

Luonnonvarakeskuksen lausunto - Tutkimuksen datanhallinnan viitearkkitehtuurin lausuntokierros 7.10.-3.11.2024

Lausunnon valmistelijat: Teppo Jylänki, Anu Kantola, Nina-Mari Salminen, Harri Pitkänen

Koska lausunto pyydetään

(<https://wiki.eduuni.fi/display/CscKtpoYtf/Tutkimuksen+datanhallinnan+viitearkkitehtuurin+lausuntokierros+7.10.-3.11.2024>) jättämään kommentteina Wiki-alustalle, tämä asiakirja sisältää yhteenvedon Luonnonvarakeskuksen nimissä jätettävistä kommenteista.

Johdanto

Tässä työssä on ymmärretty kokonaisarkkitehtuurin merkitys datan hallinnassa. Työ jäsentää organisaatiotason datanhallinnan järjestämistä tukien datan käytettävyyttä uusissa asiayhteyksissä. Työ on vaatinut eri organisaatioiden asiantuntijuutta, jotta lopputulos on käyttäjiä palveleva.

Olisi hyvä, että saisimme lyhyen, erilaisille käyttäjäryhmille (johto, päättäjät, datamanagerit, digi asiantuntijat, tutkijat jne.) suunnatun selkokielisen yhteenvedon siitä mistä tässä uudistuksessa on kysymys, miten tämä uudistus vaikuttaa kunkin käyttäjäryhmän työskentelyyn jatkossa, miten tämä muuttaa koko organisaation toimintaa liittyen datan keräämiseen, dokumentointiin, säilytykseen ja uudelleenkäyttöön, millä aikataululla uudistus jalkautetaan. Yhteenvedossa voisi olla viittaukset tarkempiin asiakohtiin, mistä voisi löytää tarkempaa määrittelyä halutessa. Samalla selkokielisen yhteenvedon lukija ymmärtäisi, miltä muutos arkkitehtuurinotaatiolla toteutettuna näyttää.

Voisiko arkkitehtuurin julkaista myös ArchiMate Model Exchange –muodossa, jos se onnistuu helposti? Kommentoitavassa materiaalissa olevat kuvat on piirretty kyseistä notaatiota hyödyntäen, joten oletettavasti taustalla on työkalu, joka pystyisi tuottamaan tuollaisen tiedoston. Tämä parantaisi yhteentoimivuutta, kun viitearkkitehtuurin osia voisi hyödyntää organisaation omassa arkkitehtuurityössä. Lisäksi avoimen lähdekoodin Archi-työkalu voi tehdä mallista selailtavan HTML-version.

Arkkitehtuurissa tunnistetaan paljon tavoitteita ja kehittämisvaatimuksia. Varmastikaan nämä kaikki eivät ole keskenään samanarvoisia. Joitakin on jo eroteltu keskeisiin vs. ei-keskeisiin, mutta tämän jälkeenkin vaatimuksia jää niin paljon, että on vaikea nähdä, mistä kannattaisi lähteä liikkeelle. Monet kehittämisvaatimukset näyttävät kietoutuvan kokonaisuudeksi koneluetun datanhallintasuunnitelman ympärille. Löytyykö tähän liittymättömistä kehittämisvaatimuksista tai strategisista tavoitteista joitain merkittäviä "matalalla roikkuvia hedelmiä", joihin voisi ja kannattaisi tarttua ja joita voisi edistää ilman massiivista kehittämisprojektia?

Kannatamme Mari Elisa Kuusniemen kommenttia: "Tutkimuksen digitaalisessa transformaatioissa on tärkeää myös vanhempien paperiaineistojen digitointi. Vaikka viitearkkitehtuuri ei käsittele paperiaineistoja, voisiko digitointi- ja OCR-palvelut lisätä arkkitehtuuriin? Nämä palvelut voisivat linkittyä vaiheeseen, jossa data kerätään vai muunnetaan muotoon, jossa sitä voidaan analysoida."

Kansallisarkisto tuottaa osin samankaltaisia tai jopa samoja palveluita kuin CSC tutkimuksen tarpeeseen, ja tarjoaa palveluitaan myös tutkimukselle. Pitäisikö tämän näkyä arkkitehtuurissa jotenkin? Nyt näiden organisaatioiden palvelutarjonnassa on päällekkäisyyksiä (mm. pitkäaikaissäilytys).

Periaatteellinen arkkitehtuuri

2.1 Tutkimuksen datanhallinnan visio, toiminnallinen tavoitetila ja strategiset tavoitteet 2030

Strategisia tavoitteita on paljon ja ne ovat melko laajoja. Ovatko kaikki yhtä tärkeitä, vai voidaanko tunnistaa joitakin, joihin olisi erityisesti kiinnitettävä huomiota? Eli voisiko arkkitehtuuriin tuoda rohkeampaa näkemystä, jolla nostettaisiin 3-5 tavoitetta keskeisimmiksi vaikkapa sillä perusteella, että ne toistuvat useissa ohjaavissa strategioissa tai että ruohonjuuritasolla tavoitteiden heikko toteutuminen näkyy vahvasti käytännön työssä? Samaa priorisointia voisi peilata myös kehittämisvaatimuksiin.

Typo: 2.1.2 Toiminnallinen tavoitetila 2030

Toiminnallinen tavoitetila tarkoittaa Visiota. Tavoitetila **saavutetaan saavutettavissa** suunnitellulla toiminnalla, jota alla esitetyt strategiset tavoitteet ja tulokset kuvaavat. Tuplana kuvassa: Säädöspohja vahvistaa tutkimusdatan uudelleenkäyttöä ja saatavuutta

2.2 Arkkitehtuuriperiaatteet

"Arkkitehtuuri on yksinkertainen ja ymmärrettävä" - yksinkertaisuus ei tällä hetkellä toteudu

2.6 Kehittämisvaatimukset

Keskeiseksi merkittäviä kehittämisvaatimuksia on 43 kappaletta, ja osa niistä on hyvin yksityiskohtiin meneviä. Pitäisi vielä paremmin löytää aidosti keskeisimmät tavoitteet.

SAATAVUUS-7: onko vaatimus monikielisydestä tarpeellinen? Oleellisempaa lienee käyttää työaikaa yhden hyvän aineistokuvauksen tuottamiseen kuin ylläpitää kuvausta usealla eri kielellä. Tekoäly on käytännössä jo pitkälti ratkaissut käännösongelman, ja tulevaisuudessa voi olla, että tekoälyn tekemät käännösvirheet ovat pienempi ongelma kuin manuaalisesti ylläpidettyjen kieliversioiden eriytyminen vahingossa.

PALVELU-31: vaatimus on ylimitoitettu. Voisiko ennemmin olla "Tutkimuksen datanhallinnassa käytettävien palveluiden tietoturvaan ja tietosuojaan liittyvät riskit tulee

arvioida ja toteuttaa tarpeelliset riskien hallinnan toimenpiteet, jotka voivat sisältää palvelun auditoinnin ja sertifiointin.”

PALVELU-48: duplikaatti vaatimukselle PALVELU-31, ks. myös siihen liittyvä kommenttimme. Myös tunniste PALVELU-48 on jo toisen vaatimuksen käytössä tällä sivulla.

SAATAVUUS-17: Olisiko tässä hyödyllisempää API-rajapinnan sijaan edellyttää koneluettavaa latauspalvelua? Ehkä tätä on tarkoitettukin, mutta nyt muotoilun voi tulkita niin, että kaikesta datasta pitää pystyä hakemaan haluamiaan osia API-rajapinnan kautta ja vain joissakin tapauksissa koko tiedostoja. Koko aineiston lataaminen on kuitenkin useimmille aineistoille normaali ja ainoa tarpeellinen käyttötapaus. Myöskään maksuttomuus ei voi koskea kaikkia aineistoja.

SAATAVUUS-22: Sama vaatimus samalla tunnisteella esiintyy tällä sivulla myös korostamattomana (ei priorisoituna). Yhteisen sanaston käyttö voisi olla keskeiseksi tavoitteeksi nostettava asia, koko arkkitehtuurin tasolla. Se on keskeinen edellytys, jotta voidaan tuottaa yhteisiä palveluita, erityisesti yhteisiä tukipalveluita tai ohjemateriaalia.

OHJAUS-1: Keskeiseksi tavoitteeksi tämä on liian vaikeatulkintainen. Vaatimus pitäisi kirjoittaa uusiksi selkokielisesti.

Liiketoiminta-arkkitehtuuri

3.4 Palvelut

Mitä palveluita hyödyntäen toteutetaan käyttötapaus, jossa dataa jaetaan rajoitetusti tunnistetuille käyttäjille projektin päätyttyä? JULKAISU-1 eli datarepositoriot ei mainitse tunnistautumista mitään, ja yleensä nämä tulkitaan avoimiksi palveluiksi. SÄILYTYS-1 eli datan laskennan aikainen jakaminen liittyy projektin aikaiseen työhön. Vaikuttaisi siltä, että palvelu joko puuttuu tai jotain nykyisistä palveluista pitäisi laajentaa.

Tietoarkkitehtuuri

Ehdotetaan lisättäväksi termi ”koneluettava”, jota arkkitehtuuri käyttää laajasti ilman määritelmää. Määritelmäksi esimerkiksi ”tiedostomuoto, jonka rakenne mahdollistaa sen, että ohjelmistot pystyvät helposti yksilöimään, tunnistamaan ja poimimaan siitä tietoaaineistoja, yksittäisiä tietoja sekä niiden rakenteita” (lähde Tiedonhallintalautakunnan suositusten sanasto <https://sanastot.suomi.fi/terminology/ea095c41-cca1-49c5-98e7-c275cb930ac2/concept/07d28101-4b09-4459-8d05-56fb0409e7c4>)

Termin ”datan jakaminen” määritelmä tulisi muuttaa esim. siten kuin joku on jo lähde-sarakkeessa ehdottanut: ”Jakaminen on pääsy aineistoon hallitusti ja sovittujen käytäntöjen mukaisesti. Jakaminen vaatii usein teknisen toteutuksen.”