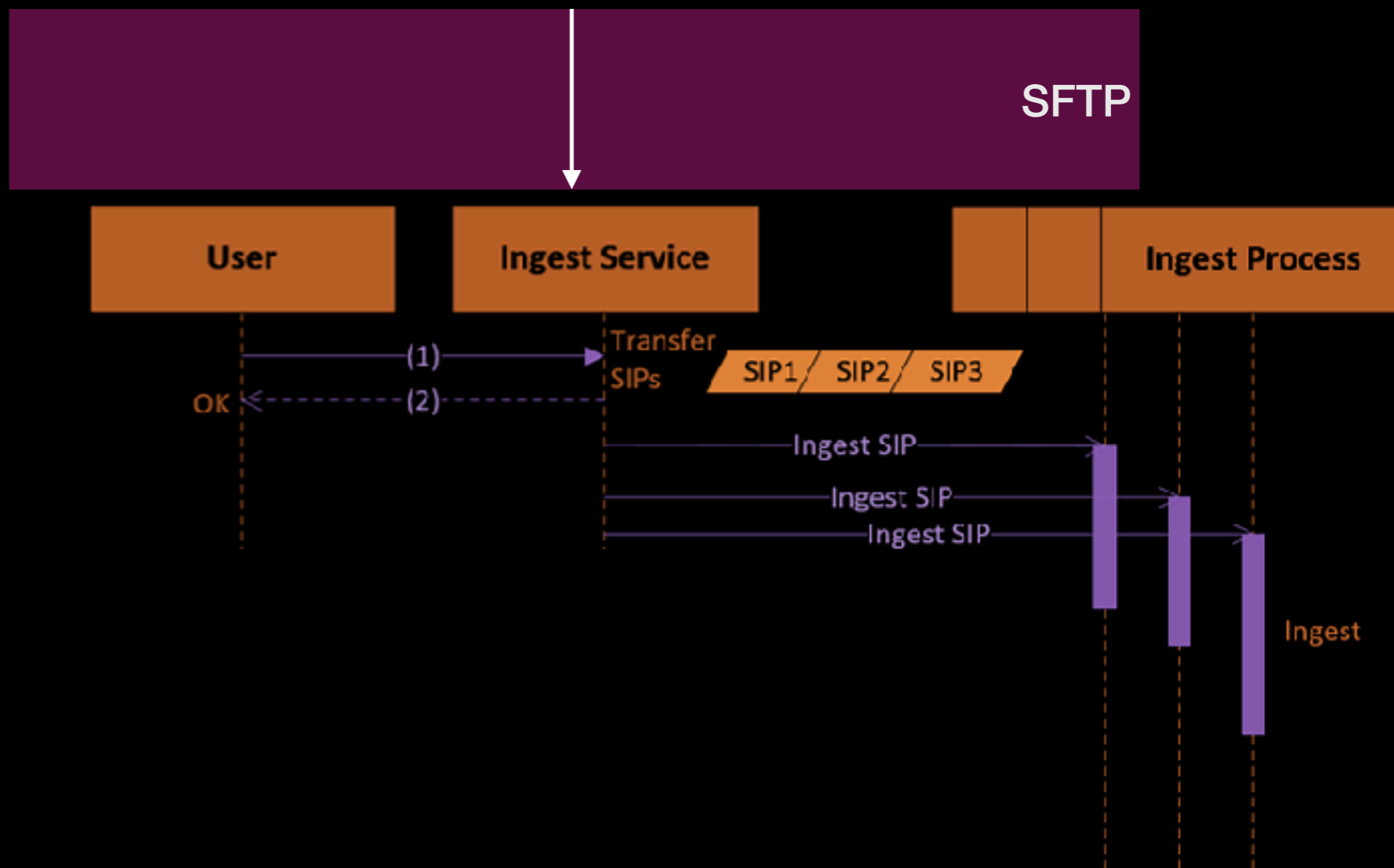
**CSC
Digivarasto**

**archive/
/assets/1/2/3/4/5/
/123456-Elokuva_01of03.tar
/123456-Elokuva_02of03.tar
/123456-Elokuva_03of03.tar**

CSC
Digivarasto

archive/
/assets/1/2/3/4/5/
/123456-Elokuva_01of03.tar
/123456-Elokuva_02of03.tar
/123456-Elokuva_03of03.tar

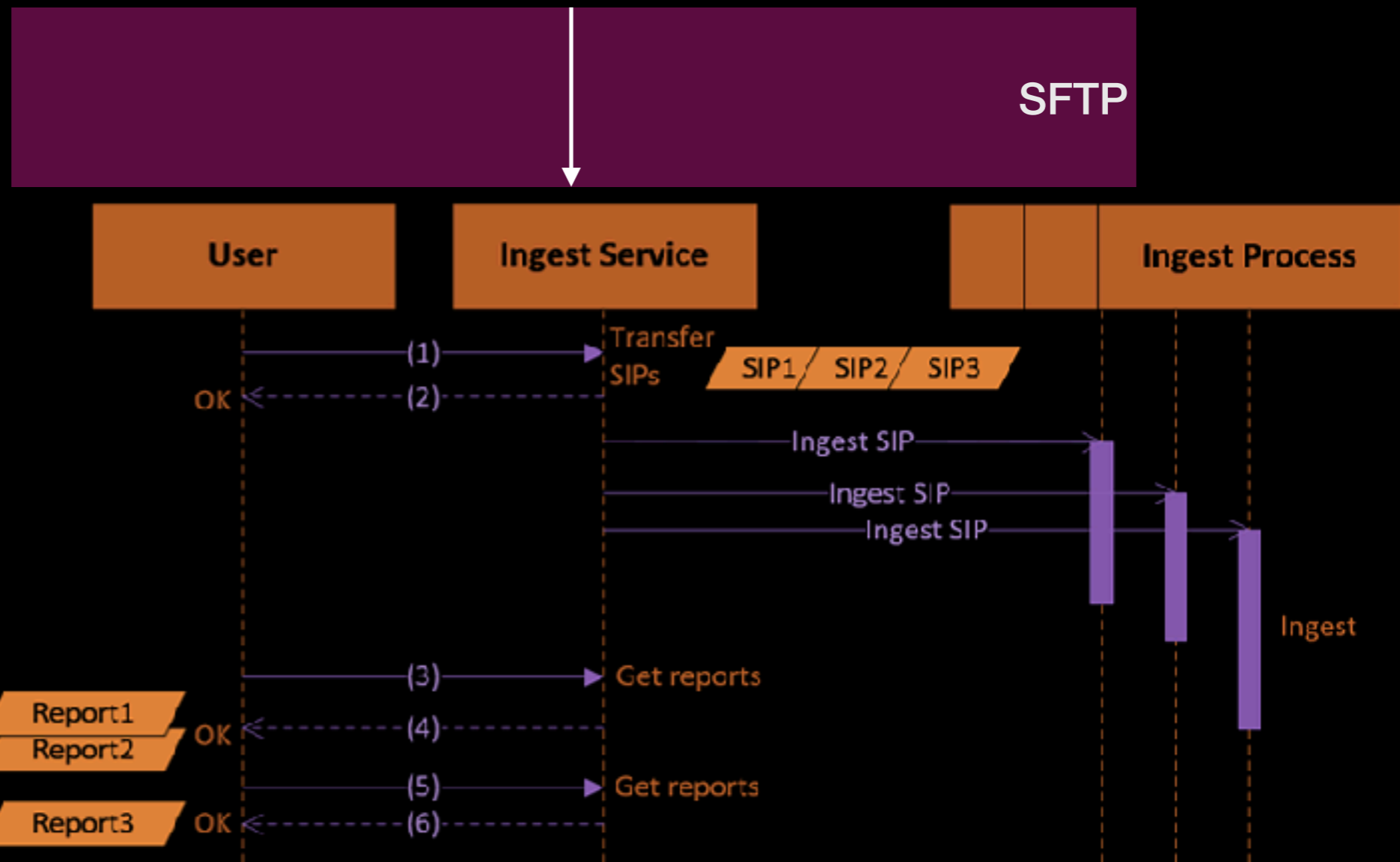
/tallennapas/
/123456-Elokuva.pas



PAS

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

archive/
/assets/1/2/3/4/5/123456-Elokuva.pas
reports/
/accepted/2018-06-12/
/123456-Elokuva_01of03.tar/...-ingest-report.xml
/123456-Elokuva_02of03.tar/...-ingest-report.xml
/123456-Elokuva_03of03.tar/...-ingest-report.xml



PAS

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

```
1 {
2   "create_date": "2017-11-23T14:03:57",
3   "format_version": "2.0",
4   "total_size": 1218930425856,
5   "format_name": "image/x-dpx",
6   "chunks": [
7     {
8       "files": [],
9       "archive_name": "1644639-RajannaenTehtaat_01of02_OC-5025_RajannaenTehtaat_osa-1.tar",
10      "name": "1644639-RajannaenTehtaat_01of02_OC-5025_RajannaenTehtaat_osa-1",
11      "nix_meta": {
12        "colorSpace": "RGB",
13        "bitsPerSampleValue": 10,
14        "imageHeight": 3164,
15        "imageWidth": 4152,
16        "bitsPerSampleUnit": "Integer",
17        "samplesPerPixel": 3,
18        "byteOrder": "big endian"
19      },
20      "sub_dir": "OC-5025_RajannaenTehtaat_osa-1",
21      "descriptive_meta": {
22        "date": "1957",
23        "identifier": "http://kavi.fi/tenho/elokuva-637335/filnikopio-1644639/",
24        "title": "Rajan\u00e4en tehtaat"
25      },
26      "preis_events": [
27        {
28          "agent_name": "SCAMITY",
29          "event_type": "scan",
30          "event_detail": "Digitointi\u00e4ri / RGB\u00e4ri / Full FA 4152x3164, CP0\u00e4ri",
31          "agent_type": "hardware",
32          "event_outcome": "success",
33          "event_datetime": "2017-11-22T12:08:54"
34        }
35      ],
36      "size": 605203218432
37    },
38    {
39      "files": [],
40      "archive_name": "1644639-RajannaenTehtaat_02of02_OC-5025_RajannaenTehtaat_osa-2.tar",
41      "name": "1644639-RajannaenTehtaat_02of02_OC-5025_RajannaenTehtaat_osa-2",
42      "nix_meta": {
43        "colorSpace": "RGB",
44        "bitsPerSampleValue": 10,
45        "imageHeight": 3164,
46        "imageWidth": 4152,
47        "bitsPerSampleUnit": "Integer",
48        "samplesPerPixel": 3,
49        "byteOrder": "big endian"
50      },
51      "sub_dir": "OC-5025_RajannaenTehtaat_osa-2",
52      "descriptive_meta": {
53        "date": "1957",
54        "identifier": "http://kavi.fi/tenho/elokuva-637335/filnikopio-1644639/",
55        "title": "Rajan\u00e4en tehtaat"
56      },
57      "preis_events": [
58        {
59          "agent_name": "SCAMITY",
60          "event_type": "scan",
61          "event_detail": "Digitointi\u00e4ri / RGB\u00e4ri / Full FA 4152x3164, CP0\u00e4ri",
62          "agent_type": "hardware",
63          "event_outcome": "success",
64          "event_datetime": "2017-11-22T12:08:54"
65        }
66      ],
67      "size": 613727207424
68    }
69  ],
70  "total_files": 23168
71 }
```

```
1 {
2   "create_date": "2017-11-23T14:03:57",
3   "format_version": "2.0",
4   "total_size": 1218930425856,
5   "format_name": "image/x-dpx",
6   "chunks": [
7     {
8       "files": [
11512       "archive_name": "1644639-RajamaenTehtaat_01of02_0C-5025_RajamaenTehtaat_osa-1.tar",
11513       "name": "1644639-RajamaenTehtaat_01of02_0C-5025_RajamaenTehtaat_osa-1",
11514       "mix_meta": {
11515         "colorSpace": "RGB",
11516         "bitsPerSampleValue": 10,
11517         "imageHeight": 3164,
11518         "imageWidth": 4152,
11519         "bitsPerSampleUnit": "integer",
11520         "samplesPerPixel": 3,
11521         "byteOrder": "big endian"
11522       },
11523       "sub_dir": "0C-5025_RajamaenTehtaat_osa-1",
11524       "descriptive meta": {
11525         "date": "2017-11-23T14:03:57",
11526         "identifier": "http://www.arkisto.fi/ark:/61906/1644639/",
11527         "title": "Rajam\u00e4en tehtaat"
11528       },
11529       "premis_events": [
11530         {
11531           "agent_name": "SCANITY",
11532           "event_type": "scan",
11533           "event_detail": "Digitointi\n\nv\u00e4ri / RGB\nFull FA 4152x3164, CPD\n",
11534           "agent_type": "hardware",
11535           "event_outcome": "success",
11536           "event_time": "2017-11-23T14:03:57"
11537         }
11538       ],
11539       "size": 605203216432
11540     },
11541     {
11542       "files": [
23208       "archive_name": "1644639-RajamaenTehtaat_02of02_0C-5025_RajamaenTehtaat_osa-2.tar",
23209       "name": "1644639-RajamaenTehtaat_02of02_0C-5025_RajamaenTehtaat_osa-2",
23210       "mix_meta": {
```

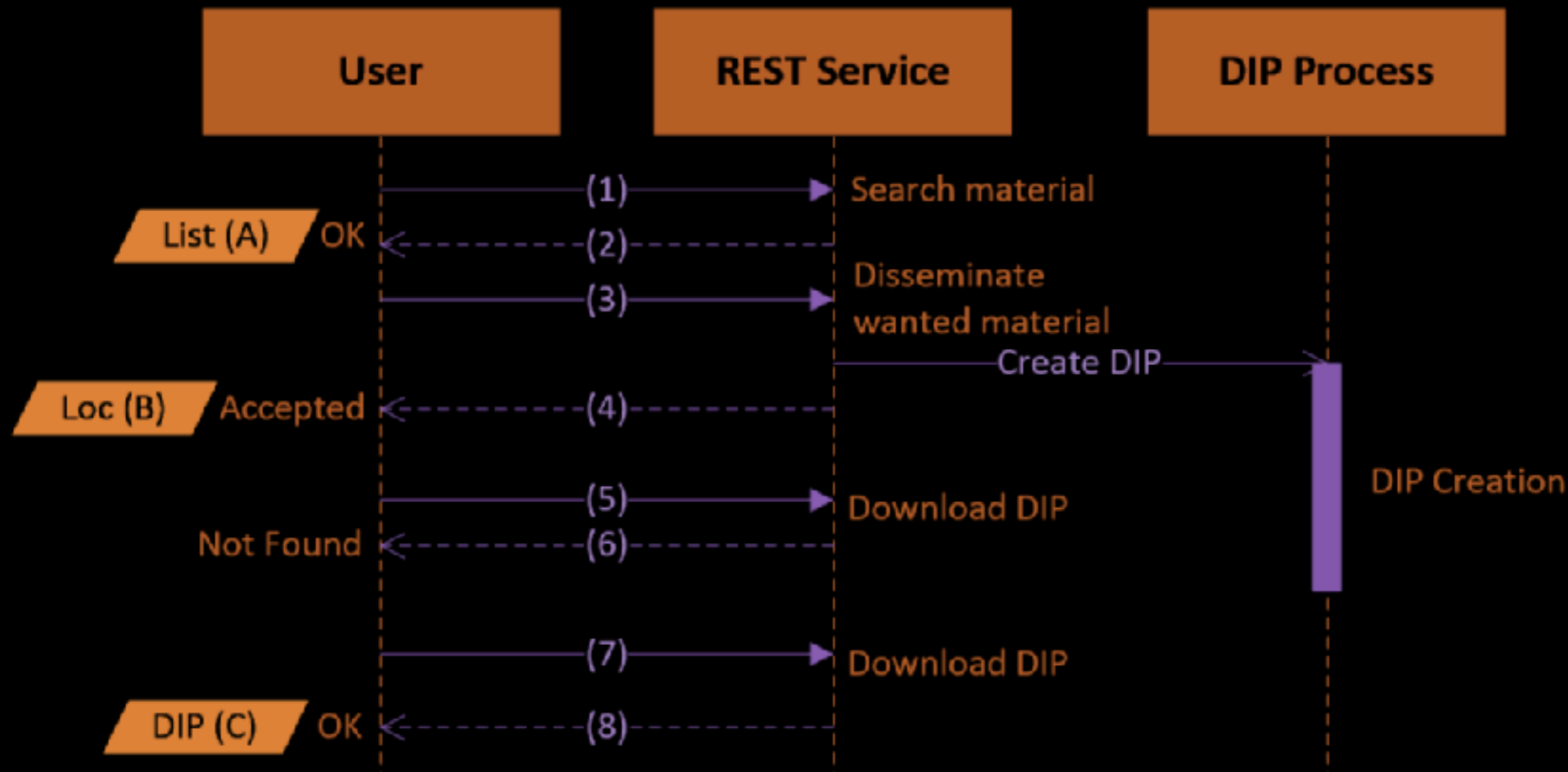
Arkistoinnin yhteydess\u00e4 REST-k\u00e4ytt\u00f6\u00e4 varten Tenhon *Workflow*-tietokantaan tallennetaan: **contentid, objid ja versio (eli aikaleima).**

<contentid> == <tenho_id>-<siivottu elokuvan nimi>
<objid> == <contentid>_<tarvittava osatieto>

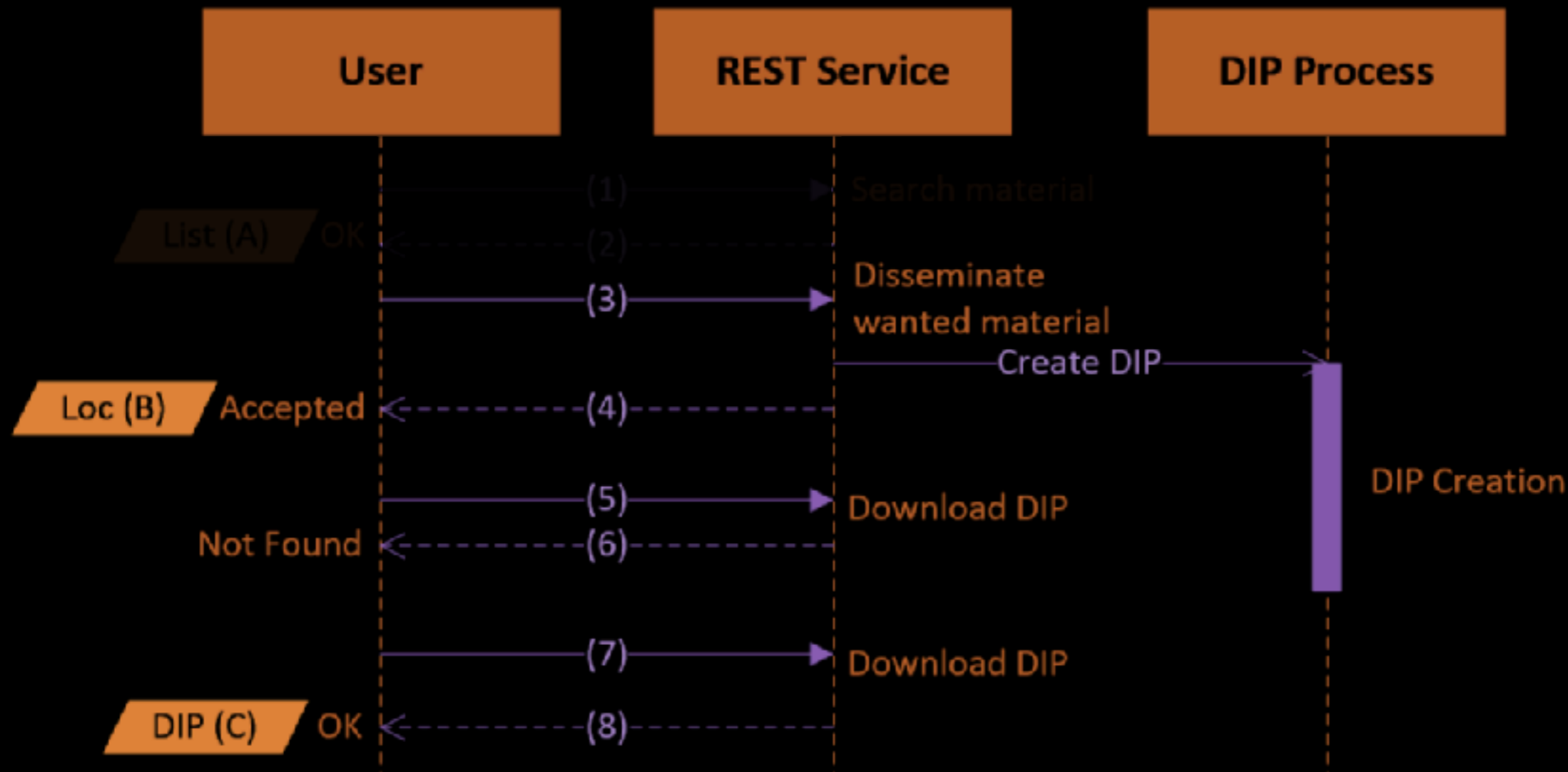

```
1 {
2   "create_date": "2017-11-23T14:03:57",
3   "format_version": "2.0",
4   "total_size": 1218930425850,
5   "format_name": "image/x-dpx",
6   "chunks": [
7     {
8       "files": [
11512     "archive_name": "1644639-RajamaenTehtaat_01of02_0C-5025_RajamaenTehtaat_osa-1.tar",
11513     "name": "1644639-RajamaenTehtaat_01of02_0C-5025_RajamaenTehtaat_osa-1",
11514     "mix_meta": {
11515       "colorSpace": "RGB",
11516       "bitsPerSampleValue": 10,
11517       "imageHeight": 316,
11518       "imageWidth": 4152,
11519       "bitsPerSampleUnit": "integer",
11520       "samplesPerPixel": 3,
11521       "byteOrder": "big endian"
11522     },
11523     "sub_dir": "0C-5025_RajamaenTehtaat_osa-1",
11524     "descriptive meta": {
11525       "title": "Rajamaen Tehtaat",
11526       "identifier": "http://www.tenho.fi/arkisto/1644639/",
11527       "title": "Rajamaen Tehtaat"
11528     },
11529     "premis_events": [
11530     {
11531       "agent_name": "SCANITY",
11532       "event_type": "scan",
11533       "event_detail": "Digitointi\n\nv\u00e4ri / RGB\nFull FA 4152x3164, CPD\n",
11534       "agent_type": "hardware",
11535       "event_outcome": "success",
11536       "event_time": "2017-11-23T14:03:57"
11537     }
11538     ],
11539     "size": 605203216452
11540   },
11541   {
11542     "files": [
23208     "archive_name": "1644639-RajamaenTehtaat_02of02_0C-5025_RajamaenTehtaat_osa-2.tar",
23209     "name": "1644639-RajamaenTehtaat_02of02_0C-5025_RajamaenTehtaat_osa-2",
23210     "mix_meta": {
```

Arkistoinnin yhteydessä REST-käyttöä varten Tenhon *Workflow*-tietokantaan tallennetaan: **contentid, objid ja versio** (eli aikaleima).

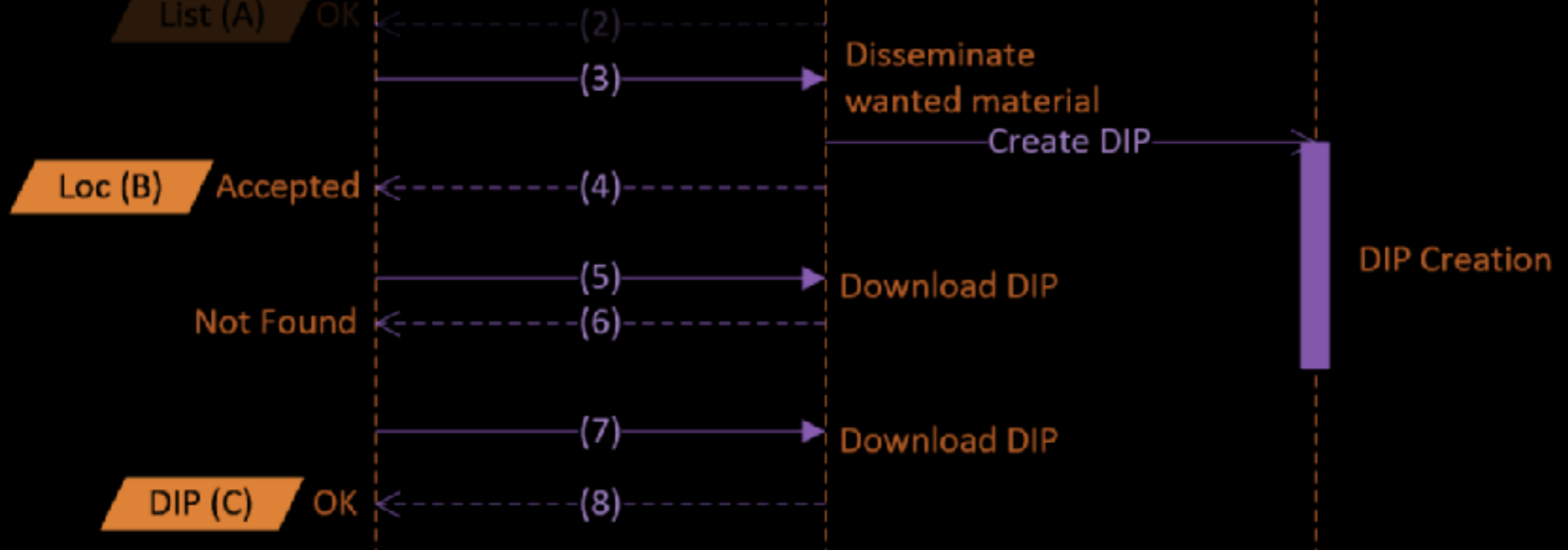
<contentid> == <tenho_id>-<siivottu elokuvan nimi>
<objid> == <contentid>_<tarvittava osatieto>



Noudot Workflow tilaa PASin REST-rajapinnasta disseminate-käyttötapauksella, parametrina tallennetut objid ja versio.



Noudot Workflow tilaa PASin REST-rajapinnasta disseminate-käyttötapauksella, parametrina tallennetut objid ja versio.



Noudot Workflow tilaa PASin REST-rajapinnasta disseminate-käyttötapauksella, parametrina tallennetut objid ja versio.

RESTin palauttaman dipid:n Workflow tallentaa Digivarastolle *noudapas*-hakemistoon .pas-tiedostoon yhdessä objid:n (=haluttu tar-tiedoston nimi *assetsissa*) kanssa.

Digivarasto pollaa RESTin `disseminated/<dipid>/status` -käyttötapausta, kunnes status on "done", ja hakee sitten dipin sftp:llä *assetsiin*.

Palautetta PAS:lle kehittäjän näkökulmasta:

- REST-rajapinnan puuttuminen testiympäristöstä on iso puute.**
- Rajapinnat on dokumentoitu hyvin, vaikka osin dokumentaatio kuvaa vielä tavoitetilaa.**
 - Ilmeisesti toistaiseksi ainoan oikeasti käytössä olevan rajapintaversioon (1.0) dokumentaatio olisi hyvä pitää helposti löydettävissä.**