



08.11.2013

Valtiovarainministeriö
PL 28
00023 VALTIONEUVOSTO

Viite VM119:00/2012, 1873/00.01.00.01/2012

Asia **OKM lausunto: Perustietovarantojen viitearkkitehtuurin luonnos**

Opetus- ja kulttuuriministeriö esittää lausuntonaan perustietovarantojen viitearkkitehtuuria koskevasta luonnoksesta seuraavan.

Opetus- ja kulttuuriministeriö on sisällyttänyt lausuntoonsa hallinnonalansa organisaatioiden lausunnoissa esitettyjä näkemyksiä.

0 YLEISTÄ

Viitearkkitehtuurin hyödynnettävyys, raportin rakenne, käytetty käsitteistö, esitystapa

Tekstissä todetaan, että perustietovarantojen viitearkkitehtuuri on osa julkisen hallinnon kokonaisarkkitehtuuria sekä siihen kuuluvaa Julkisen hallinnon tietoarkkitehtuuri -kuvausta. Tässä roolissa arkkitehtuuri on opetus- ja kulttuuriministeriön kannalta tärkeä ja sen ristiriidattomuuteen tulee panostaa.

Viitearkkitehtuurin tavoitteena on luoda yhteinen rakennemalli sille, miten julkisen hallinnon perustietovarannoissa sijaitsevia tietoja hyödynnetään eri organisaatioiden prosesseissa ja tietojärjestelmissä. Viitearkkitehtuuri muodostaa opetus- ja kulttuuriministeriön kannalta keskeisen lähtökohdan opetus- ja kulttuuriministeriön vastuulla olevan kohdealueen tietoarkkitehtuurin rakentamisessa ja kehittämisessä. Näiden tietovarantojen sisältöpalveluja tarvitaan opetus- ja kulttuuriministeriön vastuulla olevan kohdealueen toiminnoissa ja prosesseissa, joten viitearkkitehtuurin määrittelyjen selkeys ja arkkitehtuuriin pohjautuvien ratkaisujen luotettavuus on tärkeää.

Viitearkkitehtuurissa on tunnistettu viisi erilaista sidosryhmää: 1) tietovastuullinen, 2) tiedon tuottaja, 3) tiedon jatkojalostaja, 4) tiedon hyödyntäjä ja 5) rekisteröity/omien tietojensa ylläpitäjä. Dokumentissa keskitytään lähes kokonaan sidosryhmiin 1, 2 ja 4, mutta perustietovarantojen hyödyntämisen kannalta olennaisen tärkeitä ovat 3 ja 5, jotka tulisi riittävällä tavalla huomioida suunnitelmissa.

Tämänhetkisessä viitearkkitehtuurin kuvauksessa on sisäisiä ristiriitoja ja epäselvyyksiä, jotka osin johtuvat muiden viite- ym. arkkitehtuurien määrittelyjen tulkinnanvaraisuuksista. On tärkeää saada nämä ongelmat ratkaistua ennen viitearkkitehtuurin julkaisemista.

Viitearkkitehtuurista ei käy aukottomasti ilmi koskevatko määritykset muun tyyppisiä kuin yleisiä perustietovarantoja (esim. kohdealueen informaatiovarannot, strategisen päätöksenteon informaatiovarannot). Kansallisen metatietopalvelun rooli ei aukene dokumentista.

On mainio idea, että arkkitehtuurin kehittämisen osa-alueet on tässä viitearkkitehtuurissa puettu ikään kuin toiminnan muotoon korostaen sen prosessimaisuutta. Samalla voisi kuitenkin alleviivata, että arkkitehtuurin kehittäminen on myös iteratiivinen prosessi ja sitä voidaan tehdä muussakin järjestyksessä kuin dokumentin määrittelemässä. Arkkitehtuurikuvausten tarkkuustaso riippuu kulloinkin kyseessä olevasta organisaatiosta ja sen arkkitehtuurikyvykkyyden kypsyystasosta. Viitearkkitehtuurista puuttuvatkin minimikuvausvaatimukset, jotka auttaisivat priorisoimaan tehtäväkokonaisuuksia ja niiden merkitystä ulkoisen ja sisäisen kehittämisen kannalta.

1 JOHDANTO

2 KÄSITTEISTÖ

Tietovaranto;

Tutkimuksen tietoaaineistot, joihin tässä dokumentissa viitataan käsitteellä tietovaranto, ovat hajautettu asiakokonaisuus, joihin voi liittyä useampi kuin yksi tietoja hallinnoiva toimija.

Rekisteri;

Tutkimuksen tietoaaineistojen metadatakatalogit ovat tämän määritelmän mukaisesti rekistereitä.

Ne eivät ole kuitenkaan henkilötietorekistereitä termin tavanomaisessa merkityksessä.

3 PERUSTIETOVARANTOJEN KÄSIKIRJA

1 Huomioi ja kuvaa yhteentoimivuus, sidosarkkitehtuurit ja muut ohjaavat tekijät

Tekstissä todetaan, että perustietovarantojen viitearkkitehtuuri on osa julkisen hallinnon kokonaisarkkitehtuuria ja sivuaa Julkisen hallinnon tietoarkkitehtuuria ja useita muita viitearkkitehtuureja. Sivun 9 kuvassa on hahmoteltu Perustietovarantojen viitearkkitehtuurin rakennetta, sen sisältämiä elementtejä sekä suhdetta muihin julkisen hallinnon arkkitehtuurikuvauksiin. Tämän tason kuvauksen tulisi olla erityisen selkeä ja mahdolliset ristiriidat ja tulkinnanvaraisuudet tulisi dokumentoida sekä ottaa käsittelyyn kokonaisvaltaisesti.

Tiedon tuottamis- ja hyödyntämisstrategiat tulisi viitearkkitehtuurissa huomioida myös toiminnan ja teknologian tasoilla.

2 Huomioi julkisen hallinnon arkkitehtuuriperiaatteet

Tutkimuksen tietoaaineistojen kohdealueella on kehitteillä oma erillinen vain näitä tietoaaineistoja koskeva metatietopalvelunsa

3 Tunnista perustietovarantoon liittyvät sidosryhmät ja roolit

Sivulla 11 olevassa kuvassa 4 on käsitteistön osalta ristiriitoja mm. sivulla 27 olevaan kuvaan 16 verrattuna. Sidosryhmän käsitteen tulisi olla yhtenäinen ja selkeästi kuvattu koko dokumentin tasolla. Erityisesti sidosryhmien ”Rekisteröity” ja ”Omien tietojensa ylläpitäjä” määrittely herättää kysymyksiä verrattuna sivun 27 kuvaukseen.

Kuvassa 5 viitataan perustietovarantojen sidosryhmiin ja todetaan, että kuva on esimerkki eikä

tietosisällöltään kattava. Olisi hyvä, että tietosisällöltään kattava esitys olisi tarjolla, tai vähintäänkin olisi kerrottu, miltä osin kuva ei ole tietosisällöltään kattava.

Viitearkkitehtuurin rajaukset perustietovarantojen osalta ovat geneerisiä, mutta opetus- ja kulttuuriministeriön toimialan kannalta olisi olennaista pystyä arvioimaan, mitkä opetus- ja kulttuuriministeriön tälle hetkellä käyttämistä tietovarannoista (osa sellaisia jotka ostetaan käyttöön niiden ylläpidosta vastuussa olevilta tahoilta) tulevat sisältymään perustietovarantojen määrittelyyn.

Viitearkkitehtuurissa sidosryhmiin kuuluvat tiedon jatkojalostajat voivat tuottaa muutakin kuin palveluita, esim. uutta tutkimustietoa tai strategisen päätöksenteon tietoa.

Sivulla 12 mainitaan:

”Tavoitteena on, että aika tiedon synnystä sen käyttöönottoon on mahdollisimman lyhyt. Tämän seurauksena tiedon tuottajien rooli korostuu ja on mahdollista, että tästä syystä tietojen ylläpidossa siirrytään soveltuvien osien entistä enemmän hajautettuun ylläpitoon. Yhtenä keinona tavoitteen saavuttamisessa on, että perustietovarantojen tietoa ylläpidetään niiden sidosryhmien toimesta, joiden luonnollisten prosessien osana tietoa syntyy. Tiedon ylläpito integroituu osaksi näiden sidosryhmien omia toimintaprosesseja.”

Käytännössä tavoite tukisi sitä, että mm. väestötietojärjestelmän päivitysmahdollisuus laajenisi suuremmalle viranomaisjoukolle kuin vain verohallinnolle. Voitaisiin nähdä, että korkeakoululaitos ulkomaalaisten keskeisenä ensimmäisenä suomalaisena käsittelijänä toimisi väestötietojärjestelmään tallentajana ja henkilötunnuksen myöntäjänä.

4 Kuvaa perustietovarantoon liittyvät prosessit

Sivulla 16 mainitaan tiedon tuottajan ja hyödyntäjän kohdalla toiminto ”palautteen anto virhetilanteessa”, mutta jää epäselväksi missä määrin ja miten tiedon rikastamisen prosessi tulisi toimimaan osana perustietovarantojen hyödyntämistä.

Tutkimuksen tietoaaineistojen näkökulmasta tiedon julkaiseminen on tiedon tuottajan vastaama prosessi.

5 Tunnista tarkasteltavan tietovarannon sijainti julkisen hallinnon tietovarantojen ekosysteemissä

Kuvassa 7 viitataan julkisen hallinnon tietoarkkitehtuuriin ja todetaan, että siellä käytetyn kolmen tason sijasta viitearkkitehtuurissa jaetaan tietovarannot neljään luokkaan. Perustelu tälle jää epäselväksi. Lisäksi taulukon 7 sisältämät termit (kuten esimerkkinä paikkatieto) eivät lähtökohtaisesti kuvaa sitä, minkä tason/luokan tiedosta on kyse, sillä paikkatieto voi hyvin olla osa strategisen päätöksenteon tietovarantoa. Taulukon 7 informaatio tässä yhteydessä enemmänkin herättää kysymyksiä kuin antaa vastauksia.

Kuva 7;

Tutkimustieto voi toimia myös päätöksen teon tukena (tyyppiä tieto/tietämys, esim. tutkimusjulkaisu).

Sivulla 17 puhutaan sekä perustietovarannoista että perusdatavarannoista, jotka ovat ilmeisesti eri asioita, mutta tämän kaltaiset hyvin lähellä toisiaan olevat termit tuottavat merkittäviä sekaannuksen mahdollisuuksia. Käsitteistössä tulisi pyrkiä yksikäsitteisyyteen ja selkeyteen.

6 Päättää perustietovarantojen tiedon tuottamis- ja hyödyntämismalleista

Kuvassa 9 määritellyt tiedon hyödyntämismallit ovat mielenkiintoinen kehys tietojen hyödyntämisen hahmottamiseen. Tätä kehystä ei kuitenkaan systemaattisesti ole hyödynnetty dokumentissa vaan tämä jää melko irralliseksi osaksi kokonaisuutta. Lisäksi terminologia

vaatii täsmennystä, esimerkkinä lause ”ei tarkoiteta maantieteellistä hajauttamista eri organisaatioiden kesken, vaan hajauttamista eri tietokantoihin”. Tietokanta on tässä yhteydessä rajoittava ja harhaanjohtava käsite, pitäisi pikemminkin pyrkiä näkemään tietovaranto yhtenäisenä toiminnallisena kokonaisuutena eikä tietokantaan tms. rinnastettavana asiana.

Kappale 6.2: integroitu / tietovastuullisen näkökulma;

Prosessi ei välttämättä tässäkin tapauksessa jakaudu useamman toimijan kesken

7 Määrittele keskeiset käsitteet ja kuvaa tietovarannon päätietyryhmät ja muu sisältö

Metatietojen osuutta on viitearkkitehtuurissa käsitelty varsin ohuesti: Lauseilla "Tietovastuullisen vastuulla on perustietovarannon tietosisällön kuvaaminen. Tiedon yhteiskäyttöisyys edellyttää kattavia tiedon määrittelyä sekä sisältökuvauksien ja rakennemäärittelyjen laatimista." määritetään tietovastuullisen vastuuta ja velvollisuuksia yleisellä tasolla, mutta jätetään tarkemmin ohjeistamatta miten nämä tehdään. Tarkempi ohjeistus metatietopalvelusta on ilmeisesti tekeillä (s.15 ”VM Luonnos Perustietovarantojen viitearkkitehtuuri).

Metatietojen dokumentointia, ylläpitoa ja saatavuutta pitäisi konkretisoida pidemmälle, eli avata tarkemmin esimerkiksi seuraavilla tavoilla:

- 1) mitä kattavilla tiedon määrittelyillä ja sisältökuvauksilla tarkoitetaan
- 2) missä määrin perustietovarantojen yhteneväisyys edellyttää yhteistyötä eri tietovastuullisten/organisaatioiden välillä
- 3) miten tätä yhteistyötä on tarkoitus vastuuttaa sekä koordinoita.

Miksi koulutustoimijan tiedot ovat päätietyryhmänä? Suhteessa luettelon muihin toimijoihin ne tuntuvat selvästi muista eroavalta, vai onko kyse esimerkkiedosta

8 Suunnittele tietojärjestelmäarkkitehtuurin ylätasoon looginen jäsenitys

Kuva 12;

Kuvan voi tulkita siten, että perusdatavarannoista ja yleisistä informaatiovarannoista ei kerätä metatietoja yhteiseen metatietopalveluun, mikä kuitenkin yhteentoimivuuden näkökulmasta (dokumentin keskeinen tema) saattaisi olla tarpeen.

Kuva 13;

Julkaisu voi toimia myös Pull-moodissa (esim. metatietojen haravointi).

Kuva 15;

Atomaarinen palvelu viitanee kuvassa SOA arkkitehtuurin atomic service käsitteeseen.

Parempi suomennos voisi olla atominen (tai ydin) palvelu. Tämän ja koosteisen palvelun voisi myös määritellä tarkemmin kappaleessa käsitteet.

Kuva 15;

Koosteiset palvelut on kuvassa sijoitettu kansalliseen palveluväylään kuuluvaksi, mutta jäljempänä kohdassa 12.3 on maininta, ettei kansallisen tason prosessimootoria olisi suunnitteilla, jota ne kuitenkin edellyttänevät (samoin yhden luukun periaate, kappale 10.1)

Sivulla 25 viitataan kansallisen palveluväylän viitearkkitehtuuriin ja linjataan että toteutettava ratkaisu ei saa olla sen kanssa ristiriitainen. Ristiriitoja voi esiintyä myös viitearkkitehtuurien tasolla ja tulisi hahmottaa, miten perustietovarantojen viitearkkitehtuurin ja kansallisen palveluväylän viitearkkitehtuurin mahdolliset ristiriidat tullaan selvittämään ja korjaamaan. Perustietovarantoja hyödyntävien organisaatioiden kannalta nämä kaksi viitearkkitehtuuria ovat keskeisessä roolissa suunniteltaessa kohdealuekohtaisia arkkitehtuureja.

9 Huomioi järjestelmäpalveluihin liittyvät vastuut sidosryhmittäin
Kuvassa 16 on ristiriitoja suhteessa kuvaan 4 sivulla 11.
10 Huomioi perustietovarantojen palveluarkkitehtuuri
Sivulla 28 todetaan: ”Sen lisäksi, että perustiedot ovat saatavissa helposti yksittäin, pitäisi niitä myös pystyä yhdistelemään käyttäjien kannalta mielekkäiksi kokonaisuuksiksi. Sisältöpalvelukokonaisuuksien koostamisessa ”Tiedon jatkojalostaja” -rooli nousee avainasemaan.” Tämän havainnon tulisi näkyä selkeästi dokumentin kaikissa osissa, tällä hetkellä asia on piilossa ja pitkälti ilman pohdintaa siitä, mitä käytännössä tarkoittaa tiedon jatkojalostamisen roolin korostaminen. Esimerkiksi sivulla 16 mainitaan ”palautteen anto virhetilanteessa” mutta ei hahmoteta tiedon jatkojalostamisen ketjua ja tarvittavia mekanismeja tiedon rikastamisessa. Yhden luukun periaatetta noudatettaessa kertarekisteröityminen tulisi tehdä mahdolliseksi myös tutkimuksen toimialueeseen kuuluville henkilöille toimialueella käytetyn HAKA-tunnistautumisen kautta.
11 Huomioi sähköisen asioinnin viitearkkitehtuuri käyttöliittymäkerroksessa
Ovatko käyttöliittymäkerros ja siten SAVI relevantteja, kun puhutaan tietovarannoista? Olisivatko ne mahdollisen monoliittin ohjaavan vaikutuksen vuoksi syytä jättää tämän viitearkkitehtuurin ulkopuolelle ja lähteä siitä, että käyttöliittymät ja niiden sovellukset tehdään muiden ohjaavien arkkitehtuureiden puitteissa?
12 Suunnittele teknologia-arkkitehtuuri
Tunnistamisen yhteydessä viitataan VETUMA, KATSO, VIRTU ja HAKA – tunnistuspalveluihin. Missä määrin palveluväylän yhteydessä pohdittu sähköinen identiteetti kytketty tähän kohtaan? Tunnistamisen ja käyttöoikeushallinnan osana, esimerkiksi federoitaessa identiteettejä, on otettava huomioon tarve eri tunnistusten ja identiteettien yksilöimiselle samaan luonnolliseen henkilöön. Tässä henkilötunnus on heikko tapa yksilöintiin, koska se ei ole pysyvä ja toisaalta se sisältää henkilötietoja (ikä, sukupuoli). Olisi syytä ratkaista tapa pysyväälle yksilöintiavaimelle eri tunnistamistapojen ja identiteettien avaimia yhdistävänä yksinomaan teknisen tason yleisavaimena. On syytä ratkaista tunnistamisen ja käyttöoikeushallinnan osana yksilöinti myös väestötietojärjestelmän ulkopuolisten henkilöiden osalta, kuten korkeakoulujärjestelmään tulevat ja siinä vain lyhyesti vierailevat ulkomaalaiset. Mm. SAVI lähtee käytännössä yksinomaan VETUMA:n pohjalta ja on siten rajautunut niiden tunnistamiseen, joilla on henkilötunnus. Tässä esimerkiksi väestötietojärjestelmän ylläpidon laajentaminen korkeakoulujen tehtäväksi auttaisi osin, kun henkilötunnus voitaisiin luoda jo silloin, kun henkilö alkaa asioida korkeakoulujärjestelmän kanssa, joka on usein ensimmäinen Suomen julkisen hallinnon kontaktipiste ulkomaalaiselle. Edelleen yksilöintiavaimena tarvittaisiin joka tapauksessa henkilötunnuksesta riippumaton avain. Ratkaisutapana voisi olla esim. Opetushallituksen OID-standardiin pohjautuvan Oppijanumeron käyttötarkoituksen ja toteutuksen laajentaminen. Sivulla 34 olevan taulukon yllä olevassa tekstissä todetaan: ” Alla olevaan taulukkoon on koottu PERA-määrityksen rajapintatoteutusten tekniset linjaukset. Lisäksi taulukossa on joitakin linjauksia, jotka eivät ole PERA-määrityksen linjauksia.”

Dokumentin lukijan kannalta jää epäselväksi, mitkä linjaukset ovat PERA-määrityksen linjauksia ja mitkä eivät ole.
13 Sisältöpalveluiden ratkaisuvaihtoehtoja
Kopioitu tietovaranto on Korkeakoulujen valtakunnallisen tietovarannon toteutuksessa valittu ja hyväksi todettu ratkaisumalli hajanaisen ja osin vanhentuneen lähdejärjestelmäekosysteemin nostamiseksi toimivien rajapintojen taakse.
LIITTEET
LIITE 1 Ohjaavat tekijät
1. Perustietovarantojen viitearkkitehtuuri
1.1 Yhteentoimivuus ja julkisen hallinnon arkkitehtuurit
Ohjaaviin tekijöihin tulisi huomioida muutkin julkisen hallinnon standardisalkussa olevat määritykset.
1.2 Julkisen hallinnon kokonaisarkkitehtuuri
1.3 Julkisen hallinnon tietoarkkitehtuuri
1.4 Perustietovarantojen viitearkkitehtuurin rakenne ja elementit
2. Ohjaavat periaatteet
2.1 Lainsäädäntö
2.2 Strategiat, ohjelmat ja tietopolitiikka
2.3 Standardit ja suositukset
2.4 Julkisen hallinnon kokonaisarkkitehtuuriperiaatteet
Tämä on tiivis ja havainnollinen luku julkisen hallinnon arkkitehtuuriperiaatteiden vaikutuksista perustietovarantojen viitearkkitehtuuriin. Lopun perustietovarantojen näkökulmasta esitetyt huomioitavat seikat muodostavat konkreettisen tarkistuslistan.
LIITE 2 Loogiset perustietovarannot
1. Tietovarantojen jäsenyys
2. Perustietovarantojen viitearkkitehtuurin suhde loogisiin tietovarantoihin
”Henkilötietovaranto (väestötiedot) - Suomen kansalaisten ja Suomessa vakinaisesti asuvien ulkomaalaisten perustiedot” Korkeakoululaitoksen näkökulmasta henkilöitä täytyy tietovarantona hallita jo ennen kuin he vakinaisesti asuvat Suomessa. ”Todennetun osaamisen rekisteri, TOR: - - Käyttöönotto on vuonna 2014” Tässä lienee vanhentunutta tietoa. Kansallisen palveluväylän viitearkkitehtuurin luonnoksessa 0.95 on tieto, että käyttöönotto käynnistyy 2015.

LIITE 3 Tietojärjestelmäarkkitehtuurin looginen jäsenitys ja integraatioarkkitehtuuri
1. Tietojärjestelmäarkkitehtuurin looginen jäsenitys
1.1 Tietovarantojen tarjoaminen ulkoisille loppukäyttäjille
1.2 Perustietovarantorajapintojen julkaisu palveluväylässä
1.3 Perustietovarantojen hyödyntäminen
2. Integraatioarkkitehtuuri

Ylijohtaja

Håkan Mattlin

Tietohallintojohtaja

Irma Nieminen

Tiedoksi

Koulutuspolitiikan osasto
Korkeakoulu- ja tiedepolitiikan osasto
Aikuiskoulutuspolitiikan yksikkö
Kulttuuri-, liikunta- ja nuorisopolitiikan osasto
OKM Hallinnonalan virastot
OKM Arkisto