



# Digitaalisten aineistojen tapahtumat

22.4.2021 Juha Lehtonen



# Tapahtumat

- Tapahtumien avulla voidaan dokumentoida aineiston elinkaari
- Kunkin tapahtuman tarkoitus on kertoa, mitä aineistolle on tehty ja milloin
- Oleellisten tapahtumien dokumentoiminen lisää ymmärrettävyyttä ja vähentää riskejä
- Tapahtumien kuvailu lisää aineiston todistusvoimaisuutta

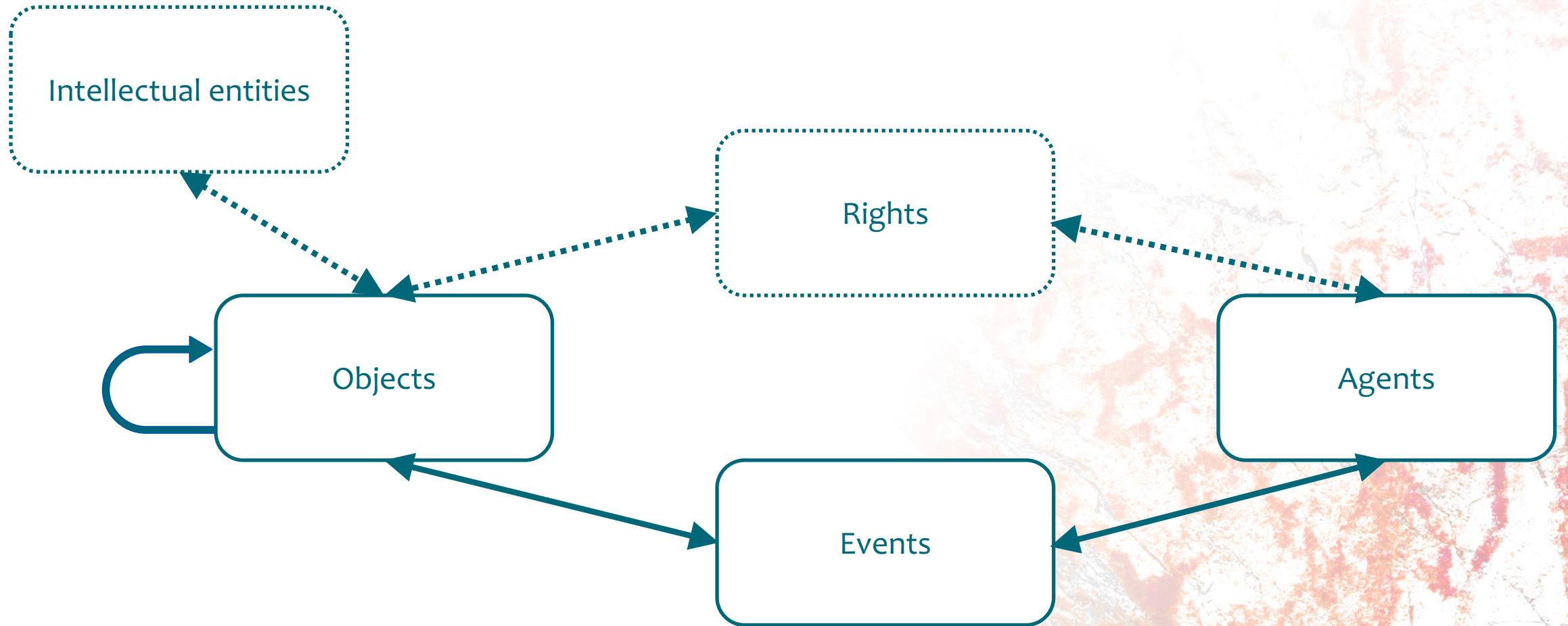
# Tapahtuma on kuin lauseen predikaatti, jolla on kuvailu

- Suomen kielen lauseessa: *subjekti-predikaatti-objekti*
  - esim. "Ohjelma validoi tiedoston."
  - "Ohjelma" (subjekti)
  - "validoi" (predikaatti)
  - "tiedoston" (objekti).
- Vastaava kuvailu tapahtumana: *agentti-tapahtuma-objekti*
  - "Kuvailtu ohjelma" (agentti, kuka tai mikä teki)
  - "validoi kuvastusti" (tapahtuma, mitä tehtiin)
  - "kuvailun tiedoston" (objekti, mihin kohdistui).

# Suomen kielen lause tapahtumakuvaileksi

- “Ohjelma VeraPDF 1.16 validoi onnistuneesti dokkari.pdf-tiedoston uudenvuodenpäivänä 2021 klo 14:00:00.”
- **Agentti**
  - Agenttina on ohjelma (“software”)
  - Ohjelman nimi on “VeraPDF”
  - Ohjelman versio on “1.16”
  - **Agentilla on joku tunniste**
- **Objekti**
  - Objektina on tiedosto (“file”)
  - Tiedoston nimi on “dokkari.pdf”
  - **Objektilla on joku tunniste**
- **Tapahtuma**
  - Tapahtumana on validointi (“validation”)
  - Validoinnin ajankohta on 01.01.2021 klo 14:00 (“2021-01-01T14:00:00+02:00”).
  - Validointi onnistui (“success”)
  - Validoinnin kuvaus: “File format validation of a digital object”
  - **Viitataan tunnisteiden avulla agenttiin ja objektiin, jotta “koko lause saadaan ehjäksi”**

# PREMIS-metatietomalli



# Esimerkki (1/2)



"Jouluaattona 2020 CSC migroi superhienolla PDF-migraattorilla siirtokelpoisen 'dokkari.pdf'-tiedoston säilytyskelpoiseksi 'dokkari\_migrated.pdf'-tiedostoksi."

- **Agentti**

- Tunniste: "csc"
- Tyyppi: "organization"
- Organisaation nimi: "CSC – IT Center for Science Ltd."

- **Agentti**

- Tunniste: "software-007"
- Tyyppi: "software"
- Ohjelman nimi: "Super Great PDF Migrator"
- Ohjelman versio: "1.0"

- **Objekti**

- Tunniste: "doc-1"
- Tyyppi: "file"
- Tiedostonimi: "dokkari.pdf"
- Tiedostomuoto: "application/pdf" (PDF-tiedosto)
- Tiedostonmuodon versio: "1.7" (siirtokelpoinen)
- Tarkistussumma: ("sha-1", "123")
- Luontiaika: "2009-01-01T14:00:00+02:00"

- **Objekti**

- Tunniste: "doc-2"
- Tyyppi: "file"
- Tiedostonimi: "dokkari\_migrated.pdf"
- Tiedostomuoto: "application/pdf" (PDF-tiedosto)
- Tiedostomuodon versio: "A-2a" (säilytyskelpoinen)
- Tarkistussumma: ("sha-1", "abc")
- Luontiaika: "2020-12-24T13:30:00+02:00"

## Esimerkki (2/2)

“Jouluaattona 2020 CSC migroi superhienolla PDF-migraattorilla siirtokelpoisen ‘dokkari.pdf’-tiedoston säilytyskelpoiseksi ‘dokkari\_migrated.pdf’-tiedostoksi.”

- Tapahtuma

- Tunniste: “event-001”
- Tyyppi: “migration”
- Aikaleima: “2020-12-24T13:30:00+02:00”
- Kuvaus: “Migration of a digital object from a file format acceptable for transfer to a recommended file format.”
- Lopputulos: “success”
- Migraatioajon tuloste: “...”

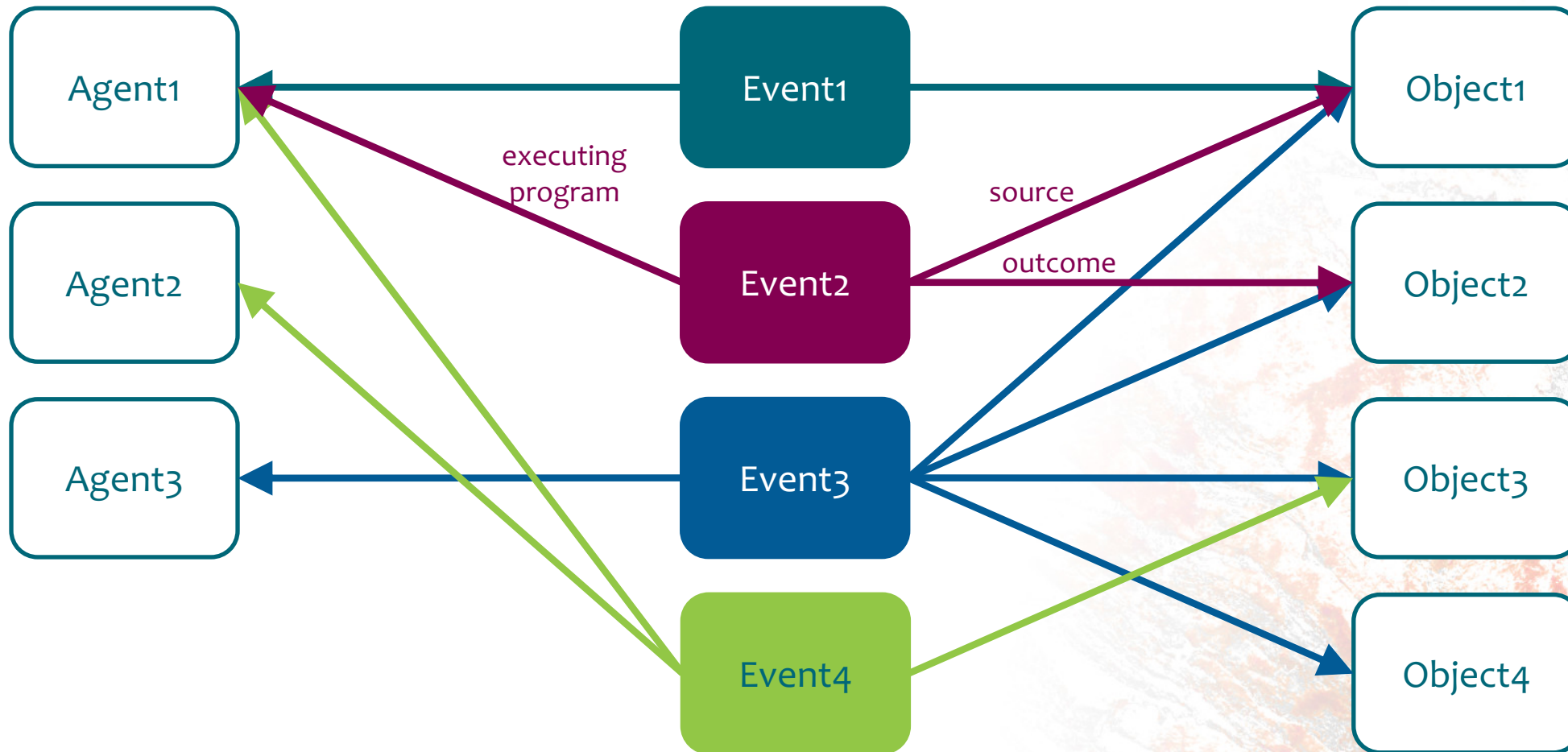
- Viittaukset agentteihin:

- Agentin tunniste: “csc”  
Agentin rooli: “implementer”
- Agentin tunniste: “software-007”  
Agentin rooli: “executing program”

- Viittaukset objekteihin:

- Objektin tunniste: “doc-1”  
Objektin rooli: “source”
- Objektin tunniste: “doc-2”  
Objektin rooli: “outcome”

# Erilaisia viittauksia





# Mitä tapahtuma sisältää?

Tapahtumatyyppi

- Seuraavissa dioissa lisää

Tunniste

- Tunnisteen tyyppi ja arvo

Aikaleima

- Voi olla myös epätarkka aika

Selitys

- Kuvaus tai muu oleellinen ihmisluettava tieto tapahtumasta

Tapahtuman tulos

- “success”, “failure”, “unknown”

Lopputuloksen lisätiedot

- Esim. ajetun ohjelman koko tuloste

Viittaukset agentteihin

- suositellaan myös agenttien roolit tapahtumassa

Viittaukset objekteihin

- suositellaan myös objektien roolit tapahtumassa

# Tapahtumatyypit ja paketointikomponentti

- Tapahtumatyyppejä on määritelty kymmeniä
  - Lista on jo aika kattava, mutta voidaan myös yhdessä sopien laajentaa
- Paketointikomponentissa on työkalut tapahtumien, agenttien ja objektien tuottamiseen.
- Siirtopakettia luotaessa paketointikomponentti luo tarvittaessa automaattisesti seuraavantyyppisiä tapahtumia (ja niille agentit):

metadata extraction

• digitaalisesta objektista poimitaan objektia koskevia teknisiä metatietoja

format identification

• selvitetään digitaalisen objektin tiedostomuoto ja versio

validation

• tiedostomuodon validointi

message digest calculation

• tiedoston tarkistussumma luodaan

creation

• uusi rakennekartta luodaan

# Tapahtumatyyppejä digitoinnissa / normalisoinnissa (esimerkiksi)

creation

- luodaan uusi kohde

normalization

- ei-PAS-tuettu objekti muunnetaan säilytyskelpoiseen tiedostomuotoon

migration

- kohde muunnetaan uuteen tiedostomuotoon (korvaa vanhan version)

modification

- tiedoston tai bittivirran muuttaminen (ei synny uutta objektia)

message digest calculation

- kohteen tarkistussumma luodaan

metadata extraction

- tietopaketista tai objektista poimitaan metatietoja

metadata modification

- kohteen metatietoihin tehdään muutoksia

validation

- kohde validoidaan

exporting

- esimerkiksi kohteesta luodaan käyttökopio

appraisal

- arvioidaan kohteen säilytyskelpoisuus

# Tapahtumatyypit PAS-vastaanotossa

transfer

- aineiston omistaja siirtää objekteja/metatietoa säilytyspaikkaan

unpacking

- kohteita puretaan säiliöformaattista

virus check

- kohde virustarkastetaan

digital signature validation

- digitaalinen allekirjoitus validoidaan

validation

- kohde validoidaan

fixity check

- kohteen eheys tarkastetaan ja varmistetaan, että kohde ei ole muuttunut

information package creation

- luodaan tietopaketti

accession

- lisätään kohde säilytyspaikkaan, ja säilytysvastuu siirtyy

- Palvelusopimukseen liittyvät ominaisuudet
- METS-dokumentti
- Tiedostomuodot (kukin tiedosto erikseen)
- Vastaanoton tapahtumat



# Mitä metatietoja agentille?

Tyyppi

- *person, organization, software, hardware*

Tunniste

- tunnisteiden tyyppi ja arvo

Nimi

- ohjelman tapauksessa nimi ja versiotieto

Lisätiedot

- esimerkiksi kuvaus konfiguraatiosta, jota ohjelmalle käytetään

(Lisäosio)

- Vapaamuotoinen paikallisesti koneluettava lisäosio

(Viittaukset)

- Paketointikomponentti luo viittaukset tapahtumaan, ei agenttiin

# Mitä metatietoja objekteille?

- Objekti-metatiedot voivat sisältää laajasti monenlaista teknistä metatietoa

Tiedoston yleiset tekniset piirteet ja ominaisuudet

Ympäristöriippuvuudet (laitteisto, ohjelmat, aputiedostot)

Digitaaliset allekirjoitukset

Erilaisia relaatioita muihin objekteihin (esim. videosäiliö sisältää kuva- ja ääniraitoja)

Viittauksia (esim. muihin objekteihin)

...

# PREMIS-metatieto-objektien tyypit

file

- tiedosto

bitstream

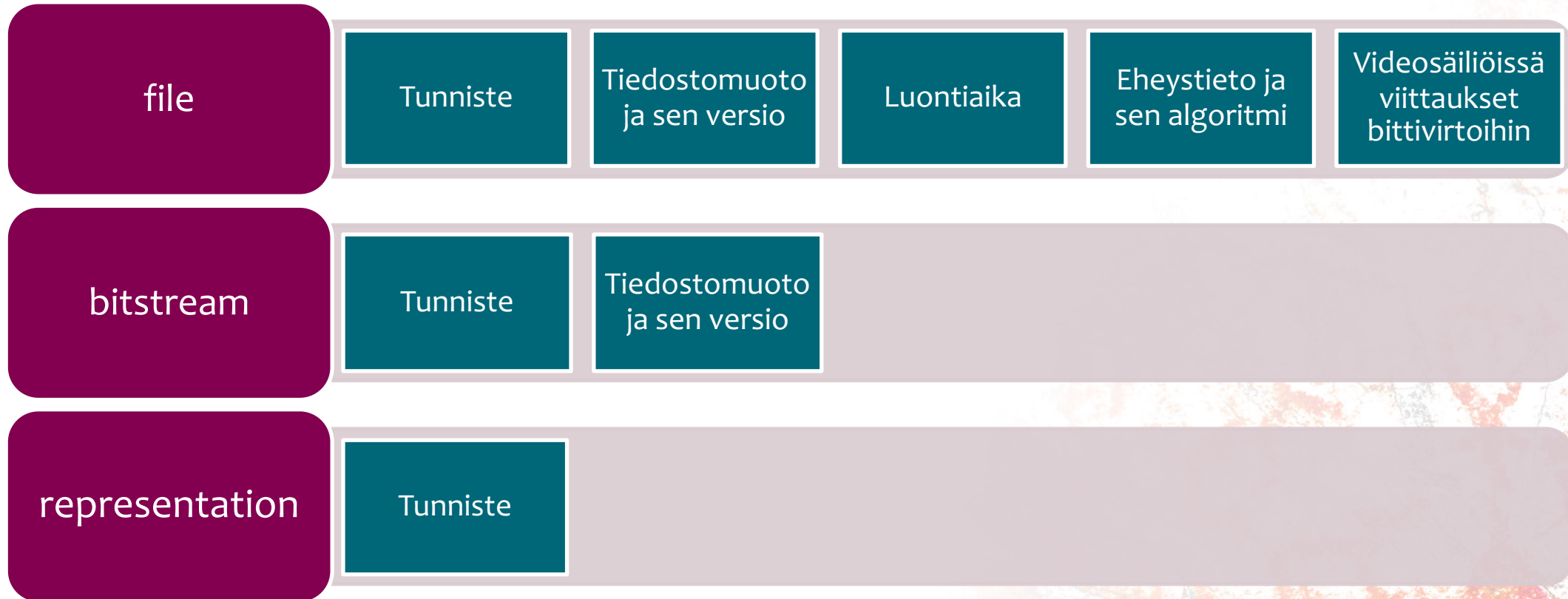
- bittivirta, esim. videosäiliön ääniraita

representation

- Esim. jonkin objektin tai kokonaisuuden edustaja

Objektin tyypistä riippuu, mitä tietoja metatiedot voivat sisältää

# Mitä objekti-metatiedoissa on pakko olla?

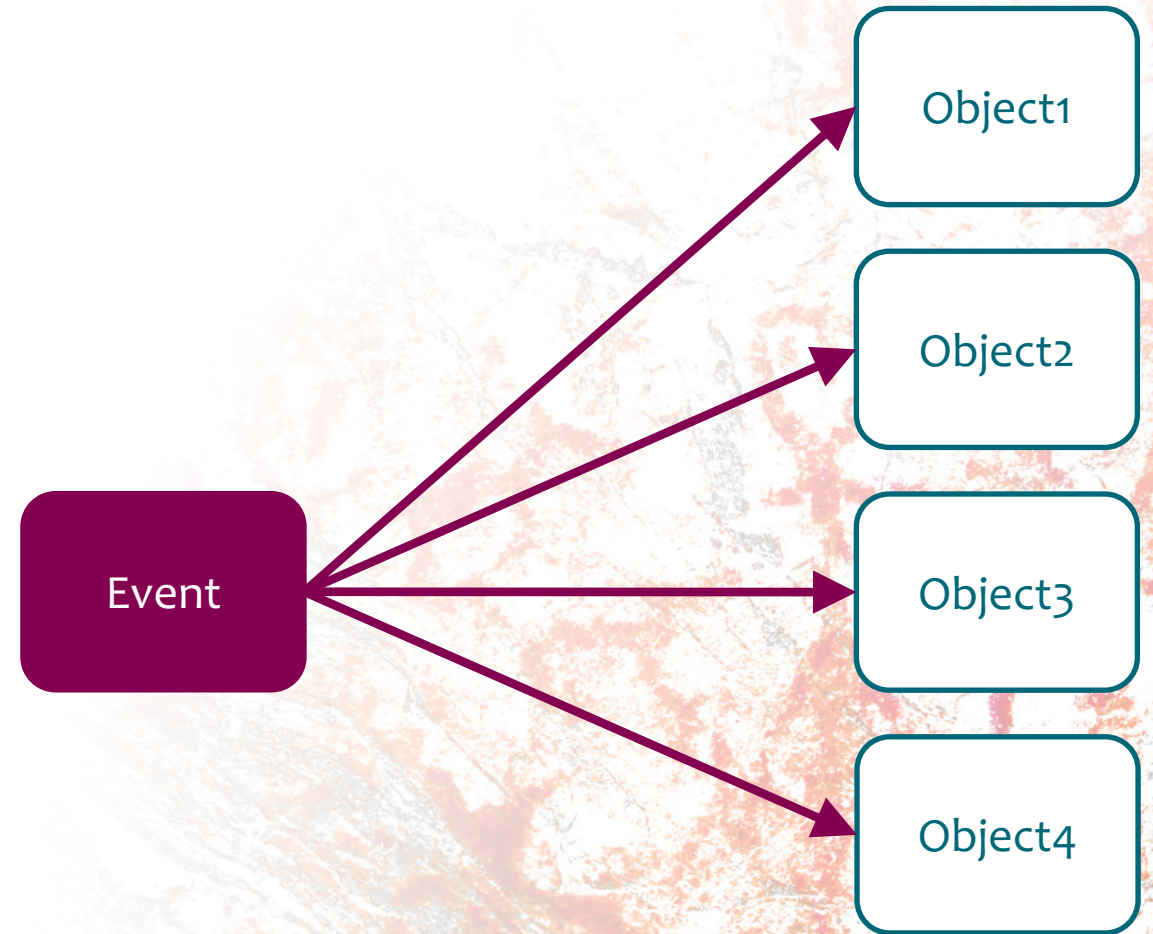


- Tässä ei kuitenkaan ole kaikki pakolliset tekniset metatiedot
  - Muut tekniset metatietomuodot määrittelyssä: ADDML, MIX, videoMD, audioMD
- Paketointikomponentti osaa tuottaa digitalisten objektien pakolliset tekniset metatiedot automaattisesti



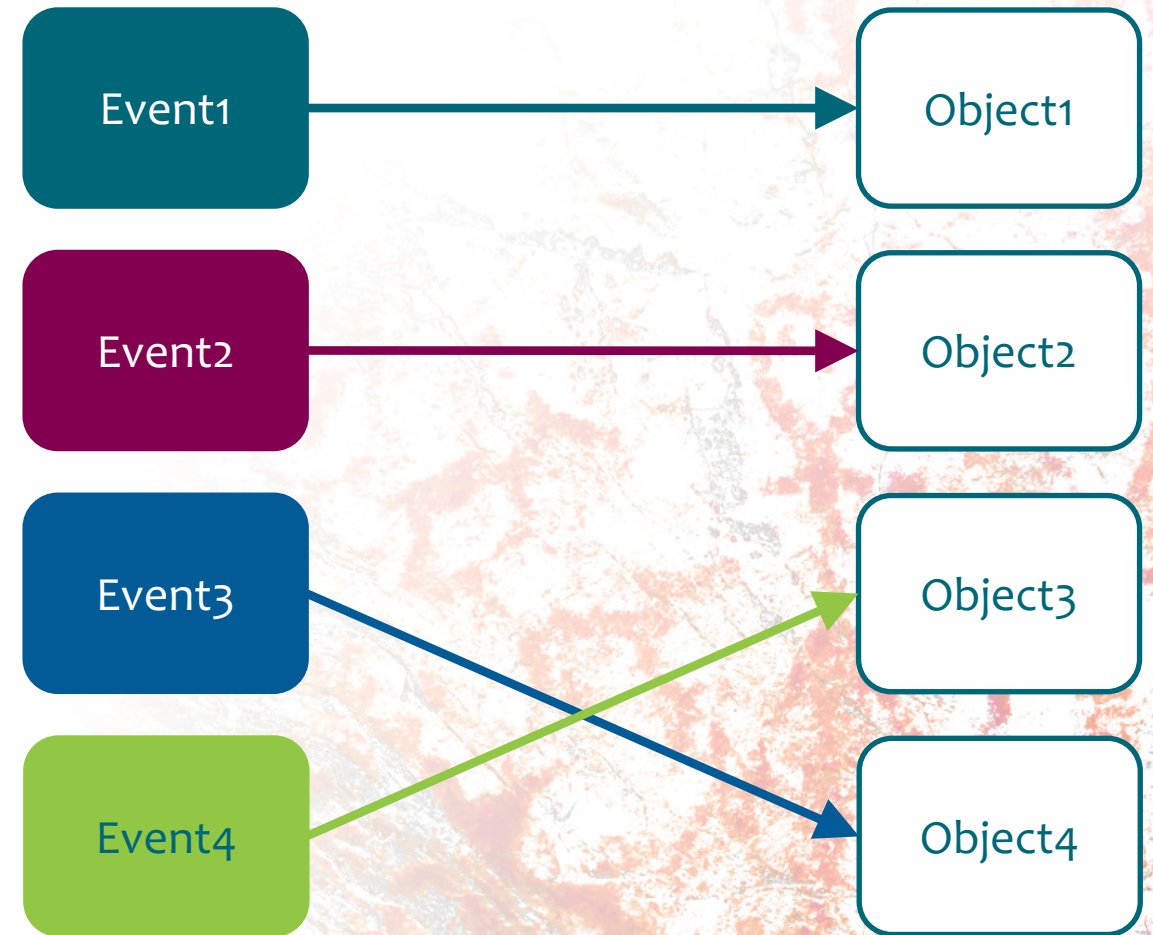
# Yksi tapahtuma vai monta tapahtumaa?

- Skenaario: Tehdään jokaiselle digitaaliselle objektille erikseen jokin yksi tietty asia
- Joskus voi olla tarpeetonta tehdä niistä jokaisesta tapahtuma, eli riittää yksi tapahtuma
- Esimerkkinä: virus- tai eheystarkastus
  - Kumpikaan ei itsessään tuota paljoakaan tarpeellista tulostetta, jos kaikki on ok.
  - Jos kaikki ei ole ok, voidaan tapahtuman lopputuloksessa luetella ongelmatapaukset.
  - Nämä tarkastukset ovat ajoittain toistuvia
  - Jos tiedostoja on paljon ja jokaisen tiedoston tarkastuksesta tehtäisiin oma tapahtumansa, syntyisi ajan mittaan valtava tapahtumaraportti, jota on vaikea hallita.



# Yksi tapahtuma vai monta tapahtumaa?

- Joskus on kuitenkin järkevää tehdä oma tapahtumansa jokaiselle digitaaliselle objektille erikseen
  - Vaikka se tuottaisikin paljon informaatiota
- Esimerkki: Tiedostomuotojen validointi
  - Eri tiedostomuodoille käytetään eri validaattoreita.
  - Jokaisen validaattorin tuloste on yksilöllinen ja yleensä tiedostokohtainen.
  - Erottelu, mikä validointituloste koskee mitään tiedostoa, on vaikeaa, jos kaikille tuotettaisiin vain yksi yhteinen tapahtuma.



# Paketointikomponentti

