



MDM-esiselvitys

Organisaatio- ja toimipaikkatiedot





23.3.2015

TIIVISTELMÄ

Yhteiskunnan toiminta on tietointensiivistä. Lainsäätäjän yhteiskunnalle asettamien velvollisuuksien täyttäminen edellyttää, että yhteiskunnan keskeisimmät tiedot (”ydintiedot”, ”master data”) ovat kunnossa.

Esimerkiksi kaikilla luonnollisilla henkilöillä on Suomessa keskitetysti hallittu yksilöllinen henkilötunnus. Henkilön ydintietojen hallintaan on yhteisesti sovitut periaatteet. Yhteiskunnan toiminnan voi rakentaa näiden luotettujen tietojen varaan.

Sen sijaan toimipaikka- ja organisaatietiedoille tällaisia käytäntöjä ei ole: esimerkiksi yksittäisellä päiväkodilla ei ole selkeää yksiselitteistä tunnistetta. Päiväkodin nimen hallintaan ei ole yhteisiä sääntöjä.

Esiselvityksen lähtökohta on, että näiden toimipaikkojen – joita kutsutaan yleisemmällä tasolla *yksiköiksi* – ydintiedot tulee saada haltuun. Yksikköjen löydettävyys Suomi.fi-kaltaisten palveluhakemistojen kautta vaatii tätä. Tietohallintolain edellyttämä yhteentoimivuuden tavoite ja kansallisen palveluväylän tehokas hyödyntäminen asettavat niin ikään vaatimuksia samaan suuntaan.

Esiselvityksessä osoitetaan, että mittavaa taustatyötä ydintietojen ja yksikköjen ydintietojen paremman hallinnan eteen on jo tehty: yhteiset sanastot, ydintietoarkkitehtuurit, yksilöllisten tunnusten (OID) suunnittelu, kaikki on ollut työn alla jo pitkään.

Lopputulos ei kuitenkaan ole ollut tyydyttävä: yhteinen ydintietonäkemyks yksikköjen hallinnasta puuttuu. Suomi.fi ja palveluväylällä ei ole yksikköydintietovarantoa käytettävissään. Tiedonhallinta on hajanaista.

Juurisyy ongelmalle on, että mittavatkin panostukset ydintietojen hallintaan liittyen ovat olleet erillisiä harjoituksiaan, vailla kokonaisnäkemyksiä. Ratkaisuksi ehdotetaan kolmeen asiaan tarttumista.

- ✓ **Tieto itse.** Ydintiedon tietalueiden määrittely sekä yleisellä tasolla (ydintietalueiden, ”data domains”) että spesifillä yksikköydintietojen tasolla. Ratkaisun keskeinen osa on yhteisen jaetun yksikköydintiedon määrittely suppeaksi sekä ei-hierarkkisen yksilöivän tunnisteen luominen.
- ✓ **Hallintamalli.** Yhteinen yksikköydintieto on realistinen tavoite vain, jos sen hallinnalle kyetään luomaan yhteiset rakenteet. Ratkaisussa ehdotetaan niin ohjaus-, omistajuus-, tuki kuin ylläpito-tensojen määrittelyä. Esimerkkejä kustakin annetaan.
- ✓ **Teknologia.** Suurin osa ydintietojen hallinnasta tulee säilymään hajautettuna. Aivan ydinosa, ”jaettu” yksikköydintieto, tulee saada keskitetyn ratkaisun alle.

Ratkaisun pilotoinnin merkitystä korostetaan: kokonaisvaltaisen ”big-bang” ratkaisuyrityksen asemesta yksikköydintietoa tulee rakentaa paloittain. Muutaman jo käynnissä olevan projektin yhteistyö on luonteva tapa ajaa yksikköydintietojen kokonaisratkaisua eteenpäin.





Sisältö

JOHDANTO	4
ESISELVITYKSEN TAVOITE	6
ESISELVITYKSEN RAKENNE	9
TAUSTAA – KESKEISIÄ KÄSITTEITÄ JA LÄHTÖKOHTIA	10
JUURISYY	15
RATKAISU	17
1 ENSIMMÄINEN ALUE – ”TIETO” ITSE	19
2 HALLINTA – DATA GOVERNANCE	26
3 TEKNOLOGIA	31
TOTEUTUS JA TIEKARTTA	35
LOPUKSI	37
MUUTOSHISTORIA	37





23.3.2015

Johdanto

Nykyaikaisen yhteiskunnan toiminta on tietointensiivistä. Niin kansalaiset kuin yritykset ja julkisen hallinnon toimijat odottavat tarvittavan tiedon olevan sekä laadukasta että helposti saatavilla. Sähköisten tiedonvälityskanavien rooli korostuu.

*Ydintiedot*¹ ovat yli yhteiskunnan ja sen toimijoiden tarvittavia keskeisimpiä tietoelementtejä. Ydintiedoille ominaista on, että niistä tulisi olla yhtenäinen käsitys läpi prosessien ja läpi organisaation, usein organisaatiorajojen ylikin.

Esimerkkinä ydintiedosta voidaan pitää yhteiskunnan jäsenten, henkilöiden, keskeisimpiä tietoja. Läpi modernin maailman yhteiskuntajärjestelmien, joitain poikkeuksia lukuun ottamatta² luonnollisille henkilöille annetaan yksilöllinen tunniste. Me Suomessa kutsumme sitä henkilötunnukseksi. Tunnuksella on sovittu formaatti, se on yksilöllinen ja tarkoitettu pysyväksi³.

Kustakin henkilöstä tallennetaan myös esimerkiksi syntymäaika ja -paikka ja nimi. Henkilön nimestä on erikseen asetettu laki⁴, jossa esimerkiksi määrätään suku- ja etunimi pakollisiksi sekä etunimien maksimimääräksi kolme. Alla oleva kuva malliajokortista konkretisoi osajoukkoa yhteisesti sovitusta tiedoista.



Kuva 1 - Luonnollisen henkilön ydintietoja

Samankaltaisia sääntöjä asetetaan juridisten henkilöiden tunnistamiselle ja nimeämiselle: yrityksillä ja yhteisöillä on yksilöllinen tunnus. Niille tulee antaa myös nimi. Nimeä koskevat yksityiskohtaiset ohjeet, joita nimen hakijan on noudatettava⁵.

¹ JHS 179 suosituksessa vuodelta 2011 Master Data -termin suomennokeksi valittiin ”ydintiedot”, mutta myös ”perustiedot” on laajasti käytössä

² Yhdistyneiden kuningaskuntien taistelu yhteistä henkilötunnusta vastaan on tunnettu poikkeus. Sosiaaliturvatunnus kyllä annetaan syntyville lapsille, mutta se ei aina ole yksilöllinen

³ Laki väestötietojärjestelmästä ja Väestörekisterikeskuksen varmennepalveluista 661/2009

⁴ <https://www.finlex.fi/fi/laki/ajantasa/1985/19850694>

⁵ <http://www.prh.fi/fi/kaupparekisteri/yritystennimet/nimiohjeet.html>





Monta muutakin yhteiskunnan keskeisiä tietoalueita on samankaltaisen kontrollin piirissä: laissa ja asetuksissa on määriteltä, mitä tietoa tarvitaan, mikä sen formaatti on ja kuka vastaa tiedon ylläpidosta. Tämä luo pohjaa yhteiskunnan toiminnalle.

Tilanne ei kuitenkaan ole yhtä selvä kaikilla tietoalueilla⁶. Esimerkiksi yhteiskunnan palvelujen tarjoavien ”toimipaikkojen”, kuten päiväkotien ja terveyskeskusten osalta, yhteisiä nimeämis- ja yksilöintitunnussääntöjä ei ole olemassa. Tampereen kaupunki kuvaa tilannetta hyvin tarjouspyynnös-sään⁷:

”Esimerkiksi samaa tarkoittavan tiedon sisällössä on eroavaisuuksia mm. tiedon formaatin ja määrittelyn osalta. Nykyiset määrittelyt tukevat vain kyseisen osa-alueen operatiivisen järjestelmän tarpeita. Lisäksi tietojen käsittelyssä on erilaisia toimintatapoja. Yhtenäisen organisaatitiedon määrittelyn ja hallintamallin puuttuminen on johtanut päällekkäisiin työnkulkuihin, tiedon heikkoon laatuun, käytettävyyteen ja saatavuuteen. Tiedon heikko laatu näkyy mm. tiedolla johtamista tukevien tietojen käytössä johdon raportoinnissa.”

Tampere ei ole yksin. Sama ongelma on läpi kunta- ja viranomaiskentän. Kun edes yksittäisen kunnan sisällä tällaiseen toimipaikkahallintaan ei ole, niin tilanne ei ole yhtään parempi yli kunta- ja muiden hallinnonalojen rajojen. Pikemminkin päinvastoin: toisin kuin kuvassa 1 olleen Meri Anna Auran kohdalla, mitään yhteisesti sovittua nimeämis- ja tunnusikäytäntöä, saati muuta ydintiedon määrittelyä yhteisellä tavalla, ei ole *edes yhden kaupungin sisällä*. asiat ovat kontekstiriippuvaisia: hallittavien tietojen sisältö ja ylläpitokäytännöt ovat järjestelmästä tai viranomaisesta riippuvaisia.



Kuva 2 - Yksikön tietoja

⁶ ”Tietoalue” käännetty englannin kielestä Data Domain, viitaten tässä yhteydessä ydintiedon hallinnan alaisiin tietokokonaisuuksiin

⁷ Tarjouspyyntö, Tampereen Logistiikka Liikelaitos, 2013





Toisin kuin henkilön kohdalla, yllä oleva yksikön tietojen esimerkki ei ole vakioitu: päiväkodin nimeämiselle ei ole yhteisiä sääntöjä, päiväkodilla ei ole yksiselitteistä yksilöivää tunnusta. Ja niin edelleen.

Tämän esiselvityksen taustaoletus on, että Tampereen kaupungin tarjouspyynnössään lainauksessa kuvaama ongelma on läpitunkeva, läpi yhteiskunnan eri instituutioiden. Nykytilanne asettaa merkittäviä haasteita ja on merkittävä este yhteiskunnan ”kitkattomuuden”⁸ ja tietohallintolain edellyttämän yhteentoimivuuden tiellä.

Esiselvityksen oletuksena on, että yhteiskunnan organisaatioille ja niiden toimipaikoille – esimerkiksi päiväkodeilla, leikkipuistoilla, vanhainkodeilla – niin julkisen sektorin ylläpitämille kuin yksityisillekin – voidaan määritellä yhteisiä ydintietoja. Minimissään tämä tarkoittaa yhteisiä tunnisteita sekä yhteisiä nimikäytäntöjä. Esiselvityksessä toimipaikkoja kutsutaan *yksiköiksi*.

Näin saadaan parempi pohja niin yksikköjen tietojen löydettävyydelle ja niiden ymmärrettävyydelle kansalaisten toimesta, mutta myös viranomaisten itsensä toiminnalle. Tavoitteena on saavuttaa niin tehokkuuden kautta kustannushyötyjä, kuin parempaa palvelua kansalaisille.

Esiselvityksen tavoite

Esiselvityksen tavoitteena on pohjustaa ratkaisua organisaatio- ja toimipaikkatietojen käytännön hallinnalle:

Miten yhteiskunnan yksiköiden ydintiedot saadaan haltuun niin, että tiedot ovat löydettävissä & käytettävissä eri tarkoituksiin, niin yksikköjen tuottamia palveluita käyttäville kansalaisille ja yhteisöille kuin yksikköjä valvoville viranomaisille ja yksiköille itselleen?

Tavoite vaatii tarkennusta:

- *Yksiköillä* tarkoitetaan laajasti toimipaikka- ja toimipistekäsiteitä – kaikkia tahoja, jotka yhteiskunnan palveluita tarjoavat tai ovat yhteiskunnan toisten yksiköiden toiminnan kohteita. Karkeasti sanoen silloin, kun yksiköllä on atominen identifioinnin tarve joko yksikön asiakkaan kannalta (yksilön löytäminen, yksikköön liittyvien ohjeiden / muiden tietojen hakeminen) tai esimerkiksi valvovalla viranomaisella on tarve tunnistaa yksikkö omissa toimenpiteissään.
- *Ydintiedot* viittaavat ensisijaisesti yhteiseen yksilöivään tunnukseen sekä yhteisiin nimeämiskäytäntöihin, toissijaisesti myös muihin yksikön tietoihin.

⁸ Raportissa ”21 Polkua kitkattomaan Suomeen” kuvataan keinoja, joilla yhteiskunta pystyy vastaamaan teknologia- ja muihin muutoksiin.





Kaikkienensa haetulla ratkaisulla ydintietojen hallintaan yleisesti ja toimipaikkaydintietojen erityisesti voidaan nähdä olevan paljon muitakin tavoitteita:

- ratkaisun tulee edistää tietohallintolain edellyttämää *yhteentoimivuutta*
- ratkaisun tulisi olla joustava ja sallia ketterä kehitys – samaan aikaan ratkaisun tulee kuitenkin olla tiukassa kontrollissa
- ratkaisun tulee tukea mahdollisimman hyvin olemassa olevia standardeja
- ratkaisun tulee olla keskitetty, jotta se tukee Kansallista palveluväylää, mutta ratkaisu tai sen osa ei saa muodostaa sellaista solmukohtaa, jonka toimimattomuus häittäisi olennaisesti viranomaisten toimintaa ja kansalaisten elämää

Epäsuorina tavoitteina voidaan myös nähdä avointen ratkaisujen ja ICT-alan kilpailun edistäminen. Suljetut tietomallit ovat kilpailun esteitä.

Hidasteet tavoitteen tiellä

Tavoite on näennäisesti triviaali: jos luonnollisille henkilöille voidaan määritellä yksilöivät henkilötunnukset, henkilön nimelle yhteiset säännöt ja systematisoida prosessit ja järjestelmät niin henkilötunnusten kuin nimien ylläpitoon, niin miksei yhteiskunnan yksiköille kuten päiväkodeille, kirjastoille tai vaikka uimarannoille voida tehdä samaa?

Tavoite ei kuitenkaan ole triviaali. Tilanteeseen on monta syytä:

- Esimerkkinä käytetty ”päiväkoti” on vain yksi yksikön ilmentymä. *Yksikköjen ilmentymiä on paljon erilaisia*, mm. yleistyvät yhteispalvelupisteet, liikkuvat yksiköt (”punkkibussit” jne), virtuaaliset ”toimipaikat”, etäpalvelutoimipaikat sähköisiä kanavia pitkin (”lääkäri videon kautta”), jne. Yksityiskohtaisemmalle tasolle mentäessä tällaiset erilaiset ilmentyvät usein vaativat hieman erilaisia tiedonkäsittelytarpeita.
- Yksikkö ei ole staattinen käsite, vaan *yksiköillä on elinkaarensa*, joka usein on luonnollisen henkilön elinkaarta monimutkaisempi. Yksiköt voivat yhdistyä, lakata olemasta, muuttaa muotoaan. Niiden ydintietotarpeet voivat elää merkittävästi yli elinkaaren.
- Yhden yksikön ja sen ”sisäkkäisen” mahdollisesti *erillisen yksikön rajanveto on toimintasidonnainen eikä aina välttämättä yksiselitteinen*: onko päiväkodin yksi osa – jossa hoidetaan erityislapsia ja jolla saattaa esimerkiksi olla erilainen aukio-oloaika – eri toimipaikka kuin muu päiväkotitoiminta? Miten yksi iso yksikkö kuten yliopisto jaotellaan yksiköiksi tarkemmalla tasolla?
- *Yksikön käsite on kontekstisidonnainen*: valvovalle viranomaiselle toimipaikka voi ilmentyä eri tavalla kuin esimerkin päiväkodin palveluita tarvitsevalle pienen lapsen vanhemmalle. Myös viranomaisten toimialueiden yli toimipaikka voi ilmentyä eri tavalla. Työ- ja elinkeinoministeriön alaiset viranomaiset näkevät toimipaikan usein eri näkökulmasta kuin Sosiaali- ja terveysministeriön viranomaiset, tarviten yksikön ja sen tietojen erilaista käsittelyä.





- *Yksiköjä hallinnoidaan erilaisista lähtökohdista, erilaisilla teknologioilla, erilaisten johtamiskulttuurien alla. Pienessä kunnassa yksiköiden hallinta voi olla yhden henkilön sivutoimi, suuressa kaupungissa ylläpitoon voi olla pitkälle organisoitua ja tietotekniikan hyvin tukemaa.*

Näistä syistä päästään siihen, että tavoitellun ratkaisun tarkan tason määrittely on haastavaa. Mitä oikein haetaan? Lähestytäänkö ”ratkaisua” tietojärjestelmänäkökulmasta? Yhteisten tietojen määrittelyn ja tietomallinnuksen kulmasta? Viitearkkitehtuurinäkökulmasta? Yhteiskunnan tavoitteiden näkökulmasta? Lainsäädännön näkökulmasta?

Ratkaisua hankaloittaa yksikkö -käsitteen monitasoisuus. Katsotaanko yksiköitä valtakunnalliselta vai kunnalliselta tasolta vaiko lupa- tai valvovan viranomaisen tarpeiden kautta? Myös vastuiden kannalta Suomen lainsäädännössä palvelutuotanto ja tiedonhallinta on lainsäädännöllisesti hajautettu pieniin viranomaisyksiköihin, korostaen viranomaisten erillisyyssperiaatetta⁹.

Yhteenvetona tavoitteista voidaan todeta, että yksiköiden ydintiedon hallinnalla tavoitellaan lähes Graalin maljaa: kaikkien ratkaisujen äitiä. Onko sellaista mahdollista edes löytää?

⁹ Raportti ”Palveluarkkitehtuuria tukeva lainsäädäntö”, 2013, professori Voutilainen



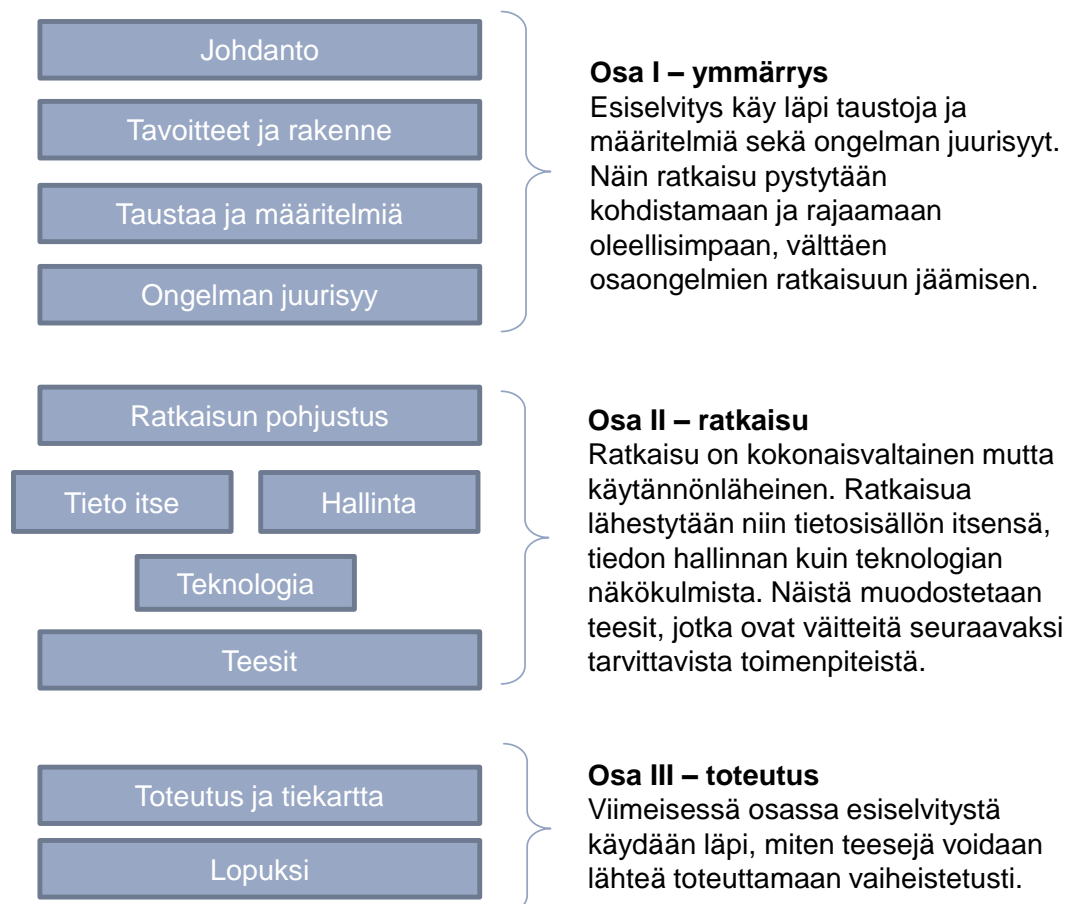


23.3.2015

Esiselvityksen rakenne

Esiselvityksessä tullaan näyttämään, että paljon työtä yksikköjen (ydin)tietojen paremman hallinnan tavoitetta kohti on jo tehty vuosien varrella. Työ on kuitenkin ollut enemmän osaongelmien ratkaisua kuin koordinoitua toimintaa kokonaisongelman ratkaisua kohti.

Esiselvitys rakentuu alla kuvatulla tavalla.





23.3.2015

Taustaa – keskeisiä käsitteitä ja lähtökohtia

Yksikköjen ydintietojen hallinnasta ei voi lähteä liikkeelle irrallisena kokonaisuutena. Se nivoutuu keskeisesti muuhun ydintietojen hallintaan. Lähdemme liikenteeseen ydintietojen hallinnasta yleisellä tasolla, josta edetään yksikköjen ydintietoihin.

Lähtökohta

”Ydintiedot” ja ”ydintietojen hallinta” ovat suhteellisen nuoria käsitteitä. Googlen trendihausta voi tarkastella MDM-käsitteen nousua huhtikuussa 2005. Data Governance, (ydin)tiedon hallintamalli, on syntynyt vasta vuonna 2008.¹⁰ Suomen tietotekniikan tilannetta hyvin peilaavassa Talentumin julkaisemassa Tietoviikossa Master Data koskevat artikkelit yleistyivät vuonna 2011.¹¹ Yliopistomaailmassa puolestaan ydintietojen hallintaan ja tiedon laatuun liittyvät lopputyöt ovat yleistyneet vasta aivan parin viimeisen vuoden aikana.¹²

Tähän taustaan nähden julkinen sektori ei kulje jälkijoukoissa: kokonaisarkkitehtuuria käsittevä JHS 179, jossa mm. Master Data Management määritellään ja suomennetaan sanaksi ”ydintieto” – on julkaistu helmikuussa 2011. Kuten tässä kappaleessa myöhemmin kuvataan, ydintietojen hallintaa on käsitelty laajasti myös muissa yhteyksissä.

Määritelmistä

JHS 179 määrittelee ydintiedon ja sen hallinnan seuraavasti:

”Master Dataksi eli ydintiedoksi luokitellaan tieto, jota käytetään useassa käyttökohteessa samanaikaisena tai jota useampi prosessi/toiminto tarvitsee tai hyödyntää. Ydintieto on pysyväisluonteista tietoa, joka kuvaa tietokokonaisuuksia, joista on tai josta tulisi olla yhtenäinen käsitys koko organisaatiossa ja joka on samalla avaintietoa organisaation toiminnalle. Ydintietoa on yleisesti asiakas-, tuote-, henkilöstö-, materiaali- ja toimittajatieto jne.”

”Master Data Management MDM on toiminto ja prosessi joka hallinnoi, korjaa ja ylläpitää toiminnan edellyttämän Master Datan eli ydintiedon. Sen tehtävänä on huolehtia riittävästä ydintiedon laadusta, jotta raportointi ja seuranta voidaan pitää luotettavana ja yksikäsitteisenä.”

Ydintiedon hallintaan liittyvä pohjatyö ei jää JHS 179 määritelmiin: PerustA -työssä on määritelty ydintietojen¹³ viitearkkitehtuuri perinpohjaisesti perustietovarantojen näkökulmasta. PerustA -työn yhteydessä määritellystä viitearkkitehtuurin kahdestatoista eri tasosta oheinen kuvakaappaus näyttää viisi ensimmäistä:

¹⁰ Haettavissa trends.google.com -osoitteesta

¹¹ Haettavissa tivi.talentum.com -osoitteesta

¹² Keskustelu Samuli Pekkolan, Tampereen teknillisen yliopiston tietojohdamisen professorinm, kanssa 5.12.2014

¹³ Joita tosin kutsutaan JHS 179:n määritelmän vastaisesti ”perustiedoiksi”





1. Huomioi ja kuvaa yhteentoimivuus, sidosarkkitehtuurit ja muut ohjaavat tekijät
2. Huomioi julkisen hallinnon arkkitehtuuriperiaatteet
3. Tunnista perustietovarantoon liittyvät sidosryhmät ja roolit
4. Kuvaa perustietovarantoon liittyvät prosessit
5. Tunnista tarkasteltavan tietovarannon taso tietovarantojen jäsenyyksessä

Olemassa oleva taustatyö ei rajaudu JHS- ja viitearkkitehtuuridokumentteihin: mm. Kuntaliitto on laatinut oman viitearkkitehtuurinsa ydintietojen hallintaan¹⁴. Siinä pyritään määrittelemään ydintiedon hallinnan keskeiset alueet yhteisiin attribuutteihin ja niiden harmonisointikysymyksiin asti, kuntasektorin näkökulmaa korostaen.

Kaiken kaikkiaan määritelmä- ja viitearkkitehtuurityön voi katsoa olevan kattavaa. Sinänsä määritelmää olisi mahdollista kehittää esimerkiksi ottamalla eksplisiittisesti esille keskeisiä termejä kuten tietovarallisuus ja roolit & vastuut, mutta määritelmien kehittämisen rajahyöty jäisi vähäiseksi.

Ydintiedon hallinnan keskeisten määritelmien tai viitearkkitehtuurien määrittely uudestaan on korkeintaan inkrementaalikehitystä. (joskin voidaan pitää perusteltuna, että esimerkiksi JHS 179:n MDM-liite päivitetään osana ydintietojen yhtenäisempää haltuunottoa, jota tässä esiselvityksessä myöhemmin suositellaan)

Yksikköjen ydintietojen hallinnan taustoja ja hankkeita

Työtä on tehty paitsi yleisellä tasolla ydintiedon hallinnan käsitteiden ja toimintamallien kehityksessä, mutta myös spesifisti organisaatio- ja toimipaikkatietojen käsittelyyn liittyen: JHS 159 käsittelee *kohteiden* (kohteita ovat mm. organisaatiot ja niiden osat sekä toimipaikat) tietojen hallintaa – sen ensimmäinen versio on jo vuodelta 2006. Näille määritellään yksilöintitunnusten (OID, Organization Identifier) pohjautuen globaaleihin ISO standardeihin.

Sosiaali- ja terveysalalla on hyväksytty yksilöintitunnusten käytön periaatteet jo 2007.¹⁵ Tällä työllä pyrittiin kohteiden (joita työssä kutsuttu *objekteiksi*) yksilöintiin mm. objektiluokan ja objektiluokan ilmentymien tasolla. Tämän työn tulokset ovat eläneet sosiaali- ja terveysalalla tähän päivään, mutta eivät ole laajentuneet hallinnonalojen yli.

¹⁴ ”Kuntasektorin MDM-viitearkkitehtuuri”, 2013

¹⁵ <http://www.hl7.fi/hl7-rajapintakartta/iso-oid-yksilointitunnuksen-kayton-kansalliset-periaatteet-sosiaali-ja-terveysalalla/>





Organisaatiota ja toimipaikkaa yleisellä tasolla määrittelevien sekä niihin liittyvien mallien suosittelevien dokumenttien lisäksi myös *soveltavia hankkeita* on käynnissä. Näistä mainittakoon:

- Sosiaali- ja terveysministeriön selvitys organisaatio- ja toimipaikkatiedon haltuunotosta vuodelta 2014
- Työ- ja elinkeinoministeriö käynnissä oleva ydintietojen hallinnan järjestelmä- ja toimintatapakehitys
- Suurten kaupunkien työ aiheen ympäriltä – johdantokappaleessa mainitun Tampereen hankkeen lisäksi esimerkiksi Helsingin kaupunki on työstänyt mm. palvelukarttansa kehittämisen myötä yksikkötietojensa systemaattista hallintaa. Helsinki on saanut työstään kansainvälistäkin tunnustusta¹⁶ jo toissa vuonna. Sen uusin palvelukartan betaversio on yhä edistyksellisempi.

Sanastotyö

Yhteisen sanaston määrittely sivuaa vahvasti ydintietojen hallintaa: jos haluamme luoda yhteisiä tapoja hallita yhteiskunnan eri yksiköiden ydintietoja, yhteinen sanasto (tai vähintään osajoukko sanastosta) on yksi työn edellytys.

Yhteiseen sanastokehitykseen on panostettu Suomessa: FinnOnto-hanke¹⁷ oli käynnissä Aalto-yliopiston ja Helsingin yliopiston välillä 2003-2012. Se jatkui ONKI-projektina¹⁸ Valtiovarainministeriön ja Kansalliskirjaston yhteishankkeena. Tulokset ovat nyt Finto-palvelun¹⁹ alla. Finto tarjoaa käyttöliittymän ja avoimet rajapinnat sanastojen, ontologioiden ja luokitusten käyttöön.

YHTEISKUNTA	JUHO - Julkishallinnon ontologia
	JUPO - Julkisen hallinnon palveluontologia
	Korkeakoulujen tutkimustiedonkeruussa käytettävä tieteenalaluokitus
	LIITO - Liiketoimintaontologia
	MERO - Merenkulkualan ontologia
	PUHO - Puolustushallinnon ontologia
	TSR-ontologia

Kuva 3 - Esimerkki Finto-palvelun sisältöalueista

¹⁶ <http://www.visithelsinki.fi/ammattilainen/miksi-helsinki/ajankohtaista/paakaupunkiseudun-palvelukartta-palkittiin>

¹⁷ FinnONTO – suomalaiset semanttisen webin ontologiat <http://www.seco.tkk.fi/projects/finnonto/index.fi.php/>

¹⁸ Kansalliskirjaston sivu ONKI-projektista http://www.kansalliskirjasto.fi/kirjastoala/uutiskirje/2_2013/onki.html

¹⁹ Finton kotisivu <http://finto.fi/fi/>





Kaikkien ontologia- ja sanastotyötä on tehty laajasti läpi yhteiskunnan eri toimijoiden. Tehdyn työn arvo (yksikköjen) ydintietojen hallinnan kannalta on toinen kysymys ja palaamme siihen.

Kansainvälistä vertailukohtaa

Yksikkö- ja muiden ydintietojen hallinta linkittyy laajempaan digitalisoitumisen²⁰ rooliin osana yhteiskuntien kehitystä ja toimintaa on pinnalla kaikkialla. Konsulttitoimisto McKinsey julkaisi joulukuussa 2014 artikkelin ”Public-sector digitization: The trillion-dollar challenge”, jossa hyvin tiivistetään koko maailman mittakaavassa tiedonhallinnan ongelmakenttää:

“Similarly, when systems and data are owned by different departments and functions, on a range of platforms and with differing taxonomies and access requirements, it can be difficult to invest at scale and generate sufficient economies. Silos, fragmentation, and the absence of a central owner for nationwide IT infrastructure and common components can make it hard to connect the internal “plumbing” to create a seamless experience for the end user, be it a government worker, a business user, an average citizen, or another intergovernmental office.”

Tämä tiivistää yksikköydintietojenkin ongelmakentän lintuperspektiivistä katsoen: tietoa hallitaan eri toiminnoista, hyvin erilaisissa järjestelmäalustoissa ja käyttäen erilaisia sanastoja ja taksonomioita. Saumattoman kokemuksen – jossa esimerkiksi kaikki Suomen päiväkodit tai uimarannat olisivat yhteisellä tavalla nimettyjä, yhteisessä hakemistoissa helposti löydettävissä eri päätelaitteiden yli, integroituina erilaisiin (esimerkiksi karttapalveluihin) – synnyttäminen on haastavaa.

McKinseyn artikkelissa nimetään erikseen hyviä esimerkkejä muutamissa maissa kehitetystä tiedonhallinnan käytännöistä. McKinseyn malliesimerkeikseen valitsemat maat - Turkki, Englanti, Tanska, Viro, Hollanti, Etelä-Korea ja Yhdysvallat – resonoivat hämmäntävän hyvin palvelutietovaranto (PTV) ja Suomi.fi kehitystyön aikana parin viime vuoden aikana tehdyn benchmarking-työn kanssa. Sen aikana Valtiokonttori ja muut osalliset viranomaiset ovat tehneet vertailevaa tutkimusta Viron, Tanskan ja Englannin kanssa²¹.

Tutkimusta on tehty laajemmasta kuin vain ydintietojen alueen näkökulmasta. Ydintiedot – erityisesti toimipaikkoihin ja niistä tarjottaviin palveluihin liittyen – ovat kuitenkin olleet keskeisessä osassa:

²⁰ ”Digitalisoitumisen/digitalisaation määritelmään ei lähdetä syvällisemmin tänä esiselvityksen puitteissa – karkeasti määritellen se viittaa laajemmin toiminnan ja prosessien sähköistymiseen ja monikanavaistumiseen, jonka avulla organisaatioiden toiminta sekä tehostuu että muuttuu läpinäkyvämmäksi ja hallitummaksi. Esiselvityksen kannalta riittää todeta, että ydintiedoilla on äärimmäisen keskeinen rooli koko digitalisoitumiskehityksessä.

²¹ Useita dokumentteja, mm.. ”UK Digital Strategy – highlights”, Marko Latvanen, ”2014-11-13 GOV.UK”, Pirkko Romakaniemi, ”Borger.dk – the Danish citizen portal”, Morten Meyerhoff Nielsen, Rasmus Espholm, ”Borger.dk teemapäivä 30.11.2012”





erityisesti kansalaisten palveluportaalin rakentamisen kannalta toimipaikkojen ja niistä tarjottavien palveluiden ydintietojen systematisointi onkin koko digitasaatiokehityksen ytimessä.

Näiden kansainvälisesti menestyneiden vertailumaiden opit ydintiedonhallinnalle, toimipaikkatiedon kuin yleisemminkin, voidaan tiivistää seuraavasti:

- **Sopiminen** – keskitetty päätöksenteko; valta ja vastuu sopia keskitetysti yhteisistä käytännöistä tiedonhallintaan;
- **Tukeminen** – pelkkä sopiminen ei riitä; keskitetty tukiorganisaatio yhteisten käytäntöjen (osin teknologiankin) tukemiseen on tarpeellinen;
- **Valvonta** – pelkkä tukeminenkaan ei riitä, vaan tarvitaan keskitetty valvonta, jotta tiedonhallinnan käytännöt säilyvät kurinalaisina

Keskittämiseen pyrkimisestä esimerkkimaista mainittakoon Yhdysvalloissa tuore esimerkki vuoden 2014 heinäkuulta:

“Commerce Department will hire its first-ever Chief Data Officer. [...] The new Chief Data Officer will pull together a platform for all data sets; instigate and oversee improvements in data collection and dissemination; and ensure that data programs are coordinated, comprehensive, and strategic.”

Määritelmien ja taustojen yhteenveto

Johtopäätöksenä jo tehdyistä määritelmistä ja käynnissä olevista hankkeista voi todeta, että maaperä on kaikkea muuta kuin neutraali. Paljon arvokasta työtä on tehty läpi julkisen hallinnon, niin valtion kuin kuntien tasolla. Ydintietojen hallinnan arkkitehtuurillisen sekä käsitelmääritelmällisen pohjan, myös organisaatio- ja toimipaikkatietojen osalta, pitäisi käsiteltäessä sanoa olla jo määritelty puhki.

Taustatyö ei kuitenkaan ole johtanut tuloksiin:

- Kansallisen palveluväylän tukena *ei ole vielä(kään) hyvää ydintietoinfrastruktuuria*;
- Kansallinen Palvelutietovaranto *PTV taistelee toimipaikkahakemiston rakentamisessa*, kun yhteistä ydintietolähestymistä ei ole olemassa, ei edes aidosti yhteiskäytössä olevia yhteisiä määritelmiä, puhumattakaan vakioituista tietosisällöistä (vaikka mainittu ansiokas benchmarking-työ on ollut käynnissä jo vuosia);
- kunnat taistelevat omien ydintietohaasteidensa kanssa omien *erillisten projektiensa puitteissa*. (toteuttakoon kuitenkin, että kuntien välistä yhteistyötä on alettu viedä eteenpäin. Helsingin työn kautta aloitetun yhteistyön on hyvä esimerkkinä)





Perustavanlaatuinen kysymys on, miksi lähtökohdat riitä? Miksi (yksikkö) organisaatio- ja toimipaikatiedon hallintaa ei vain ratkaista – urheiluvälinevalmistaja Niken sloganin 'Just do it' periaatteella – jo tehdyn työn pohjalta? Tai miksi se ei ole kunnossa? Miksi koko ydintieton hallinnan kenttää ei saateta kuntoon samalla?

Tähän kysymykseen vastaaminen on oleellista ennen työn tavoitteiden saati ehdotettujen ratkaisujen määrittelemistä. Seuraava kappale avaa esiselvityksen ongelman juurisyys. Juurisyiden ymmärtämisen kautta voidaan rakentaa kestävää ja toteuttamiskelpoista ratkaisua.

Juurisyys

Juurisyys yksikköjen ydintietojen ja yleisemmällä tasolla ydintietojen hallintaan ei ole, että ongelmaa ei olisi tunnistettu. Juurisyys ei myöskään ole tähänastisten ydintietojen hallintaan käytetyn työn riittämätön määrä. Työtä on tehty pitkäjänteisesti. Työ on korkealaatuista ja tietoisuus ongelmasta on hyvällä tasolla.

Juurisyys on *kokonais kuvan puute, punaisen langan hukassa oleminen*: ydintietojen hallinta Suomen yhteiskunnan tasolla on tällä hetkellä *diskreettien, erillään ja eri toimijoiden laatimien, dokumenttien laittamista peräkkäin vailla niistä kumuloituvaa johdonmukaista kokonaisuutta*. Miten hyvin tai huolella ne yksittäin onkaan kirjoitettu, ei auta. Näiden dokumenttien inkrementaalikehittäminen, esimerkiksi aiemmin mainittu ydintietojen hallinnan viitearkkitehtuurin jalostaminen tai JHS 183:n toimipaikka- ja palvelukäsitteiden selkeyttäminen, ei ole yksin lääke.

Parhaatkaan yksittäiset benchmarking-suositukset maailman johtavista julkisen sektorin tiedonhallinnan maista eivät tuo päiväkotien ja muiden yhteiskunnan toimipaikkojen ydintietoja kurinalaisen hallinnan piiriin, jos ne eivät kytkeydy kokonaiskuvaan. Jos toteutetaan yksittäisiä askeleita, pahimmillaan päädytään yhä syvemmälle ydintiedon hallinnan siiloihin. Ratkotaan osaongelmia.

Vastaus on, että kokonaiskuva tulee muodostaa ja toimenpiteet tulee suunnitella niin, että ne palvelevat kokonaisuutta, eivät sen osajoukkoja.

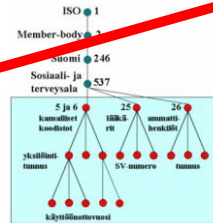
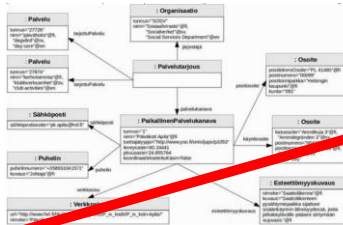
Vastauksen muodostaminen ei kuitenkaan ole triviaalia: jo toimipaikkojen monimuotoisuus sekä toimipaikkojen tiedonhallinnan elinkaari- ja kontekstiriippuvuus johtaa ongelmiin. Suomen yhteiskunnan tasoisessa heterogeenisessä kokoelmassa erilaisia organisaatioita ja niiden hyvin erilaisia toimintoja ja järjestelmiä on, että kokonaisvaltaista yhtä ”oikeaa”, eri toiminnalliset tarpeet & eri kontekstit kattavaa ratkaisua ei ole.





JulkICT-toiminto

23.3.2015



Hyviä malleja,
mutta eivät
johda yksikkö-
ydintietojen
ratkaisuun

JHS 183 – palvelujen tietomalli & ryhmittely

Hyvä dokumentti, mutta käsittelee sekä "palvelu" että "toimipaikka-/organisaatio" tietoa alueita, tekemättä selvää rajanvetoa.

OID-määritelmä, pohjautuen ISO-standardeihin

Standardipohjainen, perinpohjin mallinnettu. Käytännössä hierarkinen puu ei kuitenkaan ole kontekstiriippumaton – vaikka yksittäisessä kontekstissa (esim STM) palvelisikin.

Yllä olevat esimerkit kuvastavat tilannetta: vaikka kumpikin on kattava dokumentti, yleispätevää ja yhteisesti ymmärrettävää ratkaisua kontekstiriippumattomaan yksikköjen ydintietojen hallintaan niistä ei löydy.

Yksi juurisyy eittämättä myös on, että tähän asti tällä kokonaisuudella ei ole ollut selkeää omistajaa ja yhteistä rahoitusta. Omistajuus ja rahoitus ovat jakautuneet niin moneen eri hankkeeseen niin monella hallinnonalalla, että tilanteen hajanaisuus on ennalta arvattava seuraus.





23.3.2015

Ratkaisu

"Simplicity is the ultimate sophistication" (Leonardo da Vinci)

Moderni yhteiskunta ei ole yksinkertainen. Sitä ei saa pakotettua väkisin yksinkertaiseksi: jos lähemme muodostamaan ydintietojen hallinnalle liian pelkistetyn mallin, joka aidosti ei tue yhteiskunnan moninaisuutta, ratkaisu ei tule toimimaan.

Kuitenkin Leonardo da Vinci on ongelman ytimessä: yksinkertaisuus on arvo, kompleksiteetin vähentäminen pelkistämällä kokonaiskuva on ratkaisun avain. Yksikköjen ydintietojen hallinnan tehtäväkentän paloittelu loogisesti ja tehtäväalueiden vastuuttaminen yhteisesti hyväksytyllä tavalla on ratkaisu. Mallista tulee voida johtaa konkreettisia toimenpiteitä.

Tämän esiselvityksen teesinä on, että ratkaisu on **Suomen yhteiskunnan tasolle laadittu yhteinen ydintietostrategia**. Strategia ei saa jäädä paperinmakuiseksi, vaan sen tulee olla ymmärrettävä ja käytännönläheinen. Se tulee voida jalkauttaa vaiheittain. Yksikköydintiedot ovat ehdotettu lähtökohta jalkauttamiselle.

Premissit ratkaisulle

Esiselvityksen kansainvälisistä vertailukohdista kertovassa kappaleessa tulkittiin benchmarking-maista saatujen vertailukohtien johtavan kolmeen lähtökohtaan:

- ✓ **Sopiminen** – tarvitaan keskitettyä sopimista tärkeimmistä asioista, päätöksentekoa valtaa ja vastuuta sopia keskitetysti yhteisistä käytännöistä tiedonhallintaan;
- ✓ **Tukeminen** – pelkkä sopiminen ei riitä; tarvitaan keskitettyä tukiorganisaatio yhteisten käytäntöjen (osin teknologiankin) tukemiseen;
- ✓ **Valvonta** – pelkkä tukeminenkaan ei riitä, vaan tarvitaan keskitettyä valvontaa, jotta tiedonhallinnan käytännöt säilyvät kurinalaisina

"Keskitetty" sanan korostaminen voi herättää intohimoja ja pelkoja käytännöstä vieraantuneista ylemmän tason abstrakteista määrityksistä ja toimintamalleista. Pelko on aiheellinen: ydintietojen hallinta on viime kädessä hyvin paikallista toimintaa. Esimerkiksi yksikköjen ydintietoja hallitaan läpi yhteiskunnan eri toimijoiden. Valmiudet ja resurssit yksikköjen tietojen hallintaan ovat Helsingissä erilaisia kuin jossain pienessä kunnassa Lapin läänissä. Norsunluutornin kautta määritelty keskitetty malli on vaarallinen väärin käytettynä.

Samaan aikaan "keskittämiseen" pyrkiminen on itsestään selvä tavoite. Sitä ei voi eikä tule välttää: yksi päällimmäisistä tavoitteista on edistää yhteentoimivuutta. Yksikköjen ydintiedot tulee saada kansalaisten ja tietojen käyttävien viranomaisien tietoon yhdenmuotoisina, yhteisten kanavien kautta. Jos taustalla ei ole voimakkaita mekanismeja keskitetystä sopimisesta sekä myös keskitettyjen toimintamallien tukemisesta ja keskittyyä valvontaa, voidaan olla varmoja, että tiedot eivät tule yhteisesti hal-



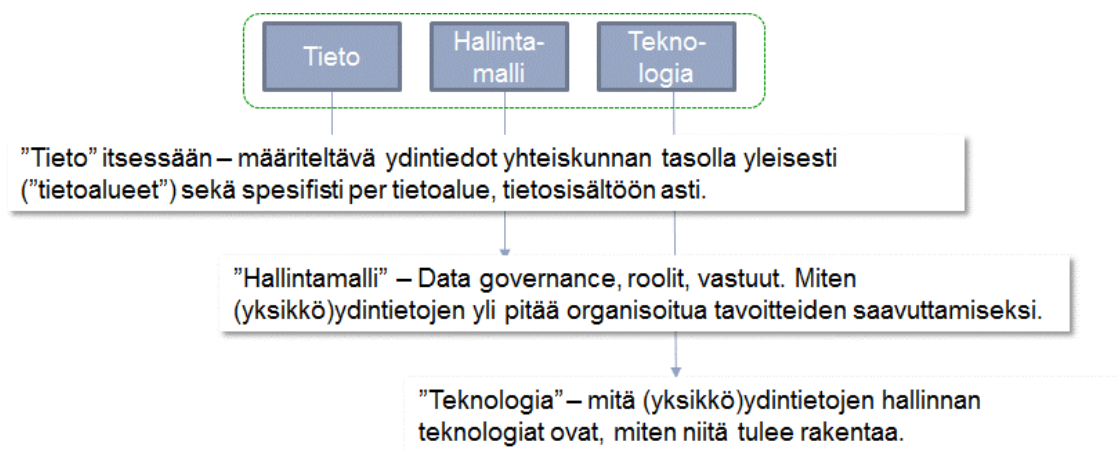


linnoiduksi. Muistettakoon aloituskappaleessa kerrottu nykyhetken lähtötila, jossa edes yksittäisten kaupunkien sisällä yksikköydintietoja ei saada pidettyä hallussa.

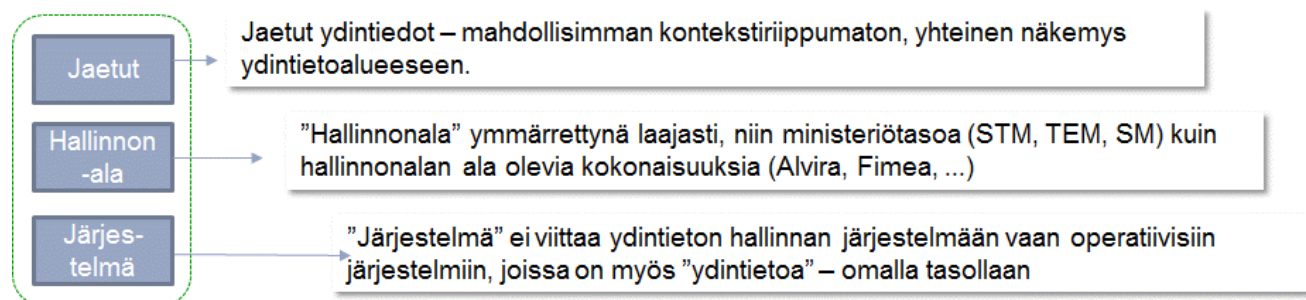
Ydinasia ratkaisun rakentamisessa on luoda samaan aikaan tiukasti keskitettyjä malleja ydintiedoista sopimiseen, niiden hallinnan tukemiseen sekä valvontaan, mutta kuitenkin sallia liikkumavara ja hyvin erilaisista lähtökohdista tehtävä työ.

Ratkaisun rakenne

Lähdemme liikkeelle ydintietojen hallinnan rakennuselementeistä – mitkä ovat ne osat, jotka tulee olla määriteltynä. Ensimmäinen askel on jakaa ydintietojen hallinnan tehtävät **alueittain**: alueiden jaotteluksi ehdotetaan:



Aluekohtainen jaottelu ei ole riittävä tapa pelkistää ratkaisua. Yksikköydintietoja – ja muita ydintietoja – tulee voida käsitellä **kerroksittain**. Yhden hallinnonalan ydintiedot – edes samalla tietoaalueella – eivät ole yhdenmukaisia toisen hallinnonalan ydintietojen kanssa. Kontekstiriippuvuus tulee pitää mielessä.





Tämä ja seuraavat kappaleet esiselvityksen käyvät kolme ulottuvuutta läpi. Kerrokset huomioidaan aina tarpeen mukaan. Ylimmäiseen, ”jaettujen ydintietojen”, tasoon keskitytään eniten.

Tavoitteena on pragmaattinen viitekehys suuren kuvan yli sekä johtaa käytännön ratkaisuehdotuksia. Käytännön ratkaisuehdotukset on numeroitu ja identifoitu marginaalissa.

1 Ensimmäinen alue – ”tieto” itse

Tässä kappaleessa käsitellään ”tietoa” ydintiedonhallinnan näkökulmasta. Epistemologisten pohdintojen asemesta tässä yhteydessä on syy rajoittaa ydintietojen käsittely käytännönläheiseksi.

- ✓ Ensimmäinen lähtökohta on pyrkiä *tunnistamaan ydintietoalueet* (data domains). Yhtä yhteistä, yhteisesti hyväksyttyä, ydintietoalueiden määritelmää ei julkisella sektorilla tällä hetkellä ole;
- ✓ seuraavaksi määritellään ”yhteisten ydintietojen” olemus tarkemmin;
- ✓ viimeiseksi päästään ”yksikköydintietojen” yksityiskohtaisempaan määrittelyyn.

Kappaleen keskittyy erityisesti jaettuun, yhteiseen, ydintietoon. Kappaleen päätteeksi vedetään yhteen suositukset myös hallinnonala- ja järjestelmäkohtaisista ydintiedoista.

Tietoalueet – data domains²²

Tällä hetkellä paras lähtökohta lähteä liikkeelle yhteiskunnan tietoalueista on PerustA -työssä tunnistetut loogiset tietovarannot, kts kuva alla.

²² Tietoalueiden kokonaisvaltainen määrittely ei ole osa tätä työtä. Aiheen nostaminen esille on kuitenkin välttämätöntä: yhteiskunnan tason yhteinen ydintietojen hallinnan kehittäminen edellyttää myös yhteistä näkemystä, mitkä ovat yhteiset ydintietoalueet. Tämän työn argumenttina on, että yksikköydintiedot ovat yksi näistä tietoalueista.





JulkICT-toiminto

23.3.2015



PerustAn tietoarkkitehtuuryössä loogiset tietovarannot on jaoteltu ”kehittäin”, karkeasti pohjautuen data-informaatio-tietämys -diskurssiin. Toimipaikkatieto ja julkishallinnon organisaatitieto on määritetty toiseksi sisemmälle kehälle, mm uutistieto- ja sopimustietovarannon rinnalle. Sisimmällä kehällä ovat perustietovarannot, mm. henkilö- ja yritys- & yhteisötietovarannot. Sisimmällä kehällä ovat yhteiskunnan ydintietovarannot.

Tehtyä jaottelua ei voida pitää loogisena: toiseksi sisimmällä kehällä sijaitsevat toimipaikat ja esimerkiksi ”uutiset” eivät ole tasoltaan analogisia. Toimipaikat (tässä esiselvityksessä yksiköt) ovat aitoa ydintietoa. Niiden voi katsoa noudattavan JHS 179:n määritelmiä: ”...jota käytetään useassa käyttökohteessa samanlaisena tai jota useampi prosessi/toiminto tarvitsee tai hyödyntää...”.

PerustA -työn perustietovarantomäärittely pohjautuukin enemmän lainsäädännön kautta asetettuihin jo olemassa olevien perustietovarantojen viranomaisvelvollisuuksiin kuin siihen, mitä tietoalueita tulee hallita ydintietoina.

Ydintietoalueiden määrittely

Tämän esiselvityksen tehtävänä ei ole kattava yhteiskunnan ydintietoalueiden määrittely. Jotta kuitenkin pääsemme kiinni ehdotukseen ”yksikkö”-ydintietoon, joudumme tekemään alustavia oletuksia tietoalue-määrittelyistä.





Esiselvityksen puitteissa seuraava yksinkertainen nelijaottelu otetaan työoletukseksi Suomen ylimmän tason *ydintietoalueista*:



Ydintietoalueita on neljä: osapuolitieto, paikkatieto, yksikkötieto ja palvelutieto (palvelujen määrittely ydintietoalueeksi on osin provokatiivista – aivan analogista ydintietoaluetta niistä ei synny, mutta palvelukäsitteiden yhtenäistäminen on yksi avain enemmän yhteisille toimintamalleille). Näiden neljän alueen sisällä yhteiset ydintietoalueet voivat jakautua useampaan tietovarantoon: esimerkiksi osapuolia koskevia perustietovarantoja on useampia, tärkeimpinä henkilötietovaranto ja yritys- ja yhteisötietovaranto. Samoin paikkatiedoille on useita ydintietovarantoja.

Näille ydintietoalueille ominaista on, että:

- Kaikkien ydintietoalueiden keskeisille määritelmille on *yhteinen, julkisen hallinnon organisatiorajat ylittävä, omistajuus* – esimerkiksi henkilötunnuksen formaatti ja henkilön nimikäsittely
- Ydintietoaluiden tietueille on *kontrolloituja ylläpitosääntöjä* – tietojen ylläpidon pitää olla kontrolloitua ja seurattua – esimerkiksi yhteistötietorekisterin käsittelysäännöt
- Ydintiedoille on *keskitetty luotettu järjestelmä* – esimerkiksi kiinteistötiedot

Esimerkiksi ”osapuoli”-ydintietoalue kattaa nämä hyvin: osapuolien ydintietoja hallitaan kahdessa päärekisterissä, luonnollisille henkilöille henkilötietorekisterissä ja juridisille henkilöille yritys- ja yhteisötietorekisterissä.

”Jaettu” ydintiedon taso on selkeä: esimerkiksi VRK:ssa on rajallinen osajoukko luonnollisten henkilöiden ydintietoja. Yksilöivä tunnus on sellainen, jonka avulla hallinnonalakohtaisia tietoja voidaan hallita.

Esimerkkinä jaettujen ydintietojen alapuolella olevista kontekstiriippuvaisista ydintietoalueista mainittakoon Valvira rekisterit terveydenhuollon ammattilaisista, esimerkiksi lääkäreistä. Lääkäreille on oma yksilöllinen tunnusnumerosa ja lääkäreistä tallennetaan tietoa paikalliseen ydintietojärjestelmään. Nämä eivät ole yhteisiä, vaan hallinnonalakohtaisia ydintietoja.





Tällaisen ei-jaetun tason ydintiedon osalta hallinta on selkeämpää: samanlaista organisaatorajat ylittävää omistajuutta, läpi Suomen eri kuntien tai toimijoiden meneviä ylläpitokäytäntöjä, ei tarvita. Valvira pystyy tekemään useimmat päätökset itsenäisesti.

Ydintietoalueiden tietosisällön määrittely

Ydintietoalueet tulee määritellä – mitä tietoa voidaan ylipäätään pitää *jaettuna ydintietona*? Mistä kulkee raja hallinnonalakohtaisiin tietoihin? Kysymys kulminoituu yhteisen ydintiedon määrittelyn laajuuteen: kuinka paljon tietoalueen tietosisältöä on syytä ottaa jaetulla tasolla haltuun. Teoriassa edellisen kappaleen Valvira-esimerkki voitaisiin nostaa ylemmälle tasolle. Miksei vain lisättäisi VRK:n väestörekisterin vastuulle myös SV-tunnusten (yksilöintitunnus nykynimeltään²³) ja muuta lääkärin ammattikuntaan liittyvää ydintietoa?

Kysymys on retorinen. Käytännössä tämä lähestyminen ei tulisi toimimaan. Tietotarpeet ovat hyvin heterogeenisiä: julkinen hallinto sekä siihen nivoutuva yksityinen toiminta on niin monitahoista, että kaiken kattava yksi yhteisesti jaettu tietomalli ei ole mahdollinen. Tiedonhallintaa tulee voida pilkkoa.

Tämän huomaa case-alueelta: TEMmin ja STM:n jo käynnissä olevista ydintiedon kehittämiseen liittyvistä tehtävistä ja JHS 183:n tapa mallintaa samaa organisaatio- ja toimipaikkatiedon data-alueita ovat jokseenkin erilaisia – ja aivan perustellusti. Myös olemassa olevat JHS 183:n implementaatiot mm. suurissa kaupungeissa soveltavat samaa dokumenttia eri tavoin.

Kaksi pääkulmaa tietosisällön määrittelyyn on:

1. Pyritään jo havaituista vaikeuksista huolimatta muodostamaan yhteinen kattava tietomalli yhteisestä ydintietoalueesta, "Enterprise (State/Society?) Data Model", joka vähintään loogisella tasolla yhdistää "kaikki" tarpeet yhteisen tietoalueen ympäriltä.
2. Pyritään redusoimaan kompleksiteettia muodostamalla yhteisestä ydintietoalueesta vain ehdoton yhteinen minimi ja mallintaa se. Kaikki muu mallinnetaan "alemmilla", hallinnonalakohtaisilla tai jopa niiden sisällä yksittäisten virastojen, yksittäisten järjestelmien tasolla tasoilla.

Tämän työn suositus on jälkimmäinen: "yhteinen ydintieto" määritellään tiukasti rajatuksi, koskemaan vain minimiosajoukkoa tiedoista. Sitä osajoukkoa, joka on aidosti "yhteinen".

"Yksikkö"-ydintietoalueeseen –mitä "yksiköt" ovat ja mikä niiden jaettu sisältö olisi?

Seuraava käytännön suositus voidaan johtaa edellisestä suosituksesta keskittyä minimijoukkoon tietoa.

²³ http://www.valvira.fi/tietopankki/rekisterit/terveydenhuollon_ammattihenkilot/julkiterhikki





- ✓ ”Yksiköillä” tarkoitetaan laajasti toimipaikka- ja toimipistekäsiteitä – kaikkia tahoja, jotka yhteiskunnan/yhteiskunnalle palveluita tarjoavat tai ovat yhteiskunnan toisten yksiköiden toiminnan kohteita. Yksikkö voi olla julkinen toimija tai yksityisen yrityksen/yhteisön toimipaikka.
- ✓ Yksikkö tulee olla määriteltynä, kun sillä on atominen identifioinnin tarve joko yksikön asiakkaan kannalta (yksilön löytäminen, yksikköön liittyvien ohjeiden / muiden tietojen hakeminen) tai esimerkiksi valvovalla viranomaisella on tarve tunnistaa yksikkö omissa toimenpiteissään.

Mikä olisi ”yksiköiden” tarkempi tietomalli? Mitä attribuuttitason tietoa ”yksiköistä” tulisi säilyttää? Näitä kysymyksiä on lähestytty useammassa hankkeessa. Alla poimittu yksityiskohtia Tampereen tällä hetkellä käynnissä olevan projektin määritelmistä sekä PTV:ssä tehtävästä työstä.

Kenttä
Toimipaikan lyhyt nimi (20 merkkiä)
Toimipaikan pitkä nimi (40 merkkiä)
Kuntalaiselle näytettävä nimi (100 merkkiä)
Katuosoite
Postinumero
Postitoimipaikka
Puhelinnumero
Postiosoite muu kuin katuosoite
Postiosoite
Postinumero
Postitoimipaikka
Toimipaikan www- sivujen osoite
Toiminnan aloituspäivä
Onko kyseessä ostopalvelu
Palveluntuottajan virallinen nimi

Tampere, käynnissä olevan organisatioprojektin ydintietomäärittelyjä.

Nimi	Teksti	1..*	Organisaation nimi eri kielillä
Kuvaus	Teksti	1..*	Organisaation kuvaus eri kielillä
Vaihtoehtoinen Nimi	Teksti	*	Organisaation itsestään käyttämä asiakasystävällinen (kansankielinen) nimi eri kielillä.
Logo tai kuvituskuva	Tiedosto	*	Organisaation tunnuskuva eri kielillä.
OID	Tunnus	1	Organisaation yksilöivä tunnus.

PTV:ssä tehtävää yhteisen tason ydintietomäärittelyä

Kummassakin tapauksessa yksikköjen tietovaatimuksia on mallinnettu pieteetillä ja kummassakin tavoitteena on toimiva ydintietojärjestelmä tukemaan yksikköjen hallintaa: Tampereella kaupungin sisäiseen työhön ja PTV:hen tukemaan tulevan Suomi.fi:n tiedonjakokanavaa.

Seuraavassa kuvassa tiivistetään aidosti jaettu, yhteinen osuus yksikköydintiedosta:





Päiväkodit

Ikääntyvien
palvelukeskukset

Uimarannat

1. Yksilöivä tunnus
2. Nimi
3. Vastuullinen juridinen henkilö
4. Rooli(t)
5. Elinkaari

- 1) Yksilöivä tunnus on edellytys yhteiskunnan tasolla jaetulle yksikköydintiedolle, niin kuin se on edellytys luonnollisten henkilöiden tiedoille. Tämän tunnuksen muodostaminen on yksi keskeisimmistä kysymyksistä.
 - a. Tämä esiselvitys argumentoi, että pyrkimys ISO-standardien mukaiseen hierarkiseen yksilöivään organisaatitunnukseen ei ole oikea lähtökohta.
 - b. Hierarkian sisältävän tai sitä indikoivan tunnusteen asemesta yksikön jaetun ydintiedon näkökulmasta tarvitaan ei-rakenteellinen, ei-hierarkkinen, ”flat list”-tyylinen tunnus.
 - c. Käytännössä tarkoittaa, että esimerkiksi ”sairaala” on yksi yksikkö ja sairaalan osasto X voi olla toinen yksikkönsä. Yksikköydintiedon kannalta niillä ei olisi hierarkkista suhdetta. Eri konteksteissa hierarkkinen suhde luonnollisesti olisi ja kontekstiriippuvaisia hierarkioita voi olla useita (kaupungin verkkosivuilla sairaalan päiväpoliklinikka voi olla omassa hierarkiassaan muiden vastaavaa toimintaa tekevien yksiköiden kanssa, vaikka hallinnollisesti ao sairaalan yksikkö olisi osa sairaalaa).
- 2) Nimi. Nimi luonnollisella kielellä tarvitaan jaetun ydintiedon tasolle. Ilman sitä yksiköiden master-luettelon käyttäminen on inhimillisesti vaikeaa.
 - a. Nimi on kuitenkin osin kontekstiriippuvainen. Toisin kuin luonnollisen henkilön kohdalle, yksikön nimelle ei voida asettaa yksyä ”oikeita” standardeja. Eri tarkoituksiin voidaan tarvita erilaisia versioita perustellusti. (esimerkiksi aiemmin kuvassa näkyvän Tampereen esimerkissä tietomalliin on varattu kaksi eripituista nimiversiota)
 - b. ”Yhteisen” nimen käytösäännöt tulee määritellä erikseen. Muut nimiversiot hallitaan niitä tarvitsevilla järjestelmissä.
- 3) Vastuullinen juridinen henkilö on jaetun yksikkötiedon edellytys. Käytännössä yksiköstä on viime kädessä vastuussa joku y-tunnuksellinen taho.





- a. Vastuu ei aina ole yksiselitteinen – näin voi käydä esimerkiksi yhteispalvelupisteessä. Pääasiallinen vastuutaho on kuitenkin voitava aina tunnistaa.
 - b. Erilaiset skenaariot ao vastuista tulee tunnistaa.
- 4) Rooli(t) yksikkökohtaisesti. Yksiköllä voi olla yksi tai useampi rooli. Roolit voivat olla kontekstisidonnaisia, mutta oleellinen kysymys on yhden ylimmän tason roolituksen tekeminen.
- a. Tämä kysymys linkittyy aiemmin mainittuun terminologia- ja ontologiakysymykseen: yksikköjen roolituksen keskeisimmän sanaston yhteinen sopiminen (”päiväkoti” vs ”lastentarha” jne)
 - b. Vaikka yksikön rooli ja ”palvelu”-käsite eivät ole täysin synonyymejä, palveluydintiedon hallinta linkittyy tähän. Palvelujen yhteistä luokittelua tehdään käynnissä olevan JHS-työn kautta parhaillaan keväällä 2015.
- 5) Elinkaari on hankala kysymys. Elinkaari on useimmiten mitä suurimmissa määrin kontekstiriippuvainen ja sen nostaminen ylimmälle jaetulle tasolle on ongelmallista.
- a. Elinkaaren huomioinen jaetulla ydintiedon tasolla vähintään siihen asti, että onko yksikkö aktiivinen vaiko tyystin lakannut toimintansa kuitenkin tarvittaneen.

Pragmaattisista syistä ”jaettuun” ydintietoon voidaan lukea yksikön osoitteet ja todennäköisesti joitain muita attribuutteja.

Lista yllä suppeine perusteluineen osoittaa ”yksikköjen” jaetun ydintiedon ongelmallisuuden. Aidosti kontekstiriippumatonta tietoa on jokseenkin vähän yhteiskunnan ”yksikön” kaltaisessa entiteetissä. Taustasyynä on tietotarpeiden heterogeenisuus: julkinen hallinto sekä siihen nivoutuva yksityinen toiminta on niin monitahoista, että minkä tahansa ydintieto (Master Data) -alueen yksiselitteinen yhteinen, eri tarpeet kattava tietomalli on haastava laatia.

Muu kuin jaettu taso

Jaettu ydintieto on pieni osa yksikköjen tietomassasta. Selkeyden vuoksi todettakoon, että jaetulle ydintiedolle osoitetut säännöt eivät välttämättä koske alemman tason ydintiedon hallintaa. Esimerkiksi argumentti ei-rakenteellisen yksilöivän tunnisteiden osalta ei päde hallinnonalakohtaisesti: kun konteksti on tiedossa, ISO-standardien mukaan rakennetun rakenteellisen OID-tunnuksen systemaattinen käyttö on suositeltavaa.

Pyrkimys standardien noudattamiseen kunkin yksittäisen kontekstin sisällä on ylipäätään suositeltavaa, mikäli soveltuvia standardeja suinkaan on olemassa.





2 Hallinta – Data Governance

Lähdimme äskeisessä kappaleessa liikkeelle tiedon itsensä määrittämisestä. Siinä tehtyjä suosituksia ei voi käytännössä toteuttaa ilman, että erikseen sovitaan yhteisestä omistajuudesta: mikäli yksikköydintietojen (tai muiden ydintietojen – tai edes ydintietojen yhteinen määrittely) ei ole hallinnallisesti kenenkään vastuulla, tilanne on ongelmallinen.

Tällä hetkellä esimerkiksi luonnollisten henkilöiden henkilötunnusten ja nimien hallinta on vastuutettu VRK:lle. Niille on olemassa hallintamalli, määriteltynä ensimmäisessä kappaleissa mainituissa laissa ja asetuksissa.

Yksikkö- tai palvelutietojen ydintiedoille vastaavaa hallintamallia ei ole. Tämä on ollut rajallinen ongelma aiemmin. Yksikkö- ja palveluydintietojen yhä keskeisempi rooli yhteiskunnan digitalisoitumisen mahdollistajana edellyttää kuitenkin hallintamallillisia toimenpiteitä niillekin: jos ja kun päiväkotien ja uimarantojen tiedot halutaan saada yhteismitallisiksi ja yhteisten palvelukanavien kautta näkyviin kansalaisille sekä Helsingissä että Utsjoella, yhteistä hallintamallia tarvitaan.

Hallintamallin pohjustus

Hallintamallia käsitteenä on pohjustettu JHS 179:ssä:

”Data Governance (jäljempänä ydintiedonhallintamalli) koostuu niistä prosesseista, joilla varmistetaan tarvittavien tietojen määrällinen hallinta organisaatiossa. Tällaisia prosesseja ovat esimerkiksi tiedonhallintaprosessit, tiedonhallintalinjausten noudattamisen- ja hyödyntämisen prosessit sekä muutostenhallintaprosessit. Ydintiedonhallintamalli on laadunvarmistusmenetelmä, jolla organisaation tärkeää tietoa huolletaan ja jonka piiriin kuuluvat kaikki ne ihmiset, prosessit ja tietojenkäsittelytekniikat, jotka tarvitaan luomaan yhtenäinen ja hyvä organisaation tiedonhallinta.”

Määritelmä on kattava, mutta ei käytännön tasolla itseään selittävä. Mitä käytännön muutoksia datan hallinnan kautta tulee vastaan? Mitä tulee tehdä, jotta Kansallisen Palveluväylän ja palvelutietovarannon (PTV) kautta saadaan muodostettua aidosti yhteinen pohja datalle, tässä tapauksessa erityisesti yksikköydintiedoille?

Yhteisen hallinnan synnyttämiselle ei ole ratkaisua tällä hetkellä. Kuntien Tieran MDM-viitearkkitehtuuri -raportti koskettaa Data Governance -näkökulmaa kevyehkösti yksittäisten kuntien näkökulmasta. Muut dokumentit sivuavat aihetta, mutta kokonaiskuva ”hallintamallille” ei ole hahmotettu.

Esiselvityksen suositukset koskevat hallintamallin määrittelyä sekä käytännön jalkauttamisen toimenpiteitä.



**Yhteinen vaatii keskittämistä**

Aiemmin selvityksessä käsitelty ulkomaisten parhaiden käytäntöjen benchmarking vahvistaa ajatusta: menestyneillä malleilla on takanaan keskitettyjä organisaatorakenteita ja toimintamalleja, ylipäättään vahva keskitetty näkemys tahtotilasta.

Esimerkiksi Englannissa gov.uk -verkkopalvelun kautta jaetun tiedon yhdenmukaisuus ja toimintamalli on keskitetysti Cabinet Officen alaisuudessa toimivassa Government Digital Service -yksikössä²⁴. Yksikölle on annettu sekä valta että vastuu yli julkisen hallinnon eri tasojen tuottaman tiedon yhdenmukaisuudesta. Yksikkö itsessään on pieni, mutta se pystyy vaikuttamaan kaikkeen yhteiskunnan digitaaliseen kanavaan tuottamaan tietoon.

Vertailukohdista erikseen mainittakoon, että myös suuret yksityisen sektorin organisaatiot koti- ja ulkomailla pyrkivät yhä enemmän katsomaan toimintaansa kokonaisvaltaiset yli prosessien, järjestelmien ja datasiilojen. ”Digitalisaation” hallinnan johtajatasoon roolit ovat myös tulossa.

Kehitys on jo menossa tähän suuntaan monella ICT:n alueella. Perinteiset rajat rikkovia uusia rakenteita tarvitaan myös muualla. Julkisen hallinnon tieto- ja viestintä tekninen toiminto (JulkICT-toiminto) on vuoden 2015 alusta jakautunut kolmeen yksikköön

- Tietopolitiikka
- Digitalisaatio
- Kyberturvallisuus ja infrastruktuuri

Yksiköille on valittu vetäjät. Yksiköillä on roolinsa yli valtionhallinnon.

Analogia ydintietojen hallintaan on selvä. Yhteisten ydintietojen hallinnan johtaminen vaatii keskittämistä. Mahdollinen ”Tietoviraston” perustaminen tulevaisuudessa Valtiovarainministeriön alle on myös mahdollisuus ydintietojen hallinnan tekijäksi.

Organisatorinen paikka ei kuitenkaan ole kynnyskysymys, kunhan yksikön valta ja vastuut saadaan määriteltyä ja yksikölle luodaan lainsäädännön kautta toimintaedellytykset.

Onnistuessaan tuleva ydintietojen hallintamalli, missä organisaatiossa se sijaitseekaan, auttaa luomaan sitoutumista yhteisten ydintietojen hallintaan niin keppien kuin porkkanoidenkin kautta.

²⁴ Yksikön blogi löytyy osoitteesta <https://gds.blog.gov.uk/about/>. Jo blogin etusivu antaa kuvan yksikön viestinnän selkeydestä ja tavoitteenasetannan käytännönläheisyydestä.

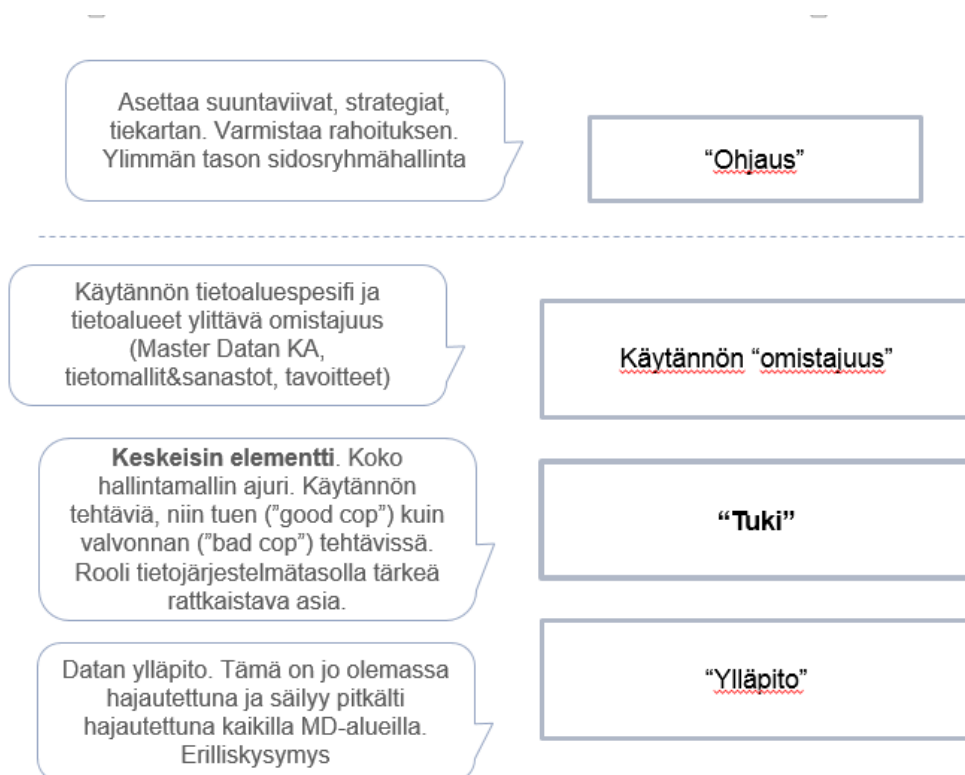




Keskitetyn hallintamallin rajaukset ja vastuut

Keskitettyä ydintietojen hallintaa tarvitaan osaksi ylintä valtionhallintoa. Mitkä ovat keskittämisen rajaukset ja vastuut? Miten hallintamalli nivoutuu esimerkiksi kansallisen palveluväylän tulevaan hallintoon / muuhun hallintoon, tulee määritellä konkreettisten esimerkkien kautta.

Lähdetään liikenteeseen yksinkertaistetusta hallintamallin kuvauksesta. Yksikköydintiedon haltuunotto vaatii neljän tason huomioimista:



"Ohjaus" on ylin keskittämisen taso. Esimerkiksi aiemmin kuvattu yhteiskunnan ydintiedon tietoauesiden määrittely – joko tämän esiselvityksen suosituksen mukaisesti tai muuten – tulee olla keskitettyä ohjattua toimintaa. Jaetun ydintiedon hallinnan suuntaviivat tulee määritellä, niin yksikkö- kuin muillekin tietoauesille. Valtiovarainministeriö – tulevaisuudessa kenties sen alla mahdollisesti oleva "Tietovirasto" on luonteva paikka.

Ohjaustason keskeisin tehtävä on luoda edellytykset varsinaiselle tekemiselle ja ylimmän tason sidosryhmähallinta, ei niinkään suunnitella yksityiskohtaisella tasolla ydintietojen määrittelyä tai hallintaa millään ydintietoauesella.





Käytännön ”omistajuus” on seuraava taso. Omistajuuden tulee olla tietaluekohtaista. Yksikköydintietojen jaetun tason määrittely ja päätöksenteko siitä – jonka rakennetta pohjustettiin edellisessä kappaleessa – on esimerkki omistajuutta vaativasta asiasta.

Omistajuuteen kuuluu mm. tietaluekohtaisen ontologian hallinta. Tehtävä tulee rajata tiukasti koskemaan erityisesti jaetun ydintiedon ja muutoin keskeisimpien käsitteiden ja niiden suhteiden tasoa. Rajanveto muun yhteiskunnan ontologiatyön kanssa voi osoittautua hankalaksi. Ero yleisemmän tason ontologiatyöhön on kuitenkin se, että tietaluekohtaisella omistajuudella on oikeus määritellä termistöä niin, että sen määritelmät ovat velvoittavia. Ilman tällaista oikeutta yksikköydintiedon saaminen yhteiseen hallintaan ei tule onnistumaan.

Omistajuuden organisatorinen sijainti on kysymys, johon tämä esiselvitys ei pyri vastaamaan yksiselitteisesti. Ehdotus on, että omistajuus annetaan ensin pilottiprojektille. Pilottiprojektin kautta löydetään luonteva lopullinen kohde. Selvityksen lopussa olevassa tiekartta-kappaleessa pohjustetaan pilotoinnin merkitystä.

”Tuki” on tulevan hallintamallin kannalta keskeinen kysymys: miten yhteistä yksikköydintiedon hallintaa voi käytännössä ajaa eteenpäin. Kenelle annetaan resurssit ja vastuu esimerkiksi jatkuvaan yksikköydintietojen laadun seurantaan? Yhteiseen toimintatapaan siirtyminen edellyttää myös, että joku ylläpitää ohjeita ja tekee käytännön jalkatyötä. Esimerkiksi aiemmin mainitussa Yhdistyneiden Kuningaskuntien esimerkissä keskitetyllä tukiyksiköllä on merkittävä rooli tässä työssä.

”Ylläpito” tulee yksikköydintietojen osalta olemaan hajautettua. Yksikköjä luodaan ja määritellään yli hallinnonalojen rajojen, pienimmistä kunnista suurimpiin, pienemmistä virastoista suurimpiin.

Hallintamallin käytännön tehtävistä tarkemmin

Rakenteet antavat viitekehyksen toiminnalle. Toiminta itsessään – mitä käytännön tehtäviä siinä on – on mallin toimivuuden kannalta keskeisin kysymys.

Seuraavat taulukot haarukoivat yllämainitun rakenteen osalle tulevia keskeisimpiä tehtäviä esimerkkein avulla.





Alue	Kuvaus	Esimerkki
Tiedon laatu	Tiedon (datan) laadun palvelut ovat keskeisin palvelu, jota koko hallintamalli edellyttää. Yhteinen, poikkiorganisatorinen kyky seurata datan laatua ja kehittää sitä, on hallintamallin pohja. Tällaisen kyvykkyyden syntyminen hajautevasti, kaikkialle (tai edes keskeisiin valtion- ja kunnallishallinnon yksiköihin), ei ole realistista.	Tuki datan harmonisointiin suuressa järjestelmäuudistushankkeessa. Tiedon laadun käytäntöjen integroiminen osaksi yhteistä projektimallia (tällaista ei vielä ole)
Tietomallinnus	Tietomallinnus muuttuu digitalisoitumisen myötä yhä kriittisemmäksi. Keskeisin ja ulospäin selvimmin näkyvä työ liittyy sanastokehitykseen, niin organisaatio- ja toimipaikka kuin muillakin alueilla. Mallinnuksen tukeminen edellyttää keskitetyn kyvykkyyden ja hallintamallin rakentamista.	Tuki sanaston yhtenäistämiseksi / laajentamiseksi jollakin alueilla.
Koulutus	Datan ylläpidon, mutta myös muiden yhteisten käytäntöjen, koulutus- ja viestintä on tukiorganisaation keskeisiä tehtäviä	Datan ylläpitäjien koulutus, jotta organisaatio- ja toimipaikkatietoja syötetään oikein, noudattaen yhteisiä sääntöjä.
Yhteiset työkalut	Yhteiset työkalut ovat mielenkiintoinen alue: rakennetaan-ko niihin kyvykkyyksiä tekniselle tasolle (Valtori?), vai enemmän neuvovalle, yhteisiä käytäntöjä muodostavalle tasolle (tämä hallintamalli). Mille työkaluille?	Toteutukseen tehtäviään, yhteinen hallintaorganisaatio tarvitsee vähintään datan laadun työkalut keskitetysti. Tarvitaanko muita?

Kutakin tasoa tulee ja voi lähteä tarkentamaan:

Alue	Tarkempaa jaottelua
Tiedon laatu	Profilointi - kyky seurata tiedon laatua (sekä pyydettyä "good cop" että valvottaessa "bad cop" roolissa)
Tiedon laatu	Datan laadun parantamisen (puhdistaminen, harmonisointi, rikastaminen) prosessit ja osaaminen.
Tiedon laatu	Järjestelmämigraatiot - prosessit, best-practice-käytännöt
Tietomallinnus	Sanastotyöt (taksonomiat, ontologiat, thesaurukset, kontrolloidut sanastot, folksonomia)
Tietomallinnus	Looginen ja fyysinen tietomallinnus, ER, UML osaaminen
Koulutus	Valtion- ja kunnallishallinnon sisällöntuottajien, järjestelmätyön
Koulutus	Järjestelmätyön koulutus, datanäkökulma
Teknologia	DQ-alusta. Tiedon laadun yhteinen alusta - niin profilointi kuin muihin tiedon laadun kohdassa mainitut
Teknologia	MDM-alusta(t). Jos yhteisiä MD-alustoja päätetään rakentaa, niiden hallinta





Alue	Tarkemmin
Tietomalli	Tietoaluekohtainen yhteinen tietomalli (huomioiden viimeksi keskustelemamme periaatteellisen rajauksen, eli hallinnonala / toiminnalliset alueet kohtaisia tietomalleja tulee säilymään)
Terminologia	Kontrolloidut sanasto, ontologiat, taksonomiat (luokittelut) per tietoalue (jälleen, yhteinen taso).
KA	Kokonaisarkkitehtuuri per tietoaletaso. Myös yli tietoalueiden.

3 Teknologia

Aiemmat ratkaisun kuvauksen pääalueet – ydintiedon itsensä määrittely ja hallintamalli tiedolle – muodostavat tarvittavat lähtökohdat teknologiakeskustelulle. Vaikka tietotekniikkakulma edellä ei tule lähteä liikkeelle, käytännössä tietotekniikka on ydintietojen menestyksellisen hallinnan edellytys.

Perustavanlaatuinen kysymys on, mitä teknologialla tarkoitetaan yhteisten ydintietojen tapauksessa. Mitä tulee lähteä ratkomaan?

Asiaa ei helpota, että ydintietojen hallinta ei ole yksittäinen teknologia. Linkittyviä teknologioita ovat mm.:

- Varsinaiset ydintiedon hallinnan MDM (Master Data Management) -ratkaisut – ja näissäkin on erilaisia arkkitehtuurivaihtoehtoja²⁵
- Datan laadun (Data Quality) teknologiat
- Sanastojen hallintaan (Business Glossary) keskittyvät teknologiat
- Metadatan hallintaratkaisut
- Liiketoimintaprosessien hallinnan teknologiat, nk. BPM (Business Process Management)
- Datan virtualisointiratkaisut

Ydintietojen hallinnan ratkaisuissa sivutaan osaa tai kaikkea näistä. Perinteiset integraatiodomainiin kuuluvat ETL- ja EAI-teknologiat kytkeytyvät myös vahvasti ydintietojen hallintaan. Niin ikään rajanvedot eri teknologioiden välillä voivat vaihdella: esimerkiksi monet MDM-ratkaisut sisältävän pitkällekin vietyjä datan laadun toiminnallisuuksia ja liiketoimintaprosessien hallinnan ja teknologioita.

Tarpeet eri ydintiedon hallinnan kerroksissa ovat myös erilaiset: ydintietojen jaetulla tasolla teknologia-tarpeet eivät välttämättä ole yhteensopivia yksittäisen hallinnonalan ydintiedon hallinnan tarpei-

²⁵ Gartnerin MDM-arkkitehtuuri- jaottelua saatavilla avoimesti.

http://mitiq.mit.edu/iqis/documents/cdoiqs_201077/papers/02_12_2c-2.pdf. Myös muita kaupallisia arkkitehtuurilähestymisiä eri toimijoilta olemassa, esimerkiksi IBM:n näkemyksiä osoitteessa <http://www.ibm.com/developerworks/data/library/techarticle/dm-0804oberhofer/>. Laajasti hyväksyttyä ei-kaupallista ydintietojen hallinnan referenssiarkkitehtuuri- jaottelua ei ole.

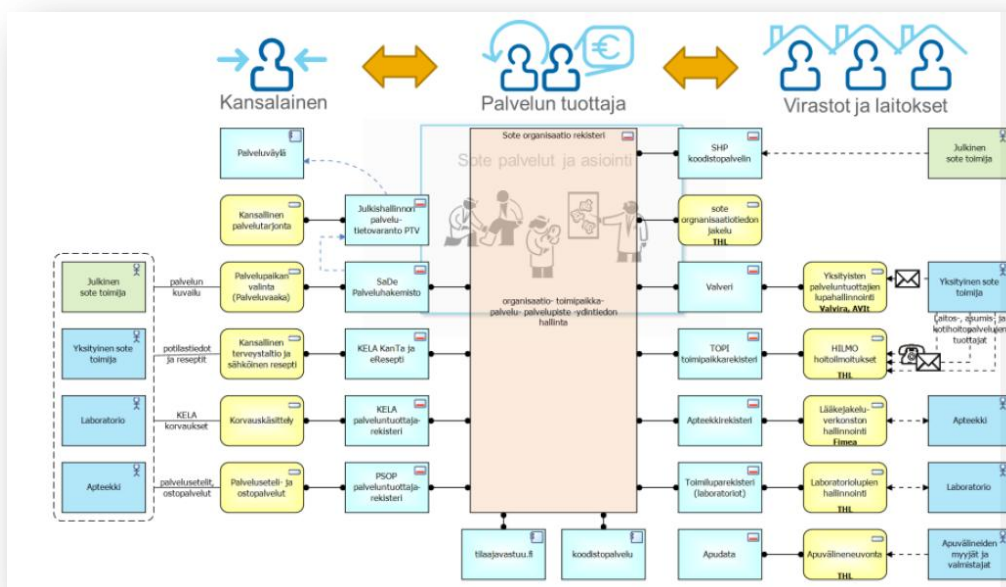


den kanssa. Kaikki kerrokset, kaikki tietalueet kattavaa yhteistä viitearkkitehtuuria ydintietojen hallinnalle ei ole tarkoituksenmukaista lähteä rakentamaan.

Tässä kappaleessa käsitellään ensin jaettua ydintietoa yksikkötiedon näkökulmasta. Sitten siirrytään hallinnonalakohtaiseen ydintietoon. Lopulta teknologialehestyminen vedetään yhteen, edelleen yksikköydintieton näkökulmasta.

Yksikköydintiedot – hallinnonala- ja järjestelmäkohtaisesti

Jaetun tason kuvaamisen asemesta (yksikkö)ydintietojen hallintaa voi lähteä kuvaamaan hallinnonalakohtaiselta tasolta. Hyvä esimerkki tällaisesta hallinnonalakohtaisesta lähestymisestä on STM:ssä suunnittelun alla vuodenvaihteessa 2014-2015 oleva MDM-työ:



Kuva 4 STM:n vdintietosuunnitelma

STM:n tavoite on löytää ratkaisu oman hallinnonalansa ydintietojen hallinnalle: miten virastot ja laitokset pystyvät ylläpitämään ydintietoja keskitetyn kanavan kautta, jotta tieto on ajantasaista kansalaisille asti. Tavoitetta lähestytään systemaattisesti. Valmistelun alla oleva ehdotus on kattava.

STM – tai muu yksittäinen hallinnonala – ei kuitenkaan voi toimia yksin: aivan niin kuin tällä hetkelläkin hallinnonalakohtaisissa järjestelmissä luotetaan luonnollisten henkilöiden osalta VRK:n henkilörekisterissä hallitsemiin henkilötunnuksiin yksilöivinä tunnisteina, myös *yksikköjen* osalta tarvitaan yhteisen jaetun tason ydintiedon määrittäykset. Mikäli niitä ei ole, STM (ja muut hallinnonalat) luovat





omat yksikkökohtaiset tunnisteensa ja esimerkiksi nimeämiskäytäntönsä. Tavoite yhteentoimivuudesta ei tule toteutumaan.

Tästä päästään jaettuun ydintietojen hallinnan tasoon.

Yksikköydintietojen hallinta – jaettu taso

On helpompi lähteä liikkeelle mitä *ei* haeta jaetulta yksikköydintiedolta: sellaista keskitettyä ratkaisua, jossa jossa MDM-järjestelmä toimisi palvelurajapinnan yli reaaliaikaisena haku- ja lukutransaktioiden kohteena yksikkötietoa tarvitseville järjestelmille, ei voi lähteä hakemaan.

Tällaisen tekeminen aiheuttaisi sekä single-point-of-failure -mahdollisuuden, joka ei ole suotavaa, että myös edellyttäisi epärealistisen määrän muutostyötä kaikissa ao. palvelua käyttävissä järjestelmissä.

Keskitetty yhteinen järjestelmä – virallinen ”master” – kuitenkin tarvitaan. Tietoa käsittelevässä kapaleessa mainittu yksikköjen yksilöivän tunnuksen generointi sekä aidosti jaettujen yhteisten tietojen – joiden määrä on hyvin rajallinen – tulee sijaita yhdessä järjestelmässä.

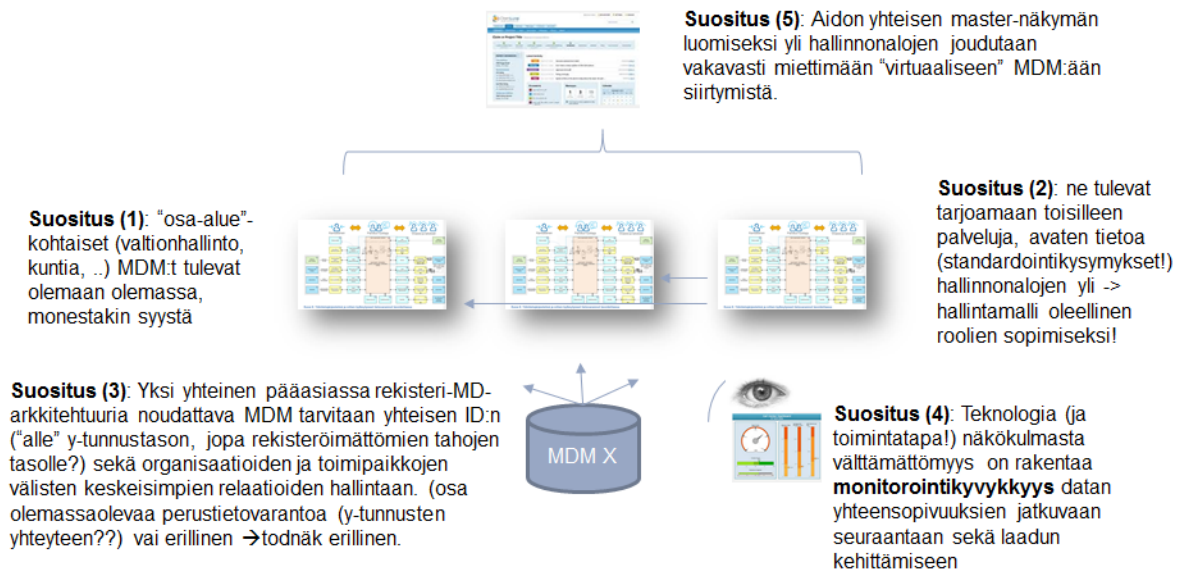
Järjestelmäjähtymistä voi kuvata seuraavien suositusten kautta (numerot viittaavat seuraavan sivun kuvaan):

- Osa-alueilla (hallinnonala-, järjestelmäkohtaisesti) tulee jatkossakin olla erilliset ydintietolähestymisensä (1). Esimerkiksi STM:n tai TEM:in yksikköydintietohankkeet ovat kumpikin vaativia jo omassa kontekstissaan. Ei ole suositeltavaa yrittää yhdistää näitä (ja muita) yhdeksi valtionhallinnon laajuiseksi ydintietohankkeeksi.
- Osa-aluekohtaisien ydintietoratkaisujen tulee kuitenkin tukea avoimia rajapintoja ja tarjota ydintietoaan myös muiden käyttöön (2).
- Kukin osa-alue ei kuitenkaan voi rakentaa kaikista perustavanlaatuisinta, ”jaettua”, yksikköydintietoa. Tätä varten tarvitaan keskitetty ydintietojen hallintaratkaisu (3).
- Yhteistä kyvykkyyttä ja yhteisiä hallintamalleja tarvitaan. Tämä asettaa vaatimuksia myös muille kuin itse ydintiedon hallintateknologioille. Näistä erikseen nostetaan esille (4) datan laadun monitorointi- ja kehitysohjelmistot. Kun ydintietojen hallinta tulee säilymään jaettujen ydintietojen ulkopuolella hajautettuna (ja myös jaetuista ydintiedoista käytännössä on useita replikoita, jotka eivät välttämättä ole reaaliaikaisesti synkronoitu), on keskitetyn hallintamallin toteuttajilla oltava mahdollisuus ja kyky tarjota dataan liittyviä palveluita. Näiden palvelujen hajauttaminen yli hallinnonalojen ei ole perusteltua.
- Viimeisenä MDM-teknologiaa sivuavana suosituksena mainittakoon yhteisen master-näkymän muodostamisen virtualisointi aidon yhteisen kaiken kattavan ydintietolähestymisen asemesta (5). Tämä on johtopäätös edellisistä suosituksista: kun vain pieni osa yksikköydintiedosta tulee olemaan yhteisessä jaetussa järjestelmässä, mutta todennäköinen tarve on nähdä yksikköjä yli hal-





linalojen (esimerkiksi mitä kaikkea lupia / valvontakohteita yksilöllä X on), on olemassa teknisiä ratkaisuja luoda virtuaalinen master-näkymä.





Toteutus ja tiekartta

Tilaus ratkaisulle yksikköydintietojen ottamiselle haltuun on kiistaton:

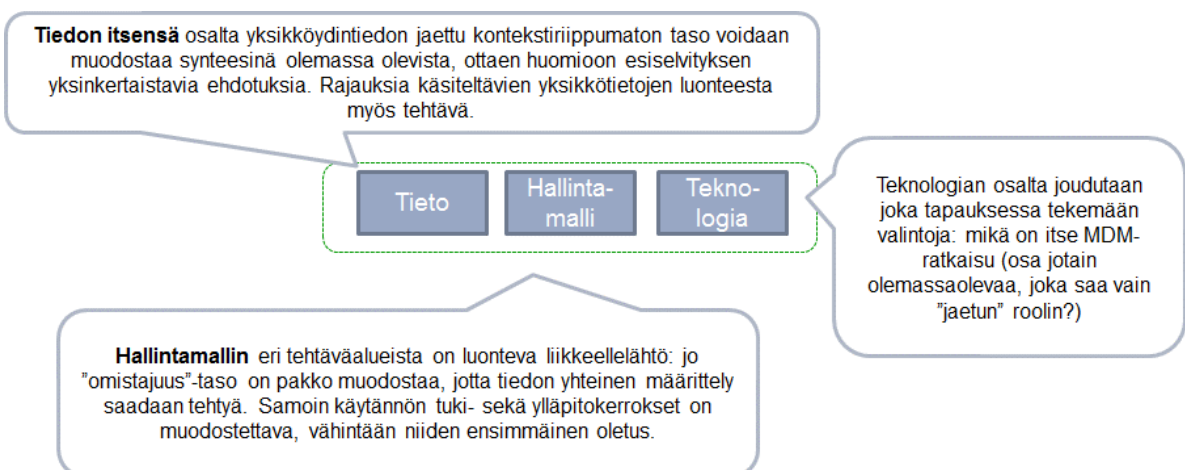
- Poliittinen tahtotila on olemassa: tietohallintolaki korostaa yhteentoimivuuden merkitystä – ja ydintietojen yhteinen hallinta on tämän kehityksen yksi pilareista;
- Palveluväyläkehitys on käynnissä: yhteentoimivuuden yhä paremmin mahdollistavat teknologiat ovat tulossa ja ne tulevat tarvitsemaan (yksikkö)ydintietojen hallintaa tuekseen;
- JulkICT:n organisoituminen vuodenvaihteessa 2015 digitalisoitumiskehityksen ympärille osoittaa myös samaan suuntaan; ja
- Tietoisuus ja tahtotila ratkaisun löytymiselle ovat käsin kosketeltavia.

Vaikka ratkaisulle on tilaus ja intuitiivisestikin ”yhteentoimivuus” on kannatettava ajatus, niin nykyisten instituutioiden pitkä historia hidastaa yhteisten ydintietojen, myös yksikkötiedon, haltuunottoa. Se, että uimaranta Porissa, päiväkotit Tampereella ja vanhainkoti Imatralla kaikki olisivat yhteisten ydintietojensa osalta samalla tavalla hallittuja, on pitkän polun takana.

Peli ei kuitenkaan ole menetetty. Vaikka mahdollisuudet yhtäaikaisen valtakunnalliseen muutokseen – jossa kaikki Suomen ”yksiköt” otetaan kerralla haltuun – ovat olemattomat, vaihteittainen ratkaisu on mahdollinen.

Käynnissä olevista yksikköydintietoa käsittelevistä tai sivuavista hankkeista on mahdollista poimia osajoukko, joiden kautta tässä esiselvityksessä pohjustettua lähestymistä ja sen suosituksia sovelletaan.

Tampereen kaupungin, STM:n, TEM:in ja PTV:n hankkeet sivuavat ja täydentävät toisiaan. Pilotti-luontoisen Proof-of-Concept (PoC) lähestymisen voi muodostaa seuraavasti:





Lähtökohdat itse **tiedon** määrittelyn osalta ovat hyvät: pohjatyötä tietomalleista ja attribuuttitason määrittämisestä on tehty jo listatuissa hankkeissa yksikköydintiedoista. Käytännönläheinen synteesi näistä sekä tämän esiselvityksen suosituksista (erityisesti ei-hierarkkisesta yhteisestä yksilöivästä tunnuksesta) antaa riittävän pohjan yksikköydintiedon pilotoinnille.

Hallintamallin osalta pilottiluontoinen lähestyminen on haasteellista. Aitojen ohjaus- ja omistajuuskerrosten rakentaminen kevyesti on sikälikin vaikeaa, että työ voi vaatia tietohallintolain soveltamista. Ainakin jos omistajuuskerroksen suosituksille halutaan sitovuutta.

Käytännön kannalta de facto -omistajuus (vaikka de jure -omistajuutta ei saataisikaan) kuitenkin tulee voida rakentaa. Kansallinen palvelutietovaranto PTV voisi tällaisen roolin ottaa: PTV:llä on kuitenkin suuri intressi saada myös palveluita toteuttavat yksiköt hallintaansa. Sikälikin PTV on luonteva, että *palvelut* itsessään on ehdotettu toiseksi ydintietetoalueeksi.

Teknologia työn osalta valintoja joudutaan joka tapauksessa tekemään jo pilotoissa. Yksi järjestelmä joutuu ottamaan ”master”-järjestelmän roolin ja myöntämään uuden ei-rakenteellisen yksilöivän tunnuksen yksiköille. Tämä esiselvitys ei ota kantaa, mikä tämä järjestelmä olisi.

Kaiken kaikkiaan toteutus on niin monivaiheinen polku, että tiekartta yhteisiin (yksikkö)ydintietoihin siirtymiseen tulee laatia erikseen

Toteutus ja lainsäädäntö

Vuonna 2011 voimaan tullut tietohallintolaki²⁶ tavoittelee julkisen hallinnon tietojärjestelmien yhteentoimivuutta. Lain tavoitteena on ”tehostaa julkisen hallinnon toimintaa sekä parantaa julkisia palveluja ja niiden saatavuutta säätämällä julkisen hallinnon tietohallinnon ohjauksesta ja tietojärjestelmien yhteentoimivuuden edistämisestä ja varmistamisesta.”

Tähän kuuluu olennaisesti kehittää ”tietojärjestelmien teknistä ja tietosisällöllistä yhteentoimivuutta muiden julkisen hallinnon viranomaisten tietojärjestelmien kanssa silloin, kun järjestelmät käyttävät samoja tietoja”, lain 3§:ää lainaten.

Vaikka lain asettama tavoite tukee (yksikkö)ydintietojen kurinalaista hallintaa, tietohallintolakia on arvosteltu voimakkaastikin²⁷ sen käytännön soveltamisen vaikeuden kannalta.

Käytännön soveltaminen nousee merkittäväksi kysymykseksi: hallintamallin ”omistajuustasolla” tietotoaluekohtaisesti tulee luoda mahdollisuus muiden yhteiskunnan toimijoiden velvoittamiseen, ei vain ohjaamiseen.

²⁶ laki julkisen hallinnon tietohallinnon ohjauksesta (634/2011)

²⁷ <http://www.tivi.fi/Uutiset/2013-11-10/Professori-tietohallintolaki-ep%C3%A4onnistui---tehty-vaikeammaksi-kuin-perustuslain-muuttaminen-3205512.html>





23.3.2015

Lopuksi

Jos ja kun yhteiskunta haluaa jakaa johdonmukaisesti tietoa niin palveluistaan kuin niitä toteuttavista yksiköistä nykyaikaisten digitaalisten kanavien kautta, niin yksikkö- kuin myös palveluydintietojen tulee olla nykyistä paremmin hallussa. Ydintietojen kurinalainen hallinta on yhteentoimivuuden edellytys.

Tämä esiselvitys lähti purkamaan nykytilannetta ja jo tehtyä taustatyötä. Näiden kautta pyrittiin tunnistaman juurisyy sekä rakentamaan tarvittavan ratkaisun keskeiset elementit.

Esiselvitys ei vastaa kaikkiin kysymyksiin – kysymyksiä on liikaa ja ne ovat komplekseja. Monta langanpäättää jää pahasti auki.

Jatkoselvitysten suosittelemisen asemesta esiselvityksen lopussa korostettakoon mieluummin liikkeelle lähtemistä edellisessä kappaleessa kuvatun Proof-of-Conceptin kautta: pilottiluontoinen konkreettinen tekeminen, jossa pyritään muodostamaan jaetulle yksikköydintiedolle pohjakäytäntöjä, vie asiaa eteenpäin.

Pilottiluontoinen tekeminen ei saa kuitenkaan jäädä yksin. Rakenteellisia muutoksia – joita hallintamallikappaleessa pohjustettiin – joudutaan tekemään. Keskitettyä omistajuutta, keskitettyjä toimintamalleja, tarvitaan. Keskittämisen rajat tulee kuitenkin tunnistaa. Suuri osa tiedonhallinnasta tulee säilymään hajautettuna.

Muutoshistoria

Versio	Päiväys	Tekijä	Tarkastaja	Hyväksyjä	Muutoshistoria
1.0	23.3.2015	Kimmo Kontra			Ensimmäinen julkaistu versio

