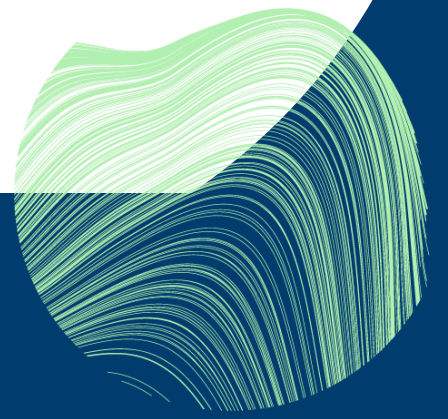
DigiFinland

VIRTA-arkkitehtuuri

Sisältö

- VIRTA-arkkitehtuuri
- Pseudonymisointi ja anonymisointi

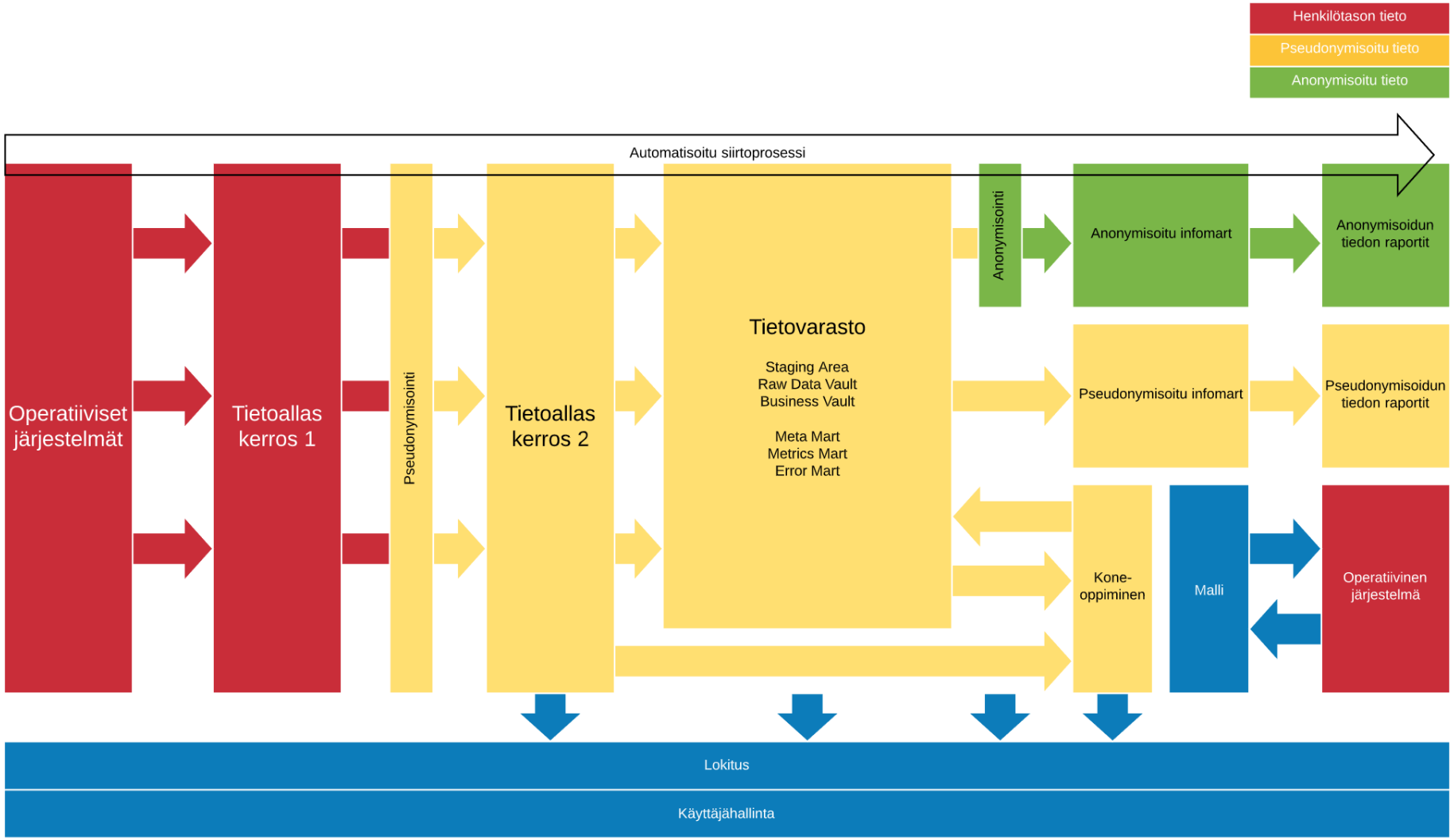
VIRTA-arkkitehtuuri



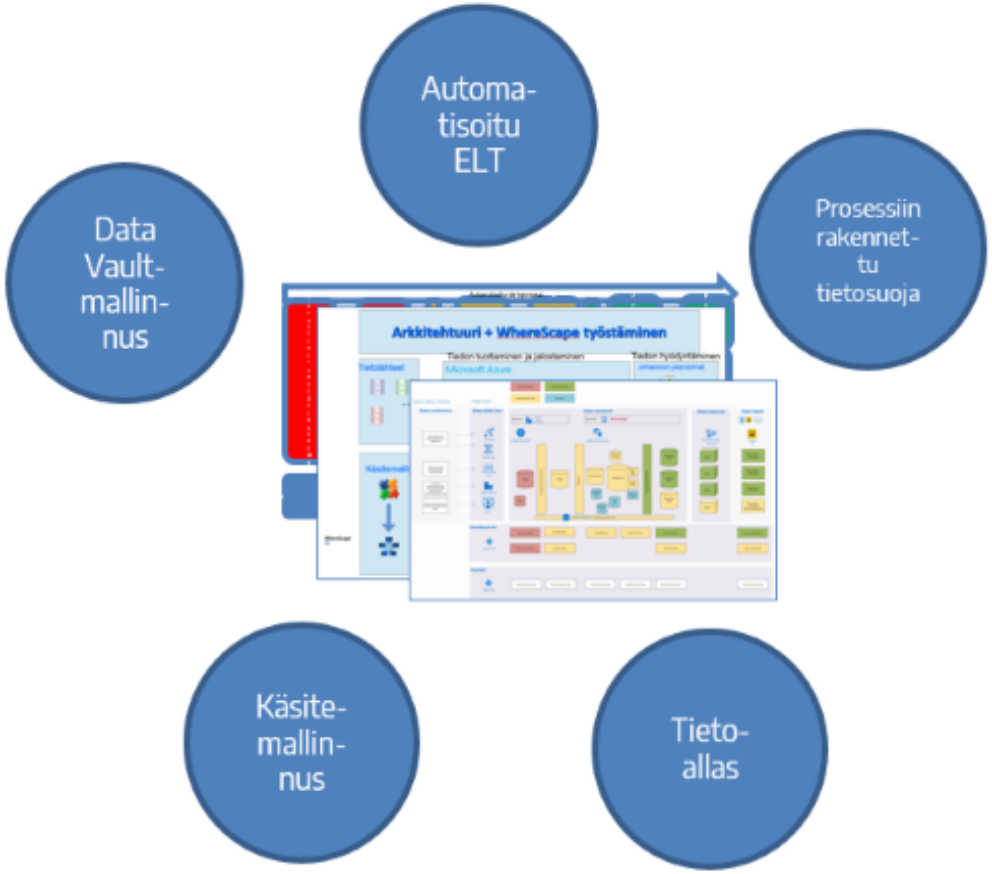
VIRTA-arkkitehtuurin perusteet

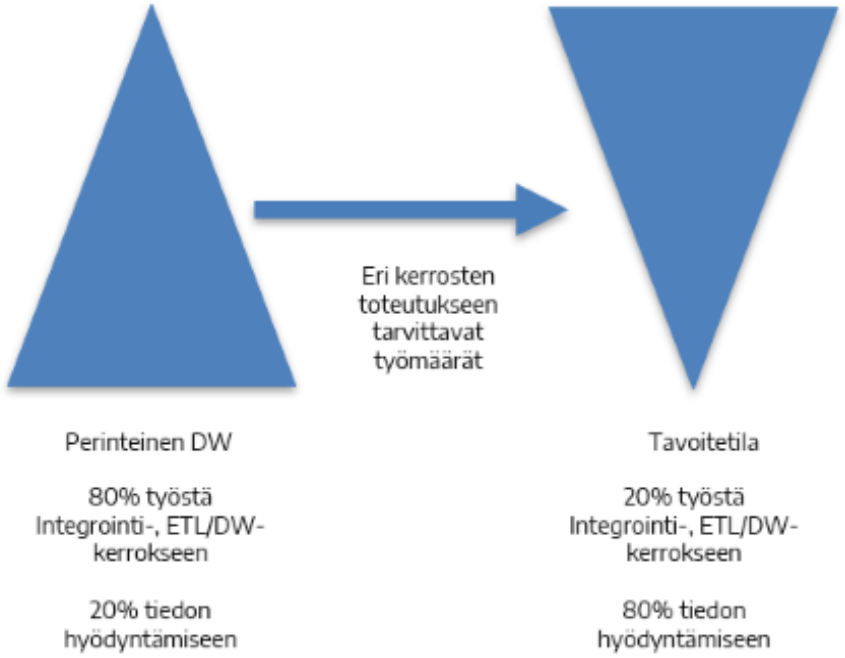
- VIRTA-arkkitehtuuria lähdettiin kehittämään 2018 osana Vimana Oy:n ja SoteDigi Oy:n yhteistä maakuntien tietojohtamishanketta
- Arkkitehtuuriin vaikuttivat olennaisesti juuri voimaantullut GDPR ja silloin vielä työn alla ollut laki sosiaali- ja terveystietojen toissijaisesta käytöstä
- Haluttiin tehdä tietojohtamisen arkkitehtuuri, jossa otetaan huomioon alusta saakka henkilötietojen käsittely
- Seuraavalla sivulla esitelty arkkitehtuuri on viitearkkitehtuuri, joka ei ota kantaa mitä komponentteja tiedolla johtamisen järjestelmässä pitää olla, vaan miten tietoja kannattaisi käsitellä (toiminta tärkeämpää kuin osat ja mitä missäkin tapahtuu)

VIRTA-arkkitehtuuri



Arkkitehtuuriperiaatteet ICT näkökulmasta

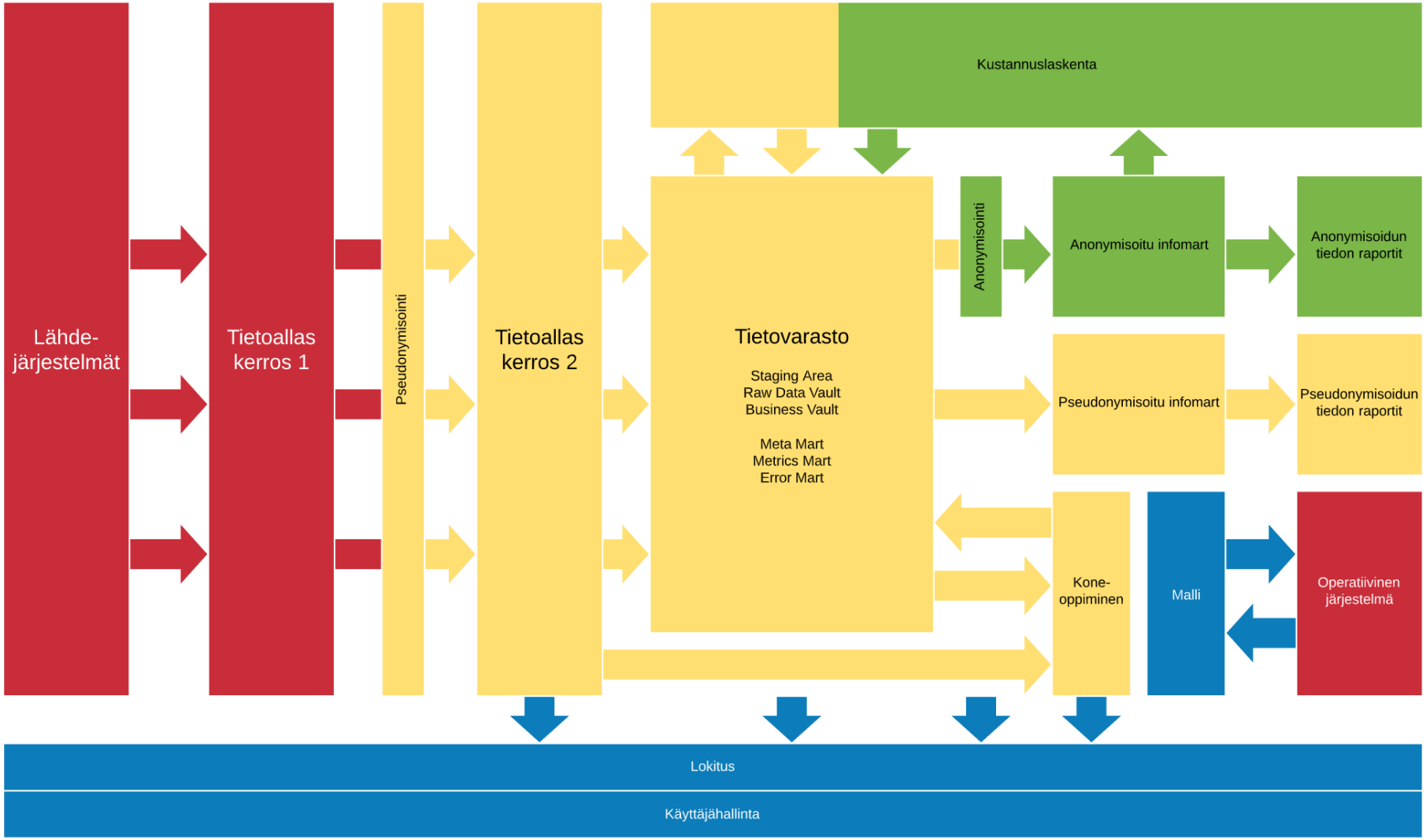




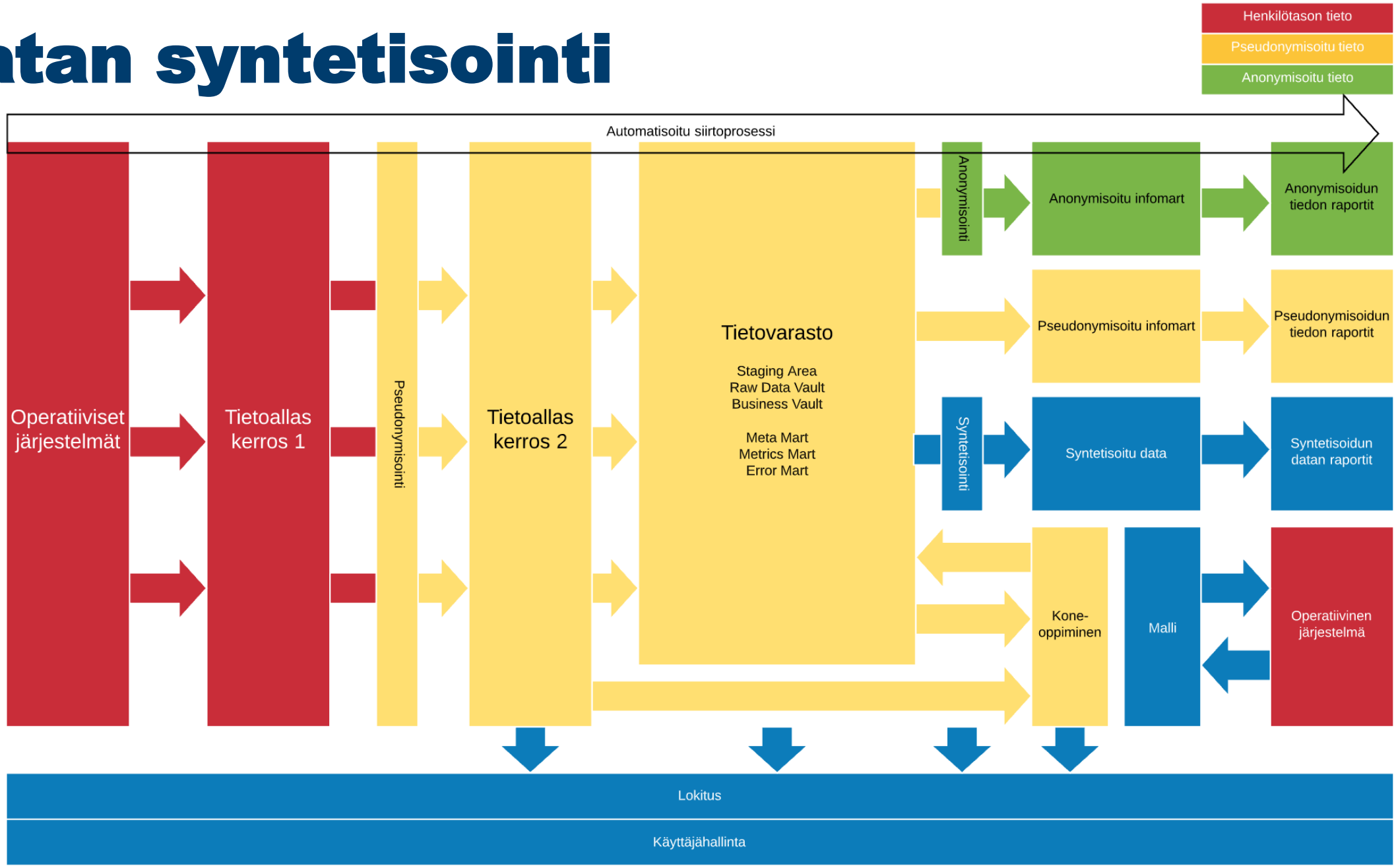
Viitearkkitehtuurin oleellisimmat vaatimukset

- Viitearkkitehtuuri yhdellä lauseella:
  - ”Kansalliseen käsitemallinnukseen pohjautuva tietoturvallinen toisiokäytön lain vaatimukset huomioiva automaattinen model-driven tietovarasto”
- Kansallinen käsitemallinnus:
  - DigiFinlandin Virta-hankkeen kanssa yhteensopiva mallinnus
  - Aluekohtaisissa projekteissa otetaan pohjaksi Virta-hankkeen käsitemallit
  - Sen pohjalta tehdään aluekohtainen mallinnus, keskustellaan ja käydään läpi DigiFinlandn kanssa lisäykset, laajennukset ja muutokset, jotka tulleet esille aluekohtaisessa mallinnuksessa
  - DigiFinland ottaa omaan, kansalliseen käsitemalliin mukaan alueilla kehitettyjä osia
  - Yhteensopivia malleja on kehitetty / kehitteillä: Ikäihmiset (Tampere, KeuSote, Siun sote, joiden pohjalla Virta-hankkeen ikäihmisten malli), Lapsiperheet (Espoo, pohjalla yhteiset käsitteet Virta-hankkeen käsitemallista), Pohjanmaa, Kymsote, PHHYKY, Pirkanmaa
  - Eri alueet voivat hyödyntää valmiita DigiFinlandin malleja
- Tietoturvallinen toisiokäytön lain vaatimukset huomioiva
  - Tietojen suojaamisen eri vaiheet: pseudonymisointi, anonymisointi, jäännösriski
  - Pääsyn hallinta, roolipohjaisuus
- Automaattinen model-driven tietovarasto
  - Data vault 2.0-tietomallinnuksen mukainen tietovarasto generoidaan automaattisesti muodostetusta käsitemallista
  - Tiedot mäpätään lähdejärjestelmiin ja latausohjelmat muodostetaan automaattisesti
  - Välineistö on yleisesti käytettävissä useiden toimittajien tukema (ei toimittajakohtainen toteutus)
  - Osaamisen siirto, kouluttautuminen mahdollista toisille toimittajille tai asiakkaalle

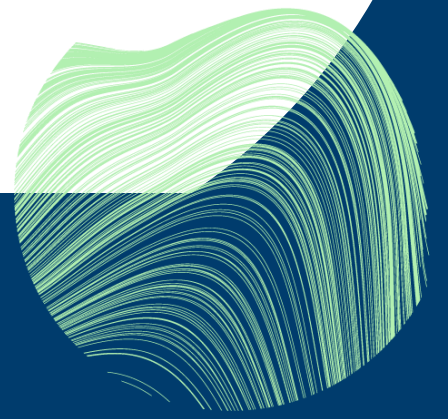
Kustannuslaskenta mukana



Datan syntetisointi



Pseudonymisointi ja anonymisointi



Sisältö

- Mitä pseudonymisointi ja anonymisointi tarkoittavat?
- Pseudonymisointi teknisesti
- Anonymisointi
- Miksi tehdään?
- Pseudonymisointi ja anonymisointi tiedolla johtamisen arkkitehtuurissa
- Anonymisoinnin toteutus käytännössä
- Jäännösriskin arviointi
- Julkaistavan tiedon suunnittelu vs. anonymisointi

Pseudonymisointi

- Pseudonymisointi tarkoittaa henkilötietojen käsittelemistä siten, että henkilötietoja ei voida enää yhdistää tiettyyn henkilöön ilman lisätietoja. Tällaiset lisätiedot täytyy säilyttää huolellisesti erillään henkilötiedoista.
- **Vaikka tiedot olisi pseudonymisoitu, niiden avulla yksilö voidaan edelleen erottaa joukosta ja yhdistää eri tietoaineistoissa.**
- Pseudonymisoidut tiedot ovat yhä henkilötietoja, ja niiden käsittelyssä on sovellettava tietosuojasäännöksiä.
- Esimerkiksi henkilötietojen koodaaminen on pseudonymisointia. Koodattuja tietoja ei voida yhdistää tiettyyn henkilöön ilman koodiavainta. Koodiavaimen haltija voi kuitenkin purkaa tietoaineiston salauksen ja helposti tunnistaa uudelleen jokaisen rekisteröidyn.
- Henkilötiedot voidaan myös suojata peitenimillä. Esimerkiksi tietokannoissa jokin henkilöön liittyvä tieto voidaan korvata toisella. Pseudonymisointi onkin tavallista muun muassa tutkimustoiminnassa ja tilastoinnissa.

Lähde: Tietosuojavaltuutetun toimisto

Anonymisointi

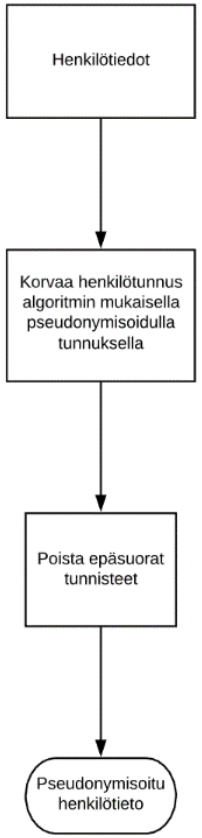
- Anonymisointi tarkoittaa henkilötietojen käsittelyä niin, että henkilöä ei enää voida tunnistaa niistä.
  - Karkeistaminen yleiselle tasolle (aggregointi)
  - Muuttaminen tilastolliseen muotoon,etteivät yksittäistä henkilöä koskevat tiedot ole enää tunnistettavassa muodossa
  - Tunnistamisen **täytyy estyä peruuttamattomasti** ja siten, että rekisterinpitäjä tai muu ulkopuolinen taho ei voi enää hallussaan olevilla tiedoilla muuttaa tietoja takaisin tunnistettaviksi.
- Anonymisoinnissa on otettava huomioon kaikki kohtuudella toteutettavissa olevat keinot, joiden avulla tiedot voitaisiin muuttaa takaisin tunnisteellisiksi.
  - Tunnistamisen mahdollisuutta arvioitaessa täytyy huomioida
    - tunnistamisesta aiheutuvat kulut
    - tunnistamiseen tarvittava aika
    - käytettävissä oleva teknologia
    - Rekisterinpitäjän on varauduttava myös siihen, että kerran tehty anonymisointi voi heiketä ajan ja teknisen kehityksen myötä
- **Anonymisoituja tietoja ei enää katsota henkilötiedoiksi. Niihin ei sovelleta tietosuojasäännöksiä.**
- Se, voidaanko jokin tieto lopulta katsoa anonyymiksi vai ei, edellyttää tapauskohtaista arviointia
  - Henkilö voi olla tunnistettavissa muutenkin kuin nimen perusteella
  - Pelkästään nimien ja muiden yksilöintitietojen poistaminen ei siis aina tarkoita, että kaikki henkilörekisterin tiedot muuttuisivat anonyymeiksi
  - Kerättävä aineisto voi sisältää henkilöstä yksityiskohtaista tietoa (esim. harvinainen sairaus tai tarpeeksi tietotyyppejä) siten, että hän on välillisesti tunnistettavissa.

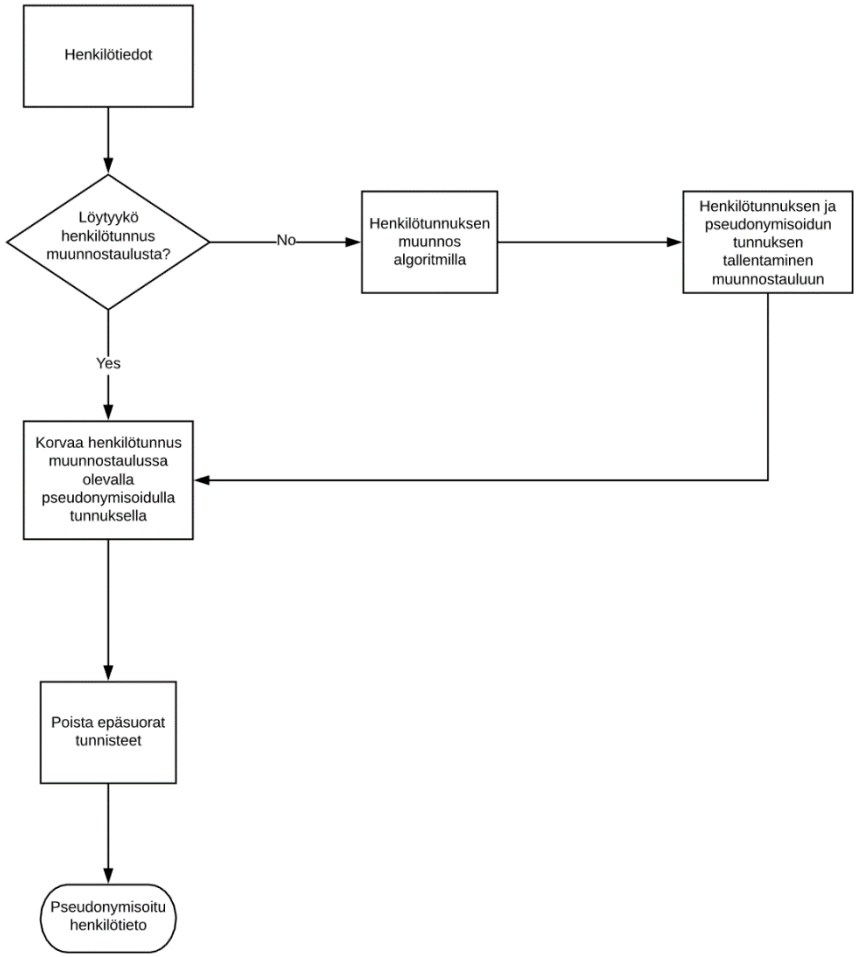
Lähde: Tietosuojavaltuutetun toimisto

Pseudonymisointi teknisesti

- Teknisesti henkilötunnus muunnetaan ajamalla se esimerkiksi hash-funktion läpi, jonka jälkeen siitä tulee ns. tiiviste, jota ei pysty palauttamaan takaisin päin:
  - 010101A001A → MD5 → 0x3C15CC0713408C26D28A4EBCDBCEBEB3
  - 010101A001A → SHA256 → 0x07611AF932EE54BE3100E8B814F7153CC08DCF2EE3189F7E1D7817A7A2612FDE
- Koska suomalaisen henkilötunnuksen muoto on vakio ja yleisesti tunnettu, niin tällä tavalla pseudottu henkilötunnus on löydettävissä helposti ajamalla kaikki mahdolliset henkilötunnukset algoritmin läpi
- Vaikeutta saadaan lisää kun laitetaan henkilötunnukseen mukaan ”suola”:
  - 010101A001A + Salainen avain MD5 → 0xA6BC3037ADCC998BCA70DCB3B46E1D85
  - 010101A001A + Salainen avain SHA256 → 0xC82BE29D20EE67B7F20618B85586B568DD4F71752CFC0F77B8585283482EE948
- Käyttämällä samaa suolaa ja algoritmia kaikkien järjestelmien kohdalla henkilötunnukset saadaan pseudonymisoitua jokaisen järjestelmän osalta samalla tavalla ja tiedot voidaan yhdistää toisiinsa

Yksi- vai kaksisuuntainen pseudonymisointi?





Miksi tehdään?

- Vaikka tiedot ja niiden käsittely suojataan käyttöoikeuksilla ja sopimuksilla, niin silti on syytä noudattaa yleistä varovaisuutta kaiken henkilötiedon kanssa
- Työtehtävien hoidossa pitää olla mahdollista käsitellä tietoja laajasti, mutta henkilötietojen näkyvyyttä pitää rajata ”näytetään vain minimitietot”-periaatteen mukaisesti
- Vahinkoja voi sattua ja silloin on hyvä, että ollaan vahingot ovat rajallisia
- Käytännössä lait ja asetukset rajoittavat:
  - GDPR (https://tietosuoja.fi/gdpr)
  - Laki sosiaali- ja terveydenhuollon asiakastietojen sähköisestä käsittelystä (https://finlex.fi/fi/laki/ajantasa/2007/20070159)
  - Laki sosiaali- ja terveystietojen toissijaisesta käytöstä (https://finlex.fi/fi/laki/alkup/2019/20190552)

GDPR

- Tietosuoja-asetuksessa on kuusi eri perustetta, joilla henkilötietojen käsittely on mahdollista:
  - rekisteröidyn suostumus
  - sopimus
  - rekisterinpitäjän lakisääteinen velvoite
  - elintärkeiden etujen suojaaminen
  - yleistä etua koskeva tehtävä tai julkinen valta
  - rekisterinpitäjän tai kolmannen osapuolen oikeutettu etu.
- Tietosuojaperiaatteiden mukaan henkilötietoja on
  - käsiteltävä lainmukaisesti, asianmukaisesti ja rekisteröidyn kannalta läpinäkyvästi
  - kerättävä ja käsiteltävä tiettyä, nimenomaista ja laillista tarkoitusta varten
  - kerättävä vain tarpeellinen määrä henkilötietojen käsittelyn tarkoitukseen nähden
  - päivitettävä aina tarvittaessa: epätarkat ja virheelliset henkilötiedot on poistettava tai oikaistava viipymättä
  - säilytettävä muodossa, josta rekisteröity on tunnistettavissa ainoastaan niin kauan kuin on tarpeen tietojenkäsittelyn tarkoitusten toteuttamista varten
  - käsiteltävä luottamuksellisesti ja turvallisesti.

Lähde: Tietosuojavaltuutetun toimisto

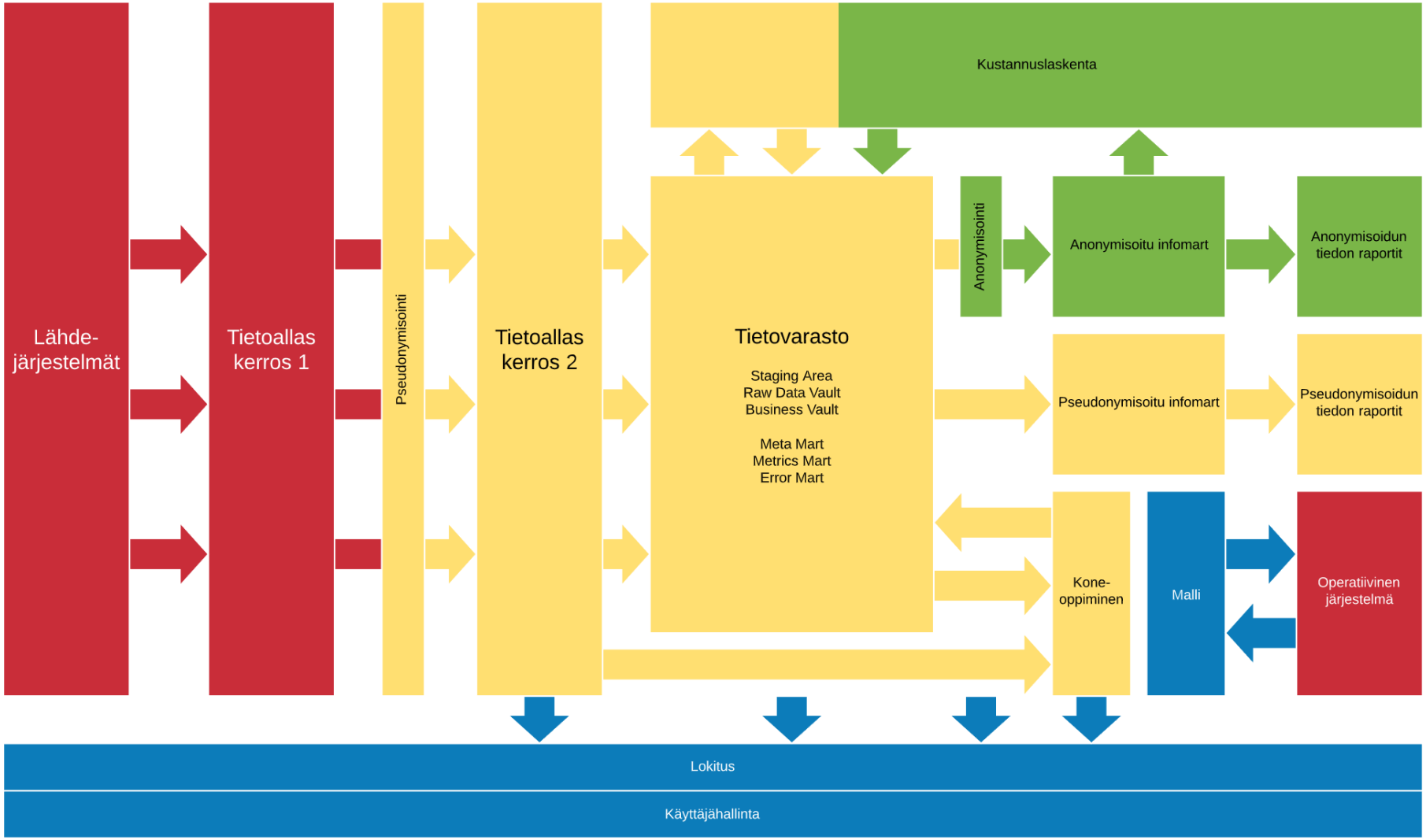
Laki sosiaali- ja terveydenhuollon asiakastietojen sähköisestä käsittelystä

- 5 §
  - Käytön ja luovutuksen seuranta
  - Sosiaalihuollon ja terveydenhuollon palvelujen antajan tulee pitää rekisteriä omien asiakastietojärjestelmiensä ja asiakasrekisteriensä käyttäjistä sekä näiden käyttöoikeuksista.
  - Palvelujen antajan **tulee kerätä asiakasrekisterikohtaisesti kaikesta asiakastietojen käytöstä ja jokaisesta asiakastietojen luovutuksesta seurantaa varten lokitiedot lokirekisteriin**. Käyttölokirekisteriin tallennetaan **tieto käytetyistä asiakastiedoista**, siitä palvelujen antajasta, jonka asiakastietoja käytetään**, asiakastietojen käyttäjästä, tietojen käyttötarkoituksesta ja käyttöajankohdasta.** Luovutuslokirekisteriin **tallennetaan tieto luovutetuista asiakastiedoista**, siitä palvelujen antajasta, jonka asiakastietoja luovutetaan, asiakastietojen luovuttajasta, tietojen **luovutustarkoituksesta, luovutuksensaajasta ja luovutusajankohdasta.** Kansaneläkelaitoksen tulee kerätä vastaavat tiedot 14 a §:ssä tarkoitettuun potilaan tiedonhallintapalveluun tallennettujen ja sen kautta näytettyjen tietojen luovuttamisesta. (28.3.2014/250)
  - Terveydenhuollon palvelujen antajien potilasasiakirjatietojen luovuttamista koskevat lokitiedot tallennetaan 14 §:ssä tarkoitettuun arkistointipalveluun.
  - Asiakastietojen käyttäjien käyttöoikeustiedot ja lokitiedot tulee hävittää, kun ne eivät enää ole tarpeen asiakastietojen käytön ja luovutuksen lainmukaisuuden seuraamiseksi. Käyttöoikeus- ja lokitiedoista sekä tietojen vähimmäissäilytysajasta voidaan säätää tarkemmin sosiaali- ja terveysministeriön asetuksella.
- 7 §
  - Suunnittelu-, tutkimus- ja tilastotiedot
  - Asiakastietojärjestelmästä tulee voida tuottaa sosiaalihuollon ja terveydenhuollon palvelujen antajan oman suunnittelun, johtamisen ja tilastoinnin, sekä valtakunnallisen tutkimus- ja tilastotoiminnan kannalta tarpeelliset tiedot ja hoidon tarpeen arviointia sekä hoitoon pääsyn ajankohtaa koskevat tiedot.

Laki sosiaali- ja terveystietojen toissijaisesta käytöstä

- 41 § Tietojohtaminen
  - Sosiaali- tai terveydenhuollon palvelunantajalla on oikeus **salassapitovelvoitteiden estämättä** sekä tietosuoja-asetuksen 9 artiklan 2 kohdan h alakohdan **nojalla käsitellä ja yhdistellä tunnisteellisesti asiakastietoja**, jotka ovat syntyneet sen omassa toiminnassa tai ovat sen omiin rekistereihin tallennettuja, jos se on **välttämätöntä** palvelunantajan vastuulla toteutettavan **palvelutoiminnan tuottamista, seurantaa, arviointia, suunnittelua, kehittämistä, johtamista ja valvontaa varten**.
  - Jos palvelunantajalle on tarpeen vastuullaan toteutettavan palvelutoiminnan taikka palveluketjujen arviointia, suunnittelua tai kehittämistä varten verrata omaa toimintaansa muiden palvelunantajien toimintaan, Tietolupaviranomainen voi tuottaa tarvittavan vertailuaineiston aggregoituna tilastotietona 45 §:n mukaisesti 3 §:n 9 kohdassa tarkoitetun tietopyynnön perusteella.
  - Sen lisäksi, mitä 1 ja 2 momentissa säädetään, **kunnalla tai kuntayhtymällä on tietojohtamisen tarkoituksessa oikeus käsitellä ja yhdistellä tunnisteellisesti** myös asiakastietoja, jotka on tallennettu terveydenhuoltolain (1326/2010) 9 §:n 1 momentissa tarkoitettuun yhteisrekisteriin.

Pseudonymisointi ja anonymisointi tiedolla johtamisen arkkitehtuurissa



Anonymisointi käytännössä

- Summatason faktat
- Henkilötason faktat

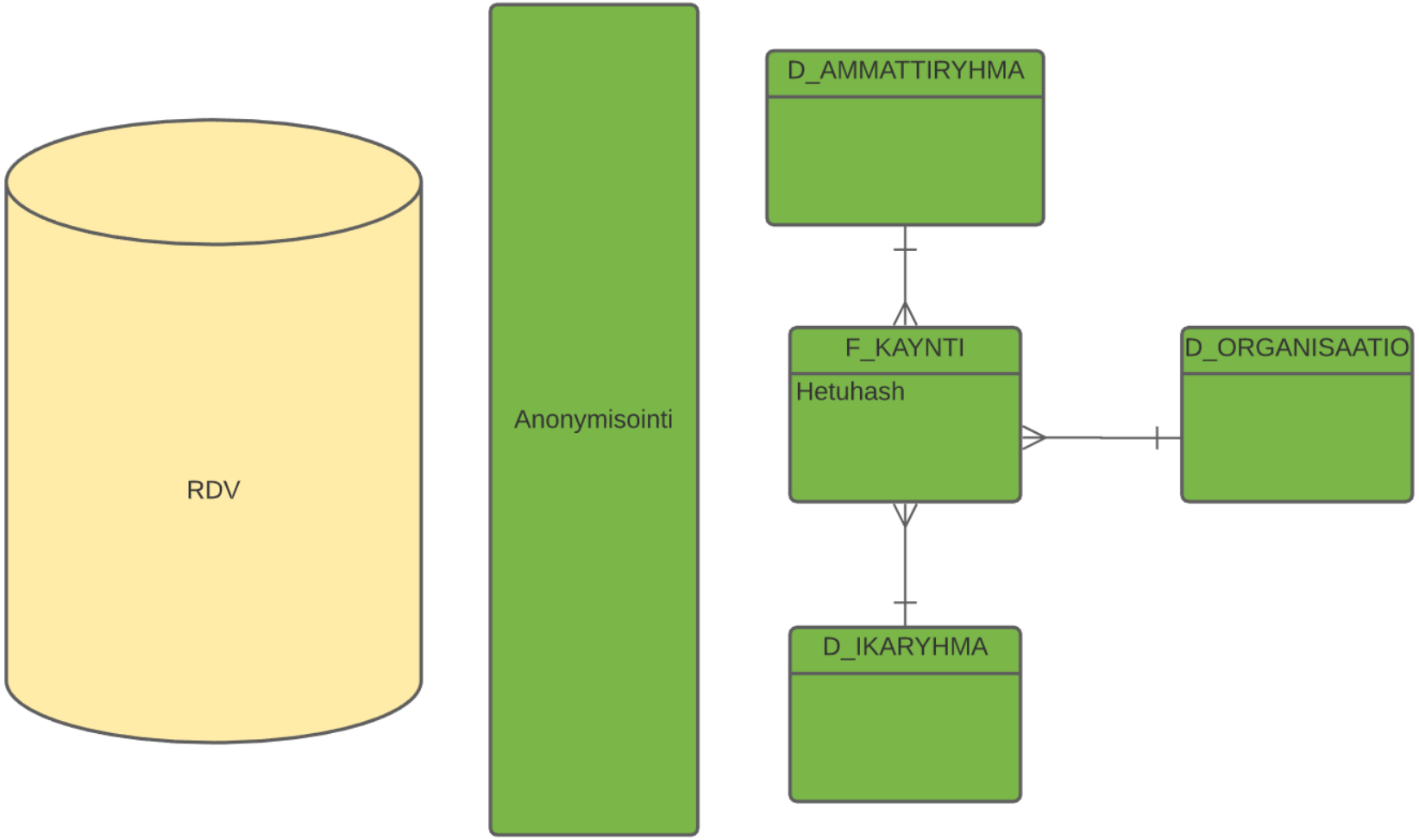
Summatason faktat

- Summatason faktoissa tiedot luokitellaan ylemmälle tasolle eikä niissä ole lainkaan henkilötunnisteita
- Näissä tulee oleelliseksi se, kuinka eri datasta voidaan löytää sellaisia yhdistelmiä, joiden kautta henkilöt voidaan epäsuorasti tunnistaa
- Liian monta luokittelua tai liian pikkutarkaksi menevät ryhmät aiheuttavat tyypillisesti tilanteet, että saadaan esille pieniä joukkoja

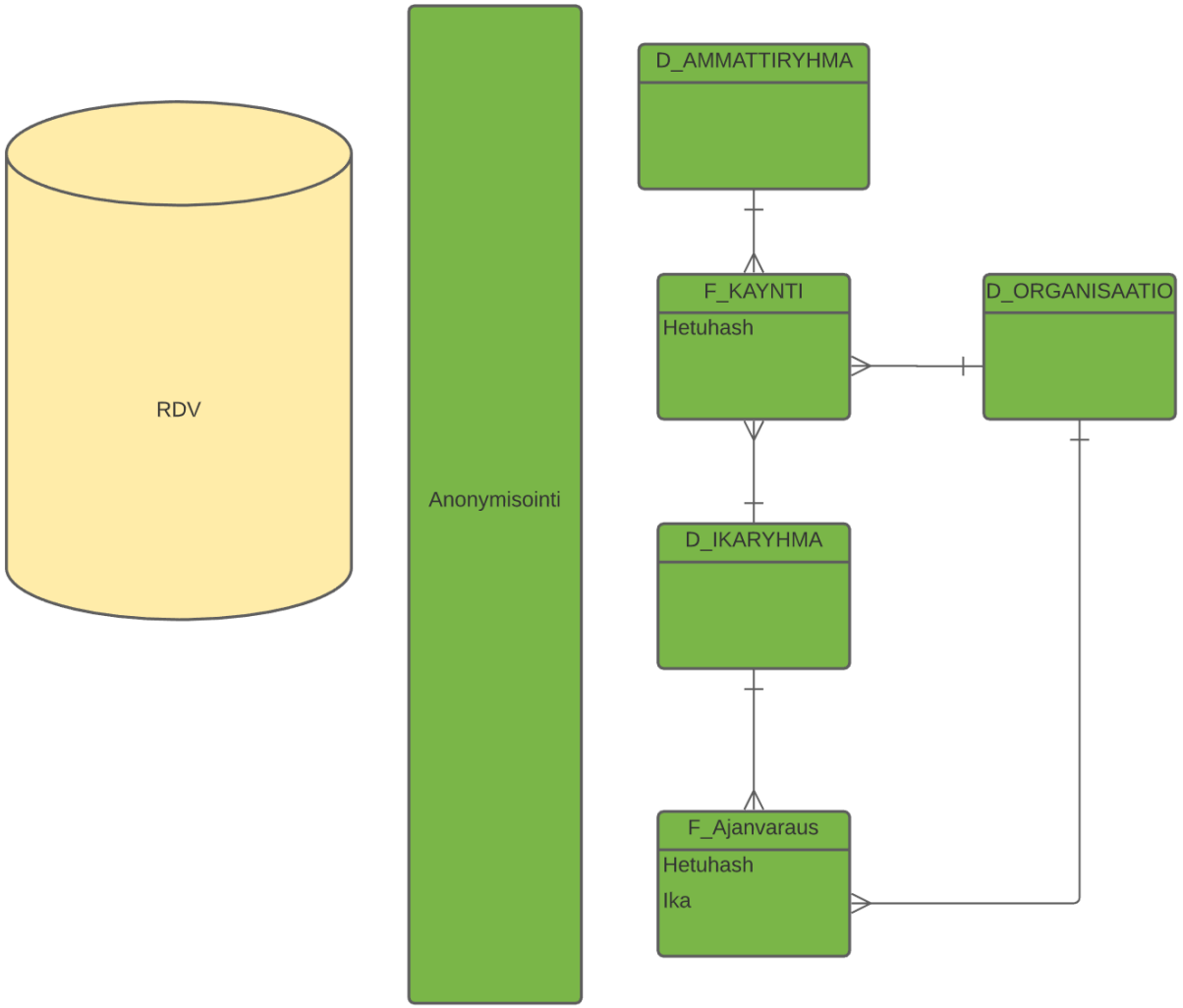
Henkilötason faktat

- Raportoinnissa on usein tarpeellista pystyä laskemaan yksittäisiä henkilöitä.
  - Tätä varten tarvitaan yksilöllinen tekninen tunniste henkilöstä, muuten raportointiohjelmistot eivät tuota laskutoimitusta pysty tekemään (distinct count-tyyppinen kaava).
- Tätä varten on mahdollista rakentaa henkilötason fakta, mutta sen anonymisointi on rakennettava huolellisesti

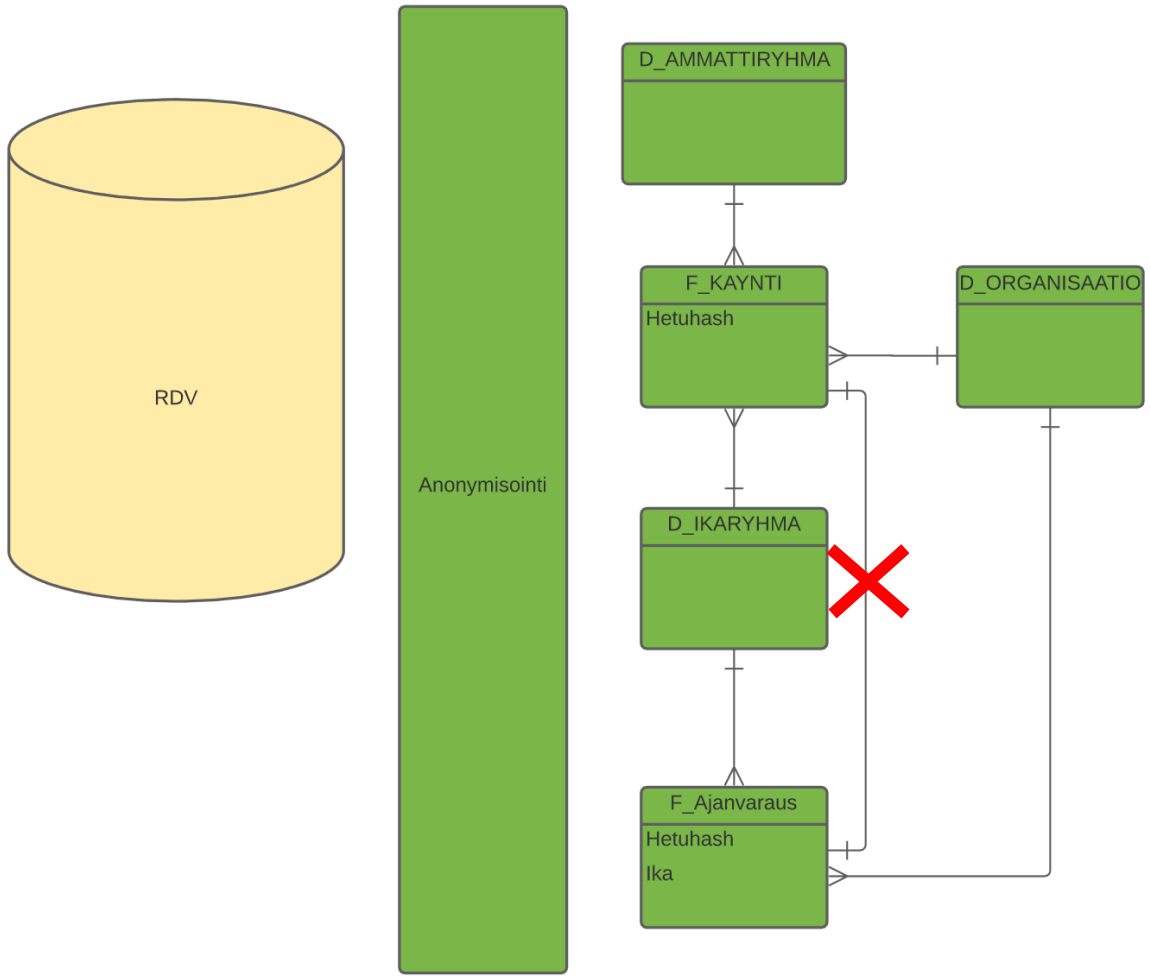
Anonymisoitu Infomart



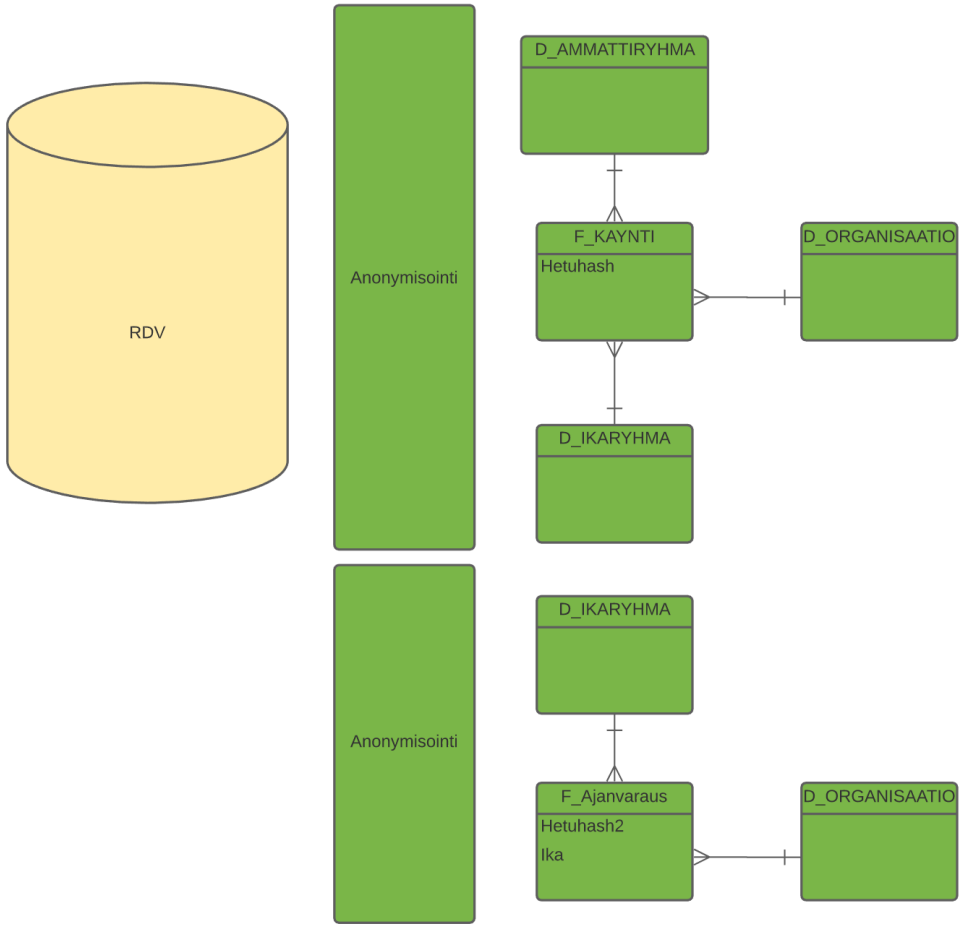
Lisää tietoa Infomartiin



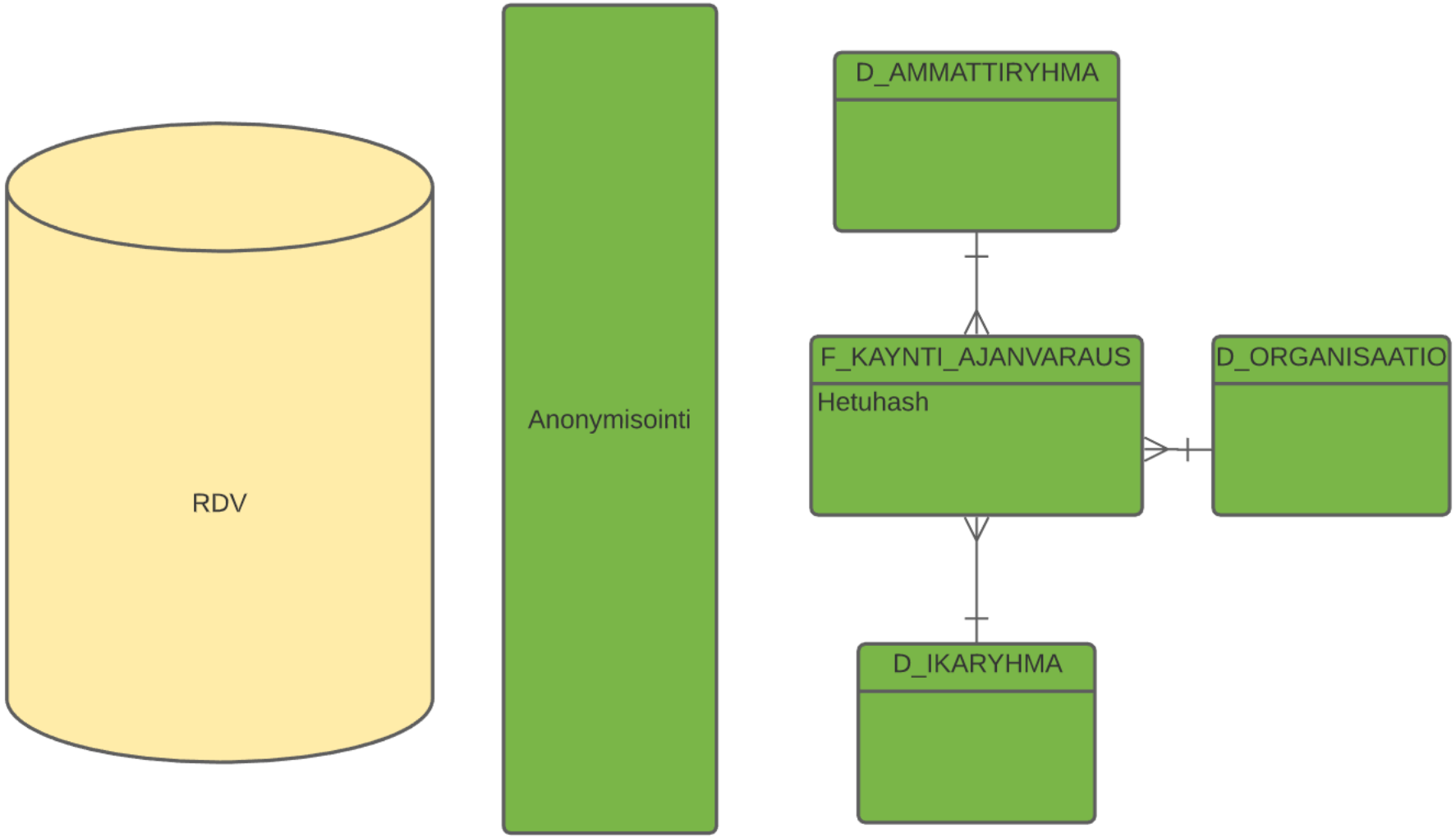
Jos käytössä sama HASH, niin faktat voidaan yhdistää



Ratkaisu: Tapauskohtainen anonymisointi ja eri hashit (myös eri dimensiot)



Yhdistetty fakta



Henkilötason faktat

- Raportoinnissa on usein tarpeellista pystyä laskemaan yksittäisiä henkilöitä.
  - Tätä varten tarvitaan yksilöllinen tekninen tunniste henkilöstä, muuten raportointiohjelmistot eivät tuota laskutoimitusta pysty tekemään (distinct count-tyyppinen kaava).
- Henkilön teknisen tunnisteen sisältävä fakta **ei ole lähtökohtaisesti anonymisoitu fakta**
  - Voidaan kuitenkin keskustella siitä, voidaanko se **jäännösriskin arvioinnin kautta sellaiseksi tulkita**
  - Jos tekninen tunniste muodostetaan faktakohtaisen suolatun hash-tunnisteen kautta, niin kyseessä on **palauttamaton tunniste**, joka **ei ole yhdistettävissä mihinkään muuhun tietoon** jota muissa julkaistuissa faktoissa, näkymissä, tietovarastossa, tietoaltaalla tai julkisissa lähteissä on
  - Näin ollen **faktaan ei pysty lisäämään mitään muuta tunnistamista helpottavaa tietoa** ja jäännösriskin arviointi voidaan tehdä pelkästään kyseisen faktan kautta
  - Mikäli jäännösriskin arvioinnissa voidaan todeta faktataulu anonyymiksi, niin silloin raportoinnissa voidaan ko. faktaa käyttää, huomioiden kuitenkin, että **jäännösriskin arviointi pitää toistaa** säännöllisesti ja tilanne arvioita uudelleen uuden datan valossa.
- Faktakohtainen suolaus kannattaa mahdollisuuksien mukaan toteuttaa kyselykohtaisesti vaihtuvalla satunnaisella arvolla
  - Samalla tämä tarkoittaa sitä, että faktaan on otettava mukaan kaikki mahdollisen henkilödimension takaa tarvittavat muut tiedot, koska henkilön uudelleen hashattua henkilön tunnistetta ei voi enää käyttää vierasavaimena.

Jäännösriskin arviointi

- **Tehdään tapauskohtaisesti** jokaiselle julkaistavalle faktataululle.
- On arvioitava mikä riski on jos faktataulun tietoja on mahdollista yhdistää muihin faktoihin ja dimensioihin tai kokonaan ulkopuoliseen dataan.
- Jäännösriskin arvioinnissa perusperiaatteena on **yrittää löytää** datasta sellaisia **yhdistelmiä**, joiden kautta **henkilöt voidaan epäsuorasti tunnistaa**
  - Tällaisia tyypillisiä yhdistelmiä ovat esimerkiksi **harvinaisemmat ammatit, sairaudet, asuinalueet**
  - Näiden kohdalla **ei välttämättä aina riitä edes se, että noudetaan tilastointieettisiä periaatteita** raporteilla, kuten tiedot vain sellaisista ryhmistä, joissa on yli 5 henkilöä
  - Interaktiivisessa raportoinnissa tyypillinen keino kiertää raportoinnissa tehtävät rajoitukset on tehdä valintoja, jotka menevät juuri noiden määrärajoitusten yli.
- Jäännösriskin arvioinnissa on oleellista se, että **arvioidaan tehtyjen luokitteluiden ja aggregointien riittävyyttä kyseisessä käyttötapauksessa**
  - Silti täytyy muistaa, että toimintaympäristöä riittävästi tunteva ammattilainen saattaa kaikesta huolimatta pystyä tunnistamaan aineistosta henkilöt. Tällöin täytyy muistaa se, että hän kykenisi siihen muutoinkin, koska hänellä nimenomaan on hyvä ja kattava tuntemus toimintaympäristöstä ja asiakkaista

Julkaistavan tiedon suunnittelu vs. anonymisointi

- Prosessi kannattaa sopia
- Lähtökohtaisesti voimme suunnitella faktat siten, että **otetaan mukaan kaikki tiedot**, jota raportoinnissa halutaan käytettävän
  - Toki suunnittelussa on jo hyvä huomioida se tosiasia, että jokainen henkilöä luokitteleva tekijä lisää mahdollisuutta tunnistaa henkilö aineistosta
- Toteutetaan fakta halutuilla tiedoilla
- **Arvioidaan toteutuksen jälkeen** erilaisilla kyselyillä miten helposti aineistosta pystyy henkilön tunnistamaan
- Tehdään jäännösriskin arvioinnista dokumentti, joka esitetään organisaation tietosuojasta vastuussa olevalle ja **pyydetään lausunto**, jossa todetaan, että anonymisointi on toteutettu **riittävällä tavalla ja merkittävää riskiä henkilön tunnistamiseen ei ole**
- **Toistetaan jäännösriskin arviointi** säännöllisin väliajoin (kun dataa tulee lisää) ja myös silloin kun faktan tai dimensioiden rakenne muuttuu