



DEPIP-määrittäminen rajapintojen toteutuksessa



CSC – Suomalainen tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon ICT-osaamiskeskus

Mikä on DEPIP?



- *Protokolla*: “Data exchange **protocol** for interoperability and preservation” (DEPIP)
- *Viitekehys*: “DEPIP specifies a standardized **framework**...”
- *Käsitteellinen tietohakemisto*: “Note that DEPIP is a **conceptual standard** to be considered as a **data dictionary**.”
- *Merkintäkieli*: “It also specifies the **syntax** and semantics of the messages...”
 - Tai sitten ei: “Other encodings may be added in the future.”
 - “The model defined in DEPIP is independent with regard to implementation issues.”
 - “However, it is impossible to foresee what the implementations will be like...”
- *Toteuttamissuunnitelma*: Sisältää luokkakaaviot tietueiden nimineen, kardinaliteetteineen ja tietotyyppeineen
- Standardi: ISO 20614:2017

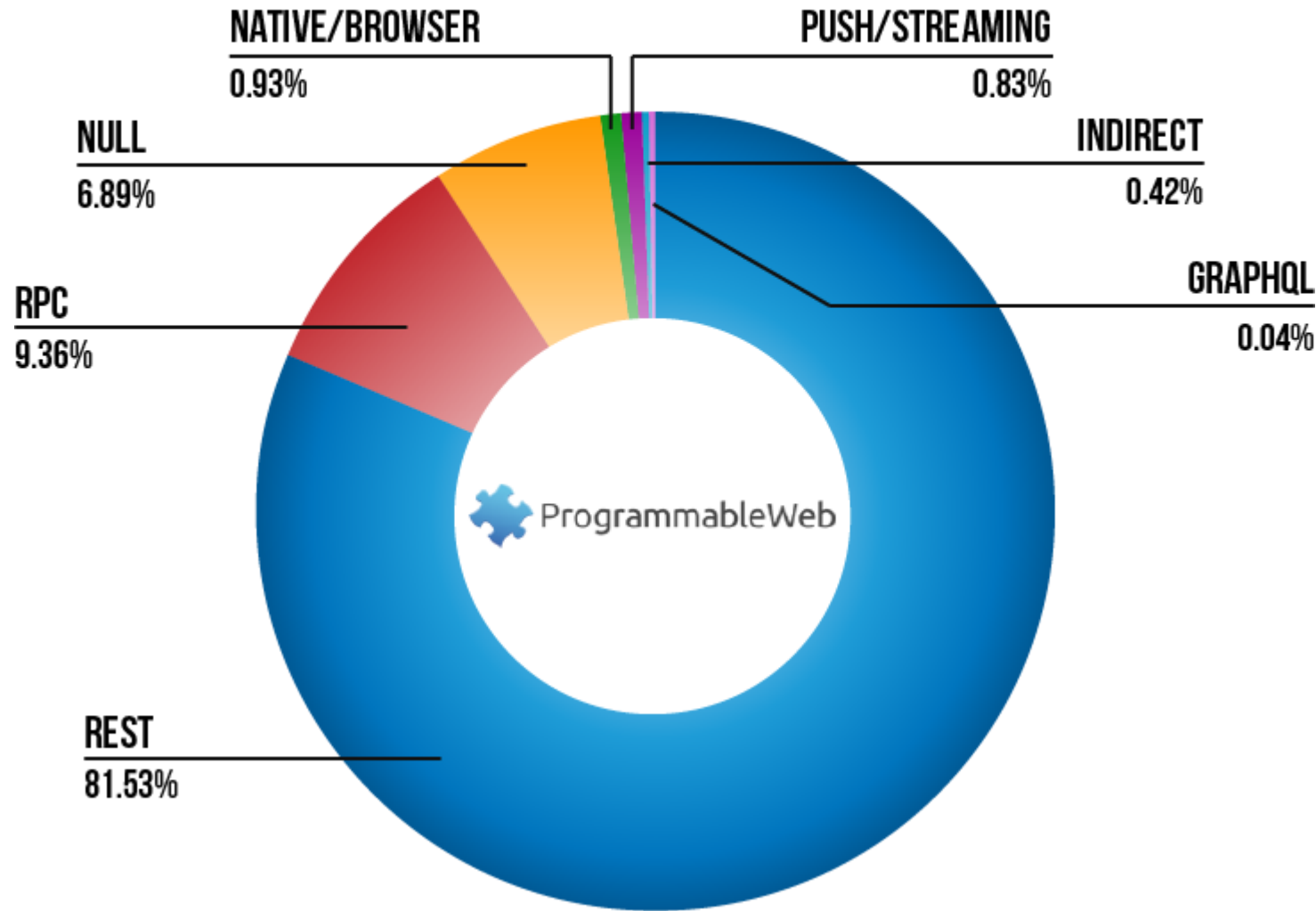
DEPIP-toiminnot

- DEPIP-standardi määrittelee seuraaville toiminnoille sellaiset eri järjestelmien väliset transaktiot, jotka ovat PAS-toiminnassa tarpeellisia:
 - Transfer: Siirtoluvan pyyntö ja aineiston siirto PAS-palveluun
 - Deliver: Aineiston jakelu PAS-palvelusta
 - Modify: Aineiston muokkaus PAS-palvelun toimesta
 - Dispose: Aineiston poisto PAS-palvelun toimesta
 - Restitute: Aineiston palautus säilytysvastuineen PAS-palvelusta takaisin alkuperäiselle toimijalle

Historiaa

- Vuonna 2006 Ranskan arkisto- ja kirjastosektorille luotiin PAS-toimintaan MEDONA-rajapinta, joka perustui (silloin yleisesti käytössä olleeseen) SOAP XML –protokollaan.
- Samana vuonna (2006) Google ilmoitti deprekoivansa SOAP-rajapintojensa tukemisen ja siirtyi REST-arkkitehtuuriin
 - Koko muu maailma seurasi perässä – niin suuret kuin pienetkin pelurit.
 - Syy muutokseen oli, että XML-ratkaisut koettiin kaikkien kannalta liian työläiksi (ja siten kalliiksi):
 - SOAP XML:ssä mikä tahansa pienikin muutos vaatii kaikilta osapuolilta korjauksia koodiin.
 - REST mahdollistaa valmiin ja laajan automatiikan mukaan ottamisen, kun se rakennetaan HTTP(S)-protokollan päälle, kuten sisäänrakennettu virheen käsittely tai välimuisti
 - Vuonna 2009 Googlen SOAP-rajapinnat suljettiin lopullisesti.
- Tänä päivänä 80% maailman aktiivisena olevista koneluettavista julkisista verkkorajapinnoista on toteutettu REST:llä
 - Osuus kasvaa koko ajan.
 - SOAP XML:n osuus on jo alle 10%, ja suuri osa näistäkin on pitkään toiminnassa olleita rajapintoja.
- MEDONA:n toteutusriippumattomasta määrittelystä DEPIP:stä tehtiin ISO-standardi 2017
 - Viittaukset SOAP XML:ään yritettiin häivyttää, mutta historian painolasti jäi standardiin vahvasti.

ARCHITECTURAL STYLES RECORDED IN PROGRAMMABLEWEB DIRECTORY



- ProgrammableWeb on merkittävä avoin ja aktiivinen tietokanta julkisista verkko-ohjelmointirajapinnoista
- Toiminut vuodesta 2005 alkaen
- SOAP-rajapinnat sisältyvät RPC-osuuteen, ja pääasiassa ne ovat tietokannan alkuajoilta
- Kuva on ProgrammableWebin tekemästä selvityksestä marraskuulta 2017



soap api
Hakutermi

rest api
Hakutermi

+ Lisää vertailu

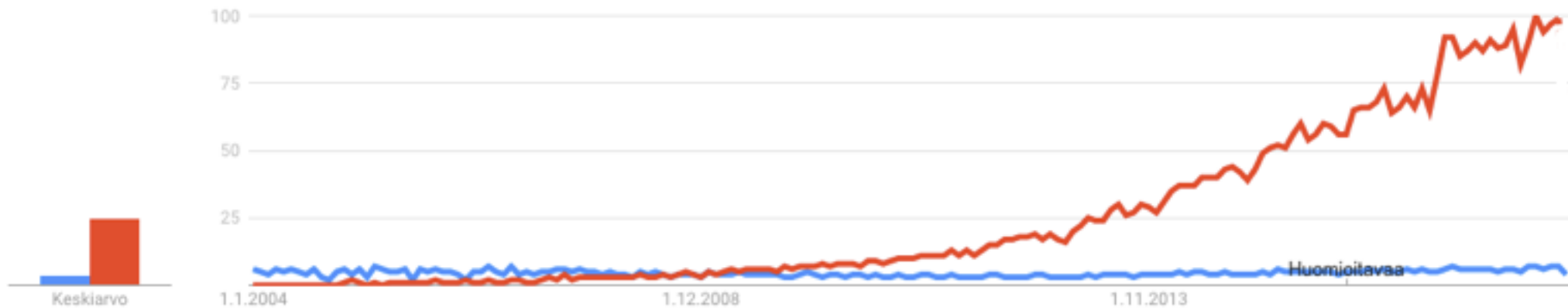
Koko maailma ▾

2004 - nykyhetki ▾

Kaikki luokat ▾

Verkkohaku ▾

Hakumäärät ajan mittaan ?



DEPIP: Asiakas-Palvelin-rajapinnat



- DEPIP-standardi ei kovinkaan hyvin selvitä sellaista tilannetta, joissa keskustelijoiden toinen osapuoli on palveluntarjoajan roolissa ja toinen on asiakkaan roolissa
- Asiakas-palvelin-rajapinnoissa lähtökohta on se, että asiakkaalla ei ole olemassa yhtään mitään palveluja palveluntarjoajaa varten
 - Palveluntarjoaja ei voi lähettää omin päin viestejä asiakkaalle suoraan, koska asiakas ei tarjoa jatkuvasti toimivaa ja turvallista palvelua, joka ottaisi viestejä vastaan
 - Sen sijaan palveluntarjoaja pystyy kyllä vastaamaan asiakkaan kysymykseen
 - Vain asiakas voi kysyä, ja jäädä odottamaan vastausta haluamansa ajan
 - Vrt. jäätelökioski: Myyjällä (palveluntarjoaja) on kyllä mahdollisuus keskustella aloitteen tehneen (tiskille tulleen) asiakkaan kanssa, mutta ei voida olettaa, että myyjä saisi asiakkaaseen yhteyttä koska tahansa samalla menetelmällä, siis esim. sen jälkeen, kun asiakas on poistunut jäätelökioskilta.
- DEPIP ei tee mitään pohdintaa, minkälainen rooli kenelläkin keskustelijalla on tai voi olla, ja siksi standardi on ohjelmistokehittäjän näkökulmasta epäselvä

DEPIP: Asiakas-Palvelin-rajapinnat



- Jos hyödyntävän organisaation odotettaisiin tarjoavan rajapintapalveluja PAS-palvelulle, se tarpeettomasti lisäisi hyödyntävän organisaation taloudellisia kustannuksia mm. ylläpidon, aikavaatimusten ja turvallisuusnäkökohtien osalta.
 - Lisäksi PAS-palvelun toteuttamat auditoinnit voisivat tällöin mahdollisesti levitä hyödyntävälle organisaatiolle asti, vaikka hyödyntävä organisaatio ei itse sitä suoranaisesti tarvitsisi
 - Hyötyjä hyödyntävälle organisaatiolle tällaisessa järjestelyssä ei käytännössä olisi, mutta työtä sitäkin enemmän
- Suomi esitti standardin luonnosvaiheessa tähän ongelmaan selvennystä. Korjausehdotus...
 - roolitti asiakkaan ja palveluntarjoajan erikseen
 - poisti asiakkaalta veloitteet tarjota palveluja palveluntarjoajalle
 - oli täysin yhteensopiva DEPIP-luonnoksessa esitetyn mallin kanssa eikä (ainakaan enempää) estänyt muita roolituksia tai tilanteita
 - oli toimintaperiaatteeltaan yhtenevä PAS-palveluiden rajapintamäärittelyssä olevien vastaavien toimintojen kanssa
- Korjausehdotusta ei sellaisenaan hyväksytty, mutta se muutettiin johdannossa seuraavaksi:
 - “Note that DEPIP is a conceptual standard to be considered as a data dictionary. The model defined in DEPIP is independent with regard to implementation issues. In the different implementations, DEPIP can be supplemented by relevant technical protocols (like HTTP) allowing implementers to handle technical exchanges between systems.”
- DEPIP ei estä Suomen korjausehdotuksen toteuttamista, eli hyödyntävältä organisaatiolta ei vaadita palveluntarjoajan roolia, mutta asiaa voisi standardissa käsitellä selvemmin.

Perusongelma: Vastauksen odottaminen vaatii lisäviestintää



- Jotkut toiminnot saattavat kestää jonkin aikaa, kuten jakelupaketin valmistuminen
 - Koska asiakas-palvelin-rajapinnassa asiakas ei tarjoa palveluja, ei valmistunutta jakelupakettia voida suoraan toimittaa asiakkaalle sen jälkeen, kun kysely ehtii aikakatketta.
 - Se voidaan kuitenkin toimittaa asiakkaan saataville niin, että asiakaspää pollaa (seuraa) valmistumista omalla aikataulullaan (seuraaminen saa olla jatkuvaa, mutta ei tarvitse olla).
- (Lähes) kaikki maailmassa toimii vastaavassa tilanteessa samalla tavalla:
 - Sähköpostiohjelma pollaa uudet viestit palvelimella olevasta sähköpostilaatikosta
 - Sosiaalisen median sovellus pollaa uudet viestivirtaan tulleet viestit some-kanavalta (palvelimelta)
 - Uutisovellukset pollaavat uudet uutiset uutislähteistä (palvelimilta)
- Jatkuvakaan pollaus ei vie kenenkään resursseja:
 - Yhden pollauksen tiedonsiirtomäärä molempiin suuntiin on vain muutamia satoja tavuja

Jatkuva pollaus käytännössä

- Alla oleva Python-koodiluonnos pollaa jakelupaketin valmistumista minuutin välein niin kauan, kunnes jakelupaketti on valmis.

```
while True:                                     # toista jatkuvasti

    response = requests.get('https://pas.csc.fi/api/2.0/<contract-id>/disseminated/<dip-id>')
    # kysy valmistumista, vastaus menee suoraan response-muuttujaan

    if response.status_code == 200 \
        and response.json()['data']['complete'] == 'true': # Jos kysely onnistui (palautti OK) ja...
        break                                             # ...DIP on vastauksen mukaan valmis...
                                                         # ...poistu ikiluupista heti (keskeytä toisto)

    time.sleep(60)                                       # muutoin odota minuutti ja kysy uudelleen
```

DEPIP: Viestien sisältö

- SOAP XML:n painolasti näkyy elementtiluettelossa vahvasti
 - Standardin elementtilistaus on tarpeettoman monimutkainen, rakenteellinen ja yksityiskohtainen ollakseen toteutusriippumaton “conceptual standard”
 - Elementtiluettelossa on jopa tyyppimäärittelyt: DateTimeType, Base64Type, xsd:decimal...
 - Esim. Base64-koodaus joillekin viestin osille tulee SOAP XML –rajapinnan perintönä
 - Standardin tulisi selvemmin todeta tyyppimäärittelyiden yms. teknisiin toteutuksiin liittyvien asioiden olevan vain esimerkkejä
- PAS-rajapintojen vastaavissa toiminnoissa on tavalla tai toisella olemassa DEPIP-standardin luetteloimat tietoelementit.
 - Osa hoituu esim. HTTPS-protokollassa automaattisesti

DEPIP: Viestien sisältö

- Suomi esitti viestien sisältöön useitakin asioita, mm. että sellaisia pakollisia elementtejä, joita käytetyssä siirtoprotokollassa on valmiina, pitäisi pystyä hyödyntämään “DEPIP-elementin” sijasta.
 - Esimerkki: Jokaisessa viestissä on pakollista esitellä keskustelun osapuolet, mutta istuntokerroksessa tämä tieto on jo valmiiksi olemassa keskustelun molemmilla osapuolilla.
 - Esimerkki: HTTP-otsakkeessa on viestin aikaleima, mutta ei käy selvästi ilmi se, edellyttääkö DEPIP-protokolla toisenlaista tietotyyppiä
 - Esimerkki: Kysely- ja vastausviesteissä on pakollisena erikseen määritellyt tunnustekentät, joilla kysely- ja vastausviestit voidaan yhdistää toisiinsa, vaikka HTTP-protokolla tekee tämän täysin automaattisesti.
- Suomen esitys tältä osin hyväksyttiin ja asiasta lisättiin dokumentin välivaihe-luonnokseen huomio.

DEPIP: Viestien sisältö



- Final draft –versiosta huomio oli poistettu, silti sähköpostinvaihdossa komitea vakuuttaa esimerkkeineen, että standardissa on kyse käsitteistä eikä syntaksista.
 - Periaatteessa viestityksen voi tehdä vaikka sähköpostitse, eikä sen tarvitse sisältää XML-rakenteita tai muita vastaavia teknisiä asioita. Käytännössä suomen kielikin riittäisi, kunhan se tavalla tai toisella sisältää vastaavat asiat, mitä DEPIP-standardi luettelee.
- Tämä on meille tärkeää, koska ei ole mitään järkeä vaatia siirtoprotokollan valmiiden, käytettyjen ja testattujen toiminnallisuuksien päälle samoja toiminnallisuuksia uudelleen.
- Tårta-på-tårtaa ei voi perustella edes sillä, että ko. tietoelementeistä jäisi parempi jälki, koska HTTPS-protokolla tallentaa jäljet automaattisesti myös silloin, kun varsinaisissa viesteissä ei noudateta yhtään mitään määrittystä.
 - Esimerkki: Jos käyttäjätunnus kaapataan, ja sieltä tulevat viestit sen jälkeen sisältävät pelkkiä murtautumisyriksiä, niin pitää edelleen tallentaa oikein lähettäjä, aikaleimat, sisällöt ja vastaavat tiedot mahdollisimman laajalti ja oikein.

Laajennettavuus

- Teknisesti PAS-palveluiden REST-rajapintoja voidaan hyvin laajentaa niin, että se toteuttaa kaikki ne toiminnot, mitä DEPIP-standardi luettelee
 - Lisäksi voidaan laajentaa myös (lähes) mitä tahansa muita tarvittavia toimintoja
 - Yleinen käytännön tarve sanelee: Jos ei tarvita jotakin toimintoa, sitä ei myöskään rajapintaan toteuteta resurssisyistä. Varsinaista teknistä estettä ei kuitenkaan ole.
- Laajennusten lisääminen DEPIP-standardiin on sen sijaan työlästä
 - ISO-standardin hallinnollinen päivityssykli on merkittäväällä tavalla hidastava tekijä
 - Varmaankin tämän vuoksi ICT-alalla vakiintuneita asioita ei tänä päivänä tavallisesti viedä mihinkään erillisiin standardointiprosesseihin, vaan yleensä näihin sitoudutaan alalla toteamalla ne maailmanlaajuisiksi *de facto* –standardeiksi. Tällaiset standardit voivat alalla kestää aikaa useita vuosikymmeniä. Päivityksiä toki tehdään, mutta yhteensopivasti mitään rikkomatta.

PAS-rajapinnan ominaisuudet, joita DEPIP ei määrittele

- DEPIP ei määrittele (mutta ei myöskään estä) seuraavia toimintoja, joita PAS-palvelun rajapinta tukee tai sen on suunniteltu tulevaisuudessa tukevan:
 - Hakuehdoilla hakeminen (“search”)
 - Aineiston päivitys hyödyntävän organisaation toimesta (SFTP, tulevaisuudessa)
 - Tilastot (tulevaisuudessa)
 - Jakelupaketin luontia ja noutoa ei selkeästi eroteta toisistaan
 - Rooleista ei käydä pohdintaa, erityisesti asiakas-palvelin-rajapintojen suhteen

DEPIP-toiminnot PAS-palveluissa



- DEPIP-rajapinnan toiminnallisuudet toteutuvat PAS-palvelussa seuraavalla tavalla:
 - Transfer request ja Transfer request reply: Luvan (aikaikkunan) pyytäminen aineiston siirtämiseksi PAS-palveluun
 - Toistaiseksi ei ole nähty tarvetta toteuttaa, koska PAS-palveluun voi siirtää aineistoa milloin tahansa.
 - Transfer ja Transfer reply: Aineiston siirto PAS-palveluun
 - PAS-palvelussa on käsitteellisesti vastaava toiminto SFTP-rajapinnassa
 - Lisäksi REST:iin lisätään tulevaisuudessa mahdollisuus noutaa vastaanoton validointiraportit
 - Delivery request ja Delivery request reply: Aineiston nouto PAS-palvelusta
 - PAS-palvelussa on käsitteellisesti vastaava toiminto REST-rajapinnassa
 - DEPIP ei selkeällä tavalla erota toisistaan jakelupaketin muodostusta ja siirtoa
 - DEPIP:in toiminnallisuudessa URI:n palauttaminen (base64-enkoodatun aineiston sijaan) on käytännössä täsmälleen sama asia, kuin PAS-palvelussa jakelupaketin muodostus
 - DEPIP ei määrittele, kuinka palautetun URI:n avulla saadaan noudettua varsinainen aineisto.
 - Sen sijaan PAS-rajapintamäärittely vie noutoprosessin loppuun asti.

DEPIP-toiminnot PAS-palveluissa



- Modification notification: PAS-palvelun ilmoitus hyödyntävälle organisaatiolle (PAS-palvelun tekemästä) aineiston muuttamisesta
 - Ei erillistä toteutusta, koska ei ole hallinnollista mallia, mutta teknisesti PAS-palvelun hakutoiminnoilla voidaan noutaa luettelo annetulla aikavälillä päivittyneistä aineistoista
- Disposal notification: PAS-palvelun ilmoitus hyödyntävälle organisaatiolle aineiston poistamisesta säilytyksestä
 - Ei erillistä toteutusta, koska ei ole hallinnollista mallia, mutta käänteisesti PAS-palvelun hakutoiminnolla voidaan noutaa luettelo säilytyksessä olevista aineistoista
- Restitution request ja Restitution request reply: Aineiston ja säilytysvastuun siirto PAS-palvelulta takaisin hyödyntävälle organisaatiolle
 - Ei toteutusta, koska ei ole hallinnollista mallia.
 - Keskustelu voi DEPIP-mallissa alkaa joko hyödyntävältä organisaatiolta PAS-palveluun tai toisinpäin.

DEPIP-toiminnot PAS-palveluissa

- Autorisoinnit: DEPIP:ssä on oma autorisointimallinsa jakelulle ja poistolle
 - Vaatimukset turvallisuudelle velvoittavat PAS-palveluja käyttämään laajalti tunnettuja, testattuja ja käytettyjä menetelmiä.
 - Tästä on olemassa standardointikomitean kanssa käyty erillinen sähköpostikeskustelu, jonka mukaan järjestelmä täyttää DEPIP-standardin kaikki vaatimukset silloinkin, vaikka se käyttäisi mitä tahansa muita autorisointimalleja DEPIP-standardissa esitetyn mallin sijasta.
 - Ei ole ihan selvää, missä määrin DEPIP tässä kohtaa mahdollisesti vuotaa istuntokerroksesta esitystapakerrokseen ISO/IEC 7498-1:1994-standardin (OSI-viitemallin) näkökulmasta.

Yhteenveto



- DEPIP luetteloi kyllä elementit PAS-toiminnan perusviestinnälle, mutta erityisesti siinä ontuu osapuolten roolien jako sekä itse tietoelementtien esitystapa
- DEPIP ei tuo sellaisia tietoelementtiä lisää, joita vastaavissa jo toteutetuissa PAS-palveluiden toiminnoissa ei olisi jo valmiiksi olemassa
- Kun tulkitaan DEPIP-standardia puhtaasti käsitteellisesti ja että standardi on toteutusriippumaton (näin standardointikomitea itse halusi sen nähtävän), niin PAS-palveluiden rajapinnat nykyisellään toteuttavat DEPIP-standardin minimivaatimukset.
 - Erot ovat toteutuspuolella syntaksissa:
 - “dissemination” (OAIS, PAS-palvelut) vs. “delivery” (DEPIP)
 - HTTPS-protokolla vs. tårta-på-tårta
 - Jotkut elementit ovat paketoitumäärityksessä eri muodossa
 - ...



pas-support@csc.fi

<http://digitalpreservation.fi>

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

@dpres_fi