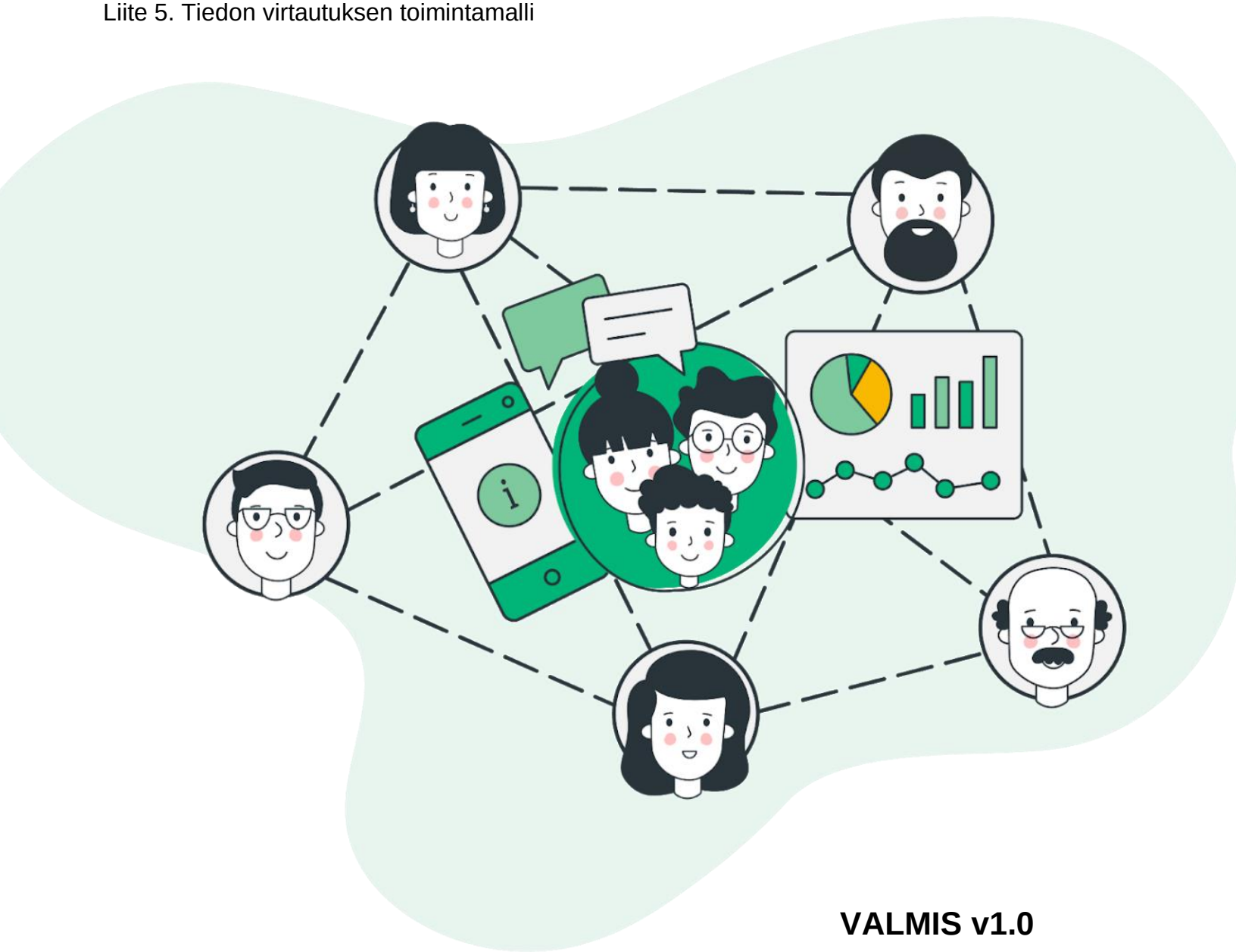
16.6.2022

Kohti ihmislähtöistä ja dynaamista palveluverkkosuunnittelua

Liite 5. Tiedon virtautuksen toimintamalli



VALMIS v1.0

Helsinki KUOPIO TURKU ÅBO GOFORE ubigu platform of trust vastuu group

Sisällys

Dokumentin versiohistoria

Versio	*Päiväys*	*Laatija*	*Muutoksen kuvaus*
0.1	31.5.2022	MK, JM, JU, OH, SK, JJ, MS, KK, PP, VJ	Ensimmäinen versio
1.0	3.6.2022	MK, JM, JU, OH, SK, JJ, MS, KK, PP, VJ	Valmis hankkeen ohjausryhmän hyväksyttäväksi
1.0	16.6.2022		Hyväksytty hankkeen ohjausryhmässä

Tiivistelmä

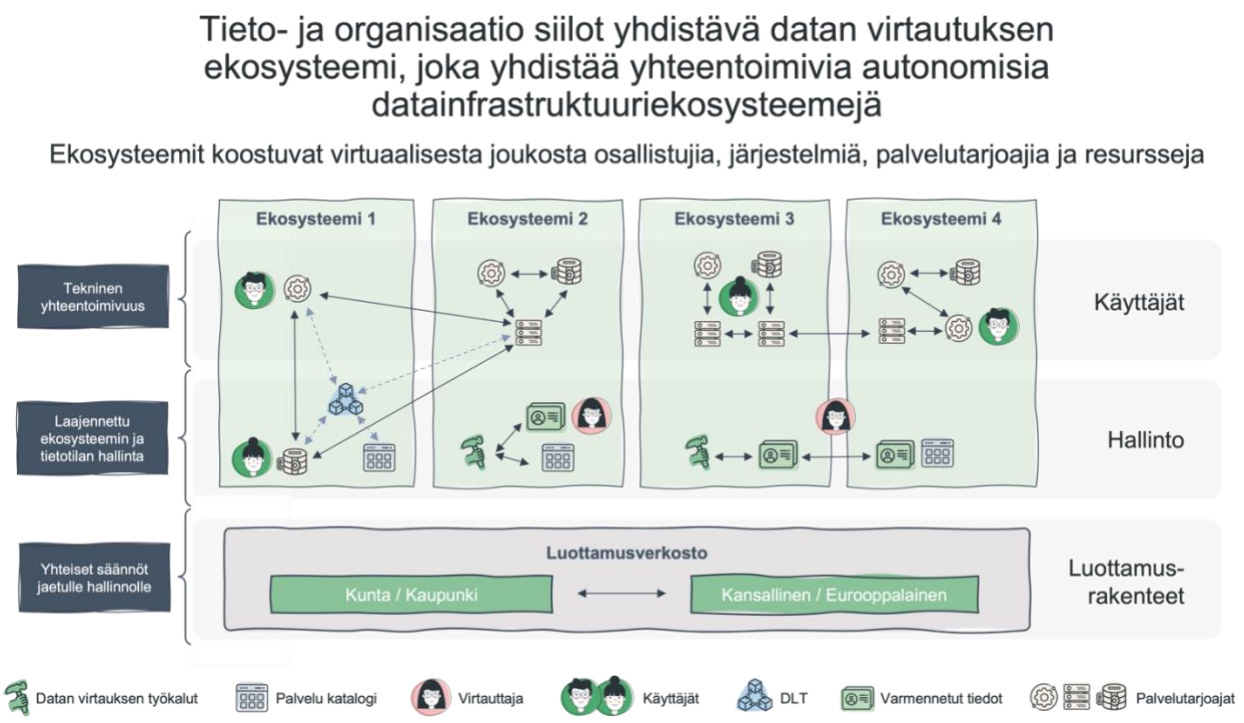
1. Johdanto

Tiedon virtautus laajassa ja kokonaisvaltaisessa kontekstissa organisaatio- ja järjestelmärajat ylittävänä ilmiönä on varsin tuore. Data-suvereniteetti ja siihen liittyvä eurooppalainen suuntaus, jota tuetaan lainsäädännöllä, luo tasapuoliset toimintaedellytykset paikallisille toimijoille.

Datatalouden kehitys leikkaa läpi koko yhteiskunnan ja vaatii ohjauksen tueksi rakenteita, kuten data-avaruuksia sekä arvoketjujen kokonaisvaltaista tarkastelua, niin taloudellisissa, sosiaalisissa, poliittisissa kuin luonnon ekosysteemeissä. Yksinkertaistettuna: murroksen aikana eläminen koskee meitä kaikkia.

Ehkäpä se tärkein johdatteleva kysymys tämän dokumentin lukijalle on: missä roolissa me ympäristöämme tarkkailemme - ymmärrämmekö ja osaammeko hallita teknologiaa ja dataa vai annammeko niiden ohjailla meitä? Tiedon virtautuksen ekosysteemien päätarkoitus on mahdollistaa uusiutuminen, hallinta ja omistajuus alati muuttuvassa toimintaympäristössä. Toisaalta organisaatioihin ja kuntiin on muodostunut siilomaisia tiedon virtautuksen ekosysteemejä, joiden yhteentoimivuuden varmistaminen ja kokonaiskuvan muodostaminen on haastavaa.

Luomalla yhdistävä kerros eri toimijoiden ja ratkaisujen välillä mahdollistetaan siirtymä kohti kestävän kehityksen tarvitsemia käytännönläheisiä ratkaisuja, joiden elinehtona voidaan pitää kokonaisvaltaista tiedon virtautusta ja tietopohjaa.



Kuva 1 - Tiedon virtautuksen eurooppalainen ulottuvuus, Gaia-X

1.1. Dokumentin sisältö ja kohderyhmä

Tämä dokumentti kuvaa tiedon virtautuksen toimintamallin ja tiedon virtautuksen mahdollistavan teknisen kohdearkkitehtuurin. Dokumentti keskittyy tarjoamaan kunnille ja muille toimijoille keinot osallistua tiedon turvallisen ja hallitun jakamisen ekosysteemeihin sekä luomaan niitä omiin organisaatioihinsa. Dokumentissa esitellään myös prosessi eri toimijoiden tietotarpeita palvelevien tietotuotteiden luomiseen ja tässä prosessissa huomioon otettavia asioita.

Tiedon virtautuksen toimintamalli tukee erityisesti seuraavien ryhmien tarpeita:

- Julkiset toimihenkilöt ja virkamiehet, jotka päättävät mitä tietoja jaetaan ja kenelle.
- Tiedon virtauttamisesta vastaavat kunnan toiminnan ja ICT:n asiantuntijat.
- Kuntien tietojen omistajat ja tiedonhallinnasta vastaavat tahot.
- Kuntien kokonaisarkkitehtuurista vastaavat tahot.
- Kuntien ICT-kehityskumppanit ja –palvelutoimittajat.
- Kuntayhteistyö.

Dokumentti tarjoaa yleiskuvauksen tiedon virtautuksen periaatteista myös kuntien strategisen, taktisen ja operatiivisen tason päätöksentekijöille sekä muille kuntien kanssa tietoja vaihtaville julkisille toimijoille, joiden tietotarpeita tiedon virtautus palvelee. Tässä dokumentissa esitettyjä periaatteita voivat hyödyntää myös kuntien kanssa yhteistyötä tekevät yritykset.

1.2. Rajaukset ja reunaehdot

Tässä luvussa esitellään lyhyesti keskeiset tätä dokumenttia koskevat rajaukset ja reunaehdot, jotka määrittävät dokumentin sisällön ja riippuvuudet.

Tässä dokumentissa esiteltävälle tiedon virtautuksen toimintamallia koskevalle ohjeistukselle on asetettu seuraavat rajaukset:

- Dokumentissa kuvataan tiedon virtautuksen ja harmonisoinnin ekosysteemin toimintamalli sekä tekninen arkkitehtuuri yleisellä tasolla, mutta tarjoten konkreettisia keinoja ja elementtejä niiden rakentamiseen.
- Dokumentti on luonteeltaan enemmän monipuolisesti aihetta tarkasteleva kuin teknisiin yksityiskohtiin pureutuva.
- Kuvataan prosessivaiheet käyttötapauksen sekä siihen liittyvän datatuotteen rakentamiseen ja hallintaan.
- Dokumentissa ei oteta kantaa, mitä tietoja virtautetaan, vaan mitä pitää ottaa huomioon tietoja virtautettaessa.
- Dokumentissa ei oteta kantaa, miten tietojärjestelmäpalvelut järjestellään loogisiin järjestelmiin, vaan kuvataan palvelujen hyödyntäminen ja palvelujen väliset suhteet.
- Dokumentti ei kannusta käyttämään tiettyjä yksittäisiä teknologiatuotteita, mutta suosittelee tiettyjen teknologioiden hyödyntämistä, jotka mahdollistavat tiedon virtautuksen tehokkuuden kasvattamisen.
- Dokumentissa käytetään esimerkkinä Platform of Trust -palvelualustaa, mutta markkinoilla on tarjolla useita vastaavia integraatio- ja palvelualustoja.

Lisäksi dokumentissa huomioidaan seuraavat reunaehdot:

- Dokumentissa esitellyt toimintamallit ja tekninen arkkitehtuuri soveltuvat tiedonhallinnaltaan ja tietojen virtautuksen osalta eri kypsyystasoilla oleville kunnille.
- Dokumentissa esitellyt toimintamallit ja tekninen arkkitehtuuri voidaan ottaa käyttöön kokonaan tai osissa.
- Toimintamallin tekniset osiot ja kuvaukset ovat kansallisesti yhteentoimivia.
- Dokumentissa esitellyt tiedon hyödyntämisen ja käsittelyn kuvaukset eivät ole ristiriidassa lainsäädännön ja julkishallinnon ohjeistusten kanssa.

1.3. Ohjaavat lait ja säädökset

Kohdearkkitehtuurin laatimishetkellä on tunnistettu tämän dokumentin esittämiä linjauksia ohjaaviksi laeiksi ja säädöksiksi seuraavat:

- Laki julkisen hallinnon tiedonhallinnasta (906/2019)
- Laki digitaalisten palvelujen tarjoamisesta (306/2019)
- Laki hallinnon yhteisistä sähköisen asioinnin tukipalveluista (571/2016)
- EU:n tietosuoja-asetus (2016/679)
- Tietosuojalaki (1050/2018)

Kohdearkkitehtuurin laatimisen yhteydessä on varmistettu, että sen esittämät linjaukset ja sovittu hallintamalli noudattavat edellä mainittuja lakeja ja säädöksiä. Mikäli jotakin tunnistetuista laeista tai säädöksistä on päivitetty, on kohdearkkitehtuurin hyödyntäjän vastuulla tarkistaa ajantasaiset säädökset ja varmistaa, että niitä noudatetaan.

1.4. Kuntien tiedon virtautuksen haasteita

Tässä luvussa listataan lyhyesti DigiPAVe 2.0-hankkeen aikana tunnistettuja nykytilan haasteita hankekaupungeissa tiedon siiloutumisesta, arkkitehtuurin hajanaisuudesta ja niistä koituvista ongelmista tiedon virtauttamiselle ja saatavuudelle.

1.4.1. Yleiset havainnot toiminta-arkkitehtuurien nykytilasta

- Palvelualueiden toiminnan ja tietokokonaisuuksien kokonaiskuvaa ei monesti ole, toiminta on useimmiten siiloutunutta.
- Palvelualueiden eri tietokokonaisuuksien yhteensovittaminen (jatkuva yhteentoimivuustyö toimialarajat ylittävästi) on kuntaorganisaatioissa vielä alkuvaiheessa.
- Palvelualueiden tiedonhallinnan ja ulkopuolisten sidosryhmien roolien oikeudet ja vastuut ovat epäselviä.
- Holistinen dataekosysteemiajattelu on uutta kunnissa.
- Dataekosysteemikyvykkyydet ovat vielä puutteellisia kuntaorganisaatioissa (käytön ja luottamusten hallinta, datatuotteen määrittely ja rakentaminen, datatuotteen julkaisu, rajapintojen ja datakatalogien hallinta jne.).

1.4.2. Yleiset havainnot tietoarkkitehtuurien nykytilasta

- Ymmärrys eri käsitteistä ja niiden välisistä yhteyksistä poikkeaa kaupunkien sisällä ja välillä.
- Yhteentoimivia hallinnonalojen tai niitä ylittäviä ylätason ontologioita, tietokomponenttikirjastoja ja soveltamisprofiileja ei juurikaan ole ollut, jotka ovat edellytyksenä tiedon virtaukselle ja yhteentoimivuudelle kaupunkien alue- ja palvelukehityksessä.
- Samoja asioita käsittelevät käsite- ja tietomallit poikkeavat toisistaan.

- Palvelualueiden tietoja ylläpidetään erillään, eivätkä ne ole kovinkaan hyvin löydettävissä, saati hyödynnettävissä prosessien välillä.
- Useita prosessien tarvitsemia tietoja ei yhdistellä automaattisesti, vaan tiedot on pidetty erillään erillisissä tietovarannoissa, joita ei ole integroitu.
- Tietojen yhdistäminen palvelualueiden välillä on haastavaa, koska ei ole yhdistävää yksilöivää tunnistetta tietojen välillä (osin VTJ-PRT- ja kustannuspaikkatunnukset).
- Tietokohteita on yksilöity osoitteiden perusteella tarkan sijaintiedon sijaan.
- Osasta keskeisiä tietoryhmiä puuttuu yksilöivät tunnisteet, mikä hankaloittaa tietojen yhdistämistä.

1.4.3. Yleiset havainnot tietojärjestelmäarkkitehtuurien nykytilasta

- Järjestelmäkohtaisia standardeja rajapintoja ei juurikaan ole (pääsääntöisesti löytyy paikkatietojärjestelmistä/perusrekistereistä).
- Osasta järjestelmäratkaisuista puuttuu myös rajapintapalvelut kokonaan, tai niitä ei voida luotettavasti hyödyntää.
- Järjestelmäratkaisuja on rakennettu pääosin palvelualuekohtaisesti ilman palvelualueiden ylittäviä yhteentoimivia tietomalleja, minkä vuoksi niiden välisten integraatioiden toteuttaminen on hankalaa.
- Suurta osaa kuntien prosesseissa tarvittavista tiedoista käsitellään taulukkolaskentaohjelmissa kuten Excelissä, joiden taulukoita ei ole integroitu muihin fyysisiin tietovarantoihin (esimerkiksi tietokannat ja tietovarastot) tietojen yhdistelemistä ja koostamista varten.
- Tietojen hakeminen ei ole aina kunnan itse toteutettavissa, vaan haku joudutaan tilaamaan järjestelmätoimittajalta.
- Tietojärjestelmäarkkitehtuureja ei ole suunniteltu kokonaisarkkitehtuurin keinoin.
- Edellä mainituista syistä johtuen nykyisillä järjestelmäratkaisuilla ei kyetä luomaan kovinkaan helposti tiedon virtausta, joka mahdollistaisi muun muassa tiedolla johtamisessa tarvittavan reaaliaikaisen ja kattavan tilannekuvan, joka tehostaisi kuntien päätöksentekoa.
- Useilla kunnilla on kuitenkin käytössään erilaisia suunnitteluanalytiikan järjestelmäratkaisuja, joilla voitaneen tarpeen mukaan yhdistellä ja esittää ainakin joitakin prosessien tarvitsemia tietoja.

1.5. Mikä ratkaisuksi?

Kuntien tietopääoma tulee jatkossa valjastaa käyttöön entistä tehokkaammin, jotta kunnat kykenevät tuottamaan siitä arvoa omien sekä julkisen sektorin yhteisten palvelutarpeiden ennakoimiseksi ja palvelujen kehittämiseksi. Kuntien tuottamaa tietoa tarvitsevat niin kunnan sisäiset, kansalliset kuin kansainvälisetkin organisaatiot sekä muut kunnat. Vastavuoroisesti kuntatasolla tarvitaan päätöksenteon tueksi tietoa edellä mainituilta organisaatiotasoilta. Tiedon jakaminen esimerkiksi kuntien välillä voi mahdollistaa ratkaisuja yhteisiin ongelmiin, jotka eivät ole maantieteellisesti sidottuja.

Jotta tietoa saadaan sitä tarvitseville tahoille turvallisesti, tulee kuntien kyetä suunnittelemaan ja toteuttamaan tietovirtoja sekä mahdollistamaan erilaisista tietoa jakavista ja hyödyntävistä toimijoista koostuvia verkostoja kunnan sisällä ja ulkopuolella. Tiedon verkostomaisessa jakamisessa useiden toimijoiden välillä korostuu tiedon omistajuuden ja tiedon jakajien luotettavuuden varmistaminen sekä halutun tiedon löytäminen. Tässä dokumentissa tarjotaan näihin haasteisiin ratkaisuksi toimintamallia tiedon tuotteistamiseksi, mikä mahdollistaa tiedon luotettavan jakamisen ja hyödyntämisen erilaisista toimijoista koostuvissa dataekosysteemeissä.

Dokumentissa esiteltyjen toimintamalli- ja arkkitehtuurikuvausten tarkoitus on luoda tiedon turvallinen ja luotettava virtauttaminen kunnille helpommin lähestyttäväksi. Myöhempänä esitetyt ratkaisut mahdollistavat kunnan ottaa vastuuta ja omistajuus tiedostaan sekä sen kehittämisestä. Ne tarjoavat myös kiintopisteet kunnan tiedon virtautuksen osaamisen, toimintamallien ja arkkitehtuurin kehittämiseen. Lisäksi dokumentti tarjoaa kunnalle työkalut arvioida, missä vaiheessa kunta on tietojensa virtauttamisen hallinnassa sekä ymmärrystä siitä, mitä sen tulisi kehittää tiedon virtautuksen mahdollistamiseksi sekä verkostomaisiin dataekosysteemeihin liittymiseksi. Kunta kykenee dokumentin perusteella myös arvioimaan, minkälaista apua sen tulisi hankkia edellä mainittujen suunnitteluun ja toteuttamiseen.

1.6. Terminologia

Dataekosysteemien ja tiedon virtauttamisen terminologian standardointi on vielä kesken, erityisesti suomenkielisten käännösten osalta. Vakiintuneen ilmaisun puuttuessa kieli elää ja uudet paremmin kuvaavat ilmaisut yleistyvät. Tämä dokumentti nojaa kansallisella tasolla Sitran ja ministeriöiden sekä eurooppalaisen datatalouden kehityksen ja sen edelläkävijäorganisaatioiden julkaisuihin. Näitä organisaatioita ovat muun muassa Gaia-X AISBL, International Data Spaces Association (IDSA) ja Open and Agile Smart Cities (OASC).

Käsite	Määritelmä	Lähde
Dataekosysteemi	Useista dataverkostoista koostuva verkosto, jossa toimijat tekevät yhteistyötä tavoitteenaan jakaa ja käyttää dataa verkoston sisällä sekä edistää innovointia ja uutta liiketoimintaa.	Sitran tulevaisuussanasto
Datakatalogi	Datakatalogi on strukturoitu metadatarekisteri, johon on yhdistetty metadataa useampien julkisten organisaatioiden hallussa olevista aineistoista. Datakatalogit voivat olla: 1) kansallisia (esimerkiksi avoindata.fi ja data.gov.uk), 2) seudullisia (Washington D.C. tai Helsinki Region Infoshare), 3) kaupunkien ylläpitämiä (San Francisco ja Tampere), 4) yksityisten tahojen ylläpitämiä (Sunlight Foundation - National Data Catalog).	Tiedon jakamisen toimintamalli
Datatalouden ekosysteemi	Datatalouden ekosysteemit muodostavat verkoston, joka koostuu dataa liiketoiminnan lähteenä käyttävistä ekosysteemin jäsenistä. Eri sidosryhmät ovat verkoston ja sen arvoketjujen kautta toisiinsa suorassa tai epäsuorassa yhteydessä. Datatalouden ekosysteemiin kuuluvat myös säännöt (viralliset tai epäviralliset), jotka määrittävät verkostossa sallitun toiminnan.	Sitran tulevaisuussanasto
Ekosysteemi	Ekosysteemit ovat yhteisen vision ympärille rakentuvia, jatkuvasti kehittyviä verkostorakenteita. Ekosysteemiin ilmaantuu uusia ominaisuuksia toimijoiden välisen vuorovaikutuksen ja riippuvuuksien kautta.	VTT:n Yhdessä kestävää kasvua - ekosysteemiopas
Federointi	Federaatiolla tarkoitetaan löyhästi kytkettyjä toistensa kanssa vuorovaikuttavaa toimijoiden joukkoa, jotka tuottavat, hyödyntävät ja tarjoavat resurssea joko suorasti tai epäsuorasti.	Gaia-X-arkkitehtuuri-dokumentti ver. 21.09
Federoitu katalogi	Federoiduilla katalogilla tarkoitetaan federoitua palvelua, joka koostaa ekosysteemin resurssien metadataa yhteisesti nähtävillä olevaan rekisteriin, jonka kautta	Gaia-X-arkkitehtuuri-dokumentti ver. 21.09:n pohjalta tehty määritelmä

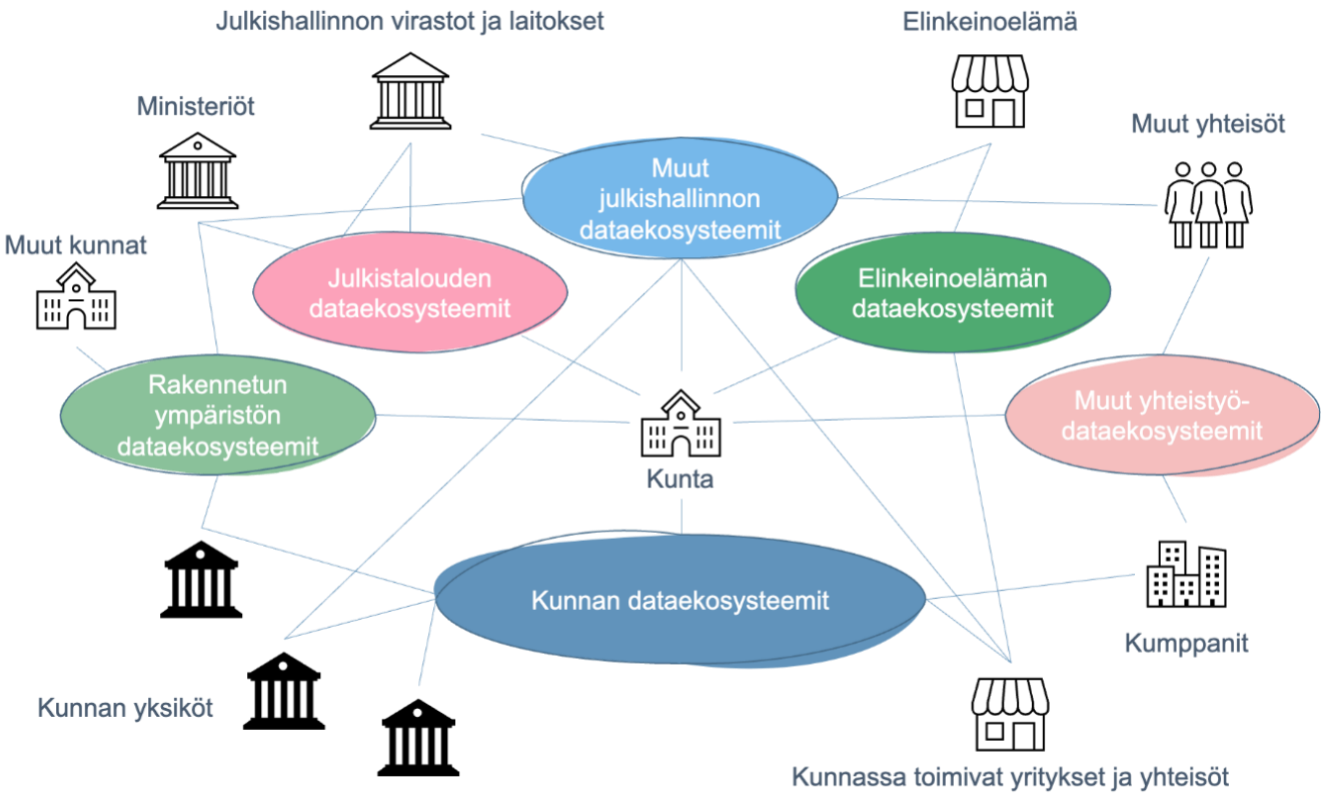
	resurssien löytäminen ja valinta mahdollistuu. Vrt. datakatalogi.	
Itsekuvaus	Itsekuvaukset tarkoittavat resurssia tai toimijaa kuvailevia metatietoja, joiden tarkoitus on antaa taustatietoja kuvauksen kohteesta.	Gaia-X-arkkitehtuuri-dokumentti ver. 21.09:n pohjalta tehty määritelmä
Käsitemalli	Tietomalli, joka kuvaa tarkastelun kohteena olevat kohdemaailman käsitteet ja niiden väliset suhteet	Yhteentoimivuusalustan sanasto
Koneluettava muoto	Tarkoittaa tiedostomuotoa, jonka rakenne mahdollistaa sen, että ohjelmistot pystyvät helposti yksilöimään, tunnistamaan ja poimimaan siitä tietoaineistoja, yksittäisiä tietoja sekä niiden rakenteita.	Tiedonhallintalaki 2§
Looginen tietomalli	Tietomalli, joka kuvaa tiedot ja tietorakenteet loogisella tasolla. Looginen tietomalli on käsitemallista tarkemmalle tasolle viety kuvaus. Looginen tietomalli kuvaa tarkastelun kohteena olevat luokat, niiden väliset assosiaatiot (suhteet) ja luokkiin sisältyvät attribuutit.	Yhteentoimivuusalustan sanasto
Looginen tietovaranto	Toiminnan tarpeiden perusteella koottu joukko tietoja. Looginen tietovaranto voi olla yhden tai useamman organisaation tuottama tai hallinnoima.	Julkisen hallinnon yhteinen sanasto
Noodi	Noodi on laskennallinen tai fyysinen entiteetti, joka hostaa, manipuloi tai vuorovaikuttaa muiden laskennallisten tai fyysisten entiteettien kanssa. Nodeja voi olla muun muassa datakeskukset, palvelimet, tietoliikenteen verkko- ja infrastruktuuripalvelut, virtuaalikoneet ja sovelluskontit.	Gaia-X-arkkitehtuuri-dokumentti ver. 21.09
Sidosarkkitehtuuri	Tarkoittaa muualla määriteltävää arkkitehtuuria, jolla on vaikutus organisaation tai muun tarkasteltavan kohteen arkkitehtuurityöhön.	JHS 179 Termit ja määritelmät
Soveltamisprofiili	Soveltamisprofiilit tarkentavat yhteisiä tietomalleja profiilien kuvaamiin käyttötarkoitukseen sopiviksi. Profiileissa määritellään muun muassa tarkennettavien koodiarvoisten ominaisuuksien arvojoukkoina toimivat koodistot ja mahdolliset rajoitukset rakenteisille ominaisuuksille.	
Tiedonhallinta	Tarkoittaa viranomaisen tehtävien hoidossa tai sen muussa toiminnassa syntyviin tarpeisiin perustuvia toimia ja tietoturvallisuustoimenpiteitä viranomaisen tietoaineistojen, niiden käsittelyvaiheiden ja tietoaineistoihin sisältyvien tietojen hallinnoimiseksi riippumatta tietoaineistojen tallentamistavasta ja muista käsittelytavoista.	Tiedonhallintalaki 2§
Tiedon virtautus	Tarkoittaa tämän dokumentin asiayhteydessä tiedon jakamisen vaiheistamista ja vaiheiden suunnittelua.	Tämä dokumentti
Tietojärjestelmäpalvelu	Tarkoittaa kokonaisuutta, johon kuuluu käyttöliittymän sisältävä loppukäyttäjäpalvelu sekä rajapinnan sisältävä automatisoitu sovelluspalvelu.	JHS 179 Termit ja määritelmät
Tietotuote / Datatuote	Tarkoittaa palvelutuotosta, joka koostuu informaatiosta. Datatuote on rakenteistettu tietotuote, jota voidaan käsitellä sähköisesti.	Julkisen hallinnon yhteinen sanasto (tietotuote)
Tekninen rajapinta	Tarkoittaa sähköisen tietojenvaihdon mahdollistavaa tiedonsiirtoratkaisua kahden tai useamman tietojärjestelmän välillä.	Tiedonhallintalaki 2§
Tietovaranto	Tarkoittaa viranomaisen tehtävien hoidossa tai muussa toiminnassa käytettäviä tietoaineistoja sisältävää kokonaisuutta, jota käsitellään tietojärjestelmän avulla (tai manuaalisesti).	Tiedonhallintalaki 2§

Tietovirta	Tiedon liikkuminen lähteestä kohteeseen määritetyn tietoprosessin mukaisesti.	Suomalainen asiasanasto- ja ontologiapalvelu
Toimintamalli	Tarkoittaa tekijöitä, jotka sääntelevät toimintaa ja sen rooleja. Käsitteen avulla kuvataan niitä sääntöjä tai ohjeita, jotka sääntelevät toimijoiden välistä toimintaa, esimerkiksi ihmissuhteita (vanhemmuus) tai sopimuksia (työsopimus).	Julkisen hallinnon yhteinen sanasto
Viitearkkitehtuuri	Tarkoittaa kehitettävään kohteeseen sovellettavaa loogisen ratkaisumallin kuvausta. Viitearkkitehtuuri tarjoaa yhteisen mallin ja käsitteistön kehitettävän kohteen arkkitehtuurin suunnitteluun ja toteuttamiseen määrittäen kohteeseen kuuluvat rakenteet ja niiden väliset suhteet. Viitearkkitehtuuri ohjaa organisaation tai kehitettävän kohteen arkkitehtuuria.	JHS 179 Termit ja määritelmät
Yhteentoimivuus	Tarkoittaa toimintaan liittyvien toimijoiden, prosessien ja tietojärjestelmien kykyä toimia ja viestiä keskenään sellaisella tavalla tai siinä laajuudessa, että ne voivat rutiininomaisesti käyttää ja ymmärtää toistensa tietoja. Yhteentoimivuus voi olla luonteeltaan *teknistä, semanttista, organisatorista tai oikeudellista.*	Yhteentoimivuusalustan sanasto **Finto**

2. Dataekosysteemin perusteet

Dataekosysteemit tai datapohjaiset toiminnan ekosysteemit ovat toiminnallisia ekosysteemejä. Toiminnallisilla ekosysteemeillä tarkoitetaan erilaisten toimijoiden yhteenliittymiä sekä niiden resurssien yhdistämistä siten, että yhdistämisellä mahdollistetaan arvontuottoa, joka ei ole mahdollista yksittäisen toimijan omilla resursseilla. Ekosysteemit muodostavatkin arvoa tuottavia riippuvuussuhteita, joista ideaalitilanteessa kaikki ekosysteemitoimijat hyötyvät.

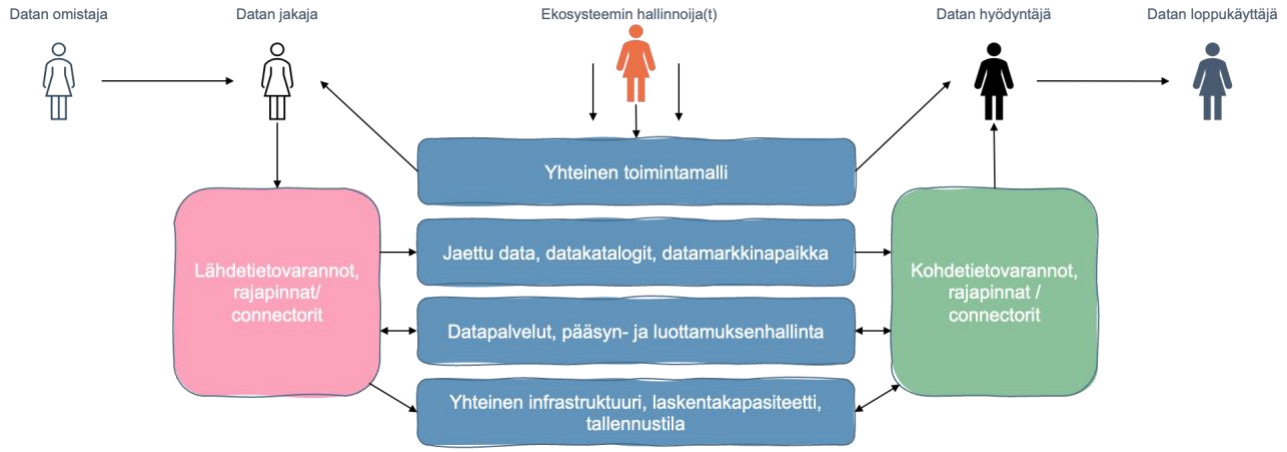
Datapohjainen toiminnan ekosysteemi tai dataekosysteemi on ekosysteemi, jossa dataa käsitellään resurssina, ja jota ekosysteemin toimijat hyödyntävät yhdessä arvon tuottamiseksi. Data on tässä tapauksessa strategista pääomaa, jolla mahdollistetaan uusia innovatiivisia ratkaisuja toiminnan ja palvelujen kehittämisessä. Avaintekijä datan menestyksekkääseen hyödyntämiseen on jakaa ja ylläpitää dataa ekosysteemin sisällä, jotta yhteisillä dataresursseilla voidaan tukea yksittäisen tai useamman ekosysteemitoimijan yhteisiä prosessiketjuja päästä päähän. Datapohjaiset toiminnan ekosysteemit tarvitsevatkin yhteiset toimintatavat ja aktiivista hallinnointia, jotta eri toimijoiden välille kyetään muodostamaan luottamussuhteita sekä varmistamaan ekosysteemiarkkitehtuurin eri toimijoiden ja komponenttien yhteentoimivuus. Yksittäinen toimija, kuten kunta tai kunnan sisäinen yksikkö, voi toimia ekosysteemin sisällä useassa eri roolissa, kuten tiedon jakajana ja hyödyntäjänä sekä osallistua useaan kunnan sisäiseen tai ulkoiseen dataekosysteemiin palvelujen kehittämiseksi.



Kuva 1 - Esimerkkejä kunnan ja muiden julkisen hallinnon toimijoiden mahdollisista ekosysteemeistä

Datan hyödyntäminen strategisena resurssina ja useiden ekosysteemitoimijoiden yhteistoiminta sekä tietotarpeet aiheuttavat erityisiä haasteita datan omistajuutta ja hyödyntämisoikeuksia koskien. Lähtökohtaisesti, mitä useampi toimija ekosysteemissä on, sitä monimutkaisempaa luottamuksen, datan omistajuuden ja käyttöoikeuksien hallinta on ilman apuvälineitä. Datan omistajuudessa onkin ekosysteemityön kannalta kysymys datan suojaamistarpeiden sekä datan avaamisen välisestä tasapainottelusta. On kuitenkin syytä huomata, että kaikkea dataa ei ole välttämätöntä tai taloudellisesti kannattavaa suojata samalla tavalla, vaan suojaamistarpeet on määritettävä datakohtaisesti. Datan suojaamista käsitellään tarkemmin myöhemmissä luvuissa.

Datan lisäksi ekosysteemiin voi kuulua myös muita yhteisiä resursseja, kuten tiedon välityksen, muokkauksen ja pääsynhallinnan mahdollistavia palveluita sekä yhteistä infrastruktuuria, kuten verkkopalveluita, tallennustilaa ja laskentaresursseja.



Kuva 2 - Dataekosysteemin yhteentoimivuuden ja luottamuksen rakennuspalikat

Seuraavissa luvuissa esitellään tarkemmin, miten ekosysteemeissä hyödynnettävä data julkaistaan, miten sitä vaihdetaan sekä miten jakelu lopulta tapahtuu. Lisäksi esitellään tarkemmin, mitä erilaisia teknisiä ratkaisuja voidaan hyödyntää dataekosysteemissä. Lopuksi esitellään vielä erilaiset sidosarkkitehtuurit, jotka toimivat perustana tälle dokumentille, ja joita voidaan hyödyntää virtautuksen suunnittelussa.

2.1. Datan vaihto, jakelu ja federointi dataekosysteemissä

Tässä luvussa esitellään periaatteellisella tasolla, minkälaisia komponentteja tai kyvykkyyksiä dataekosysteemin luomiseen tarvitaan. Lisäksi tarkastellaan, mitä yleisiä asioita niistä on syytä ottaa huomioon jakaessa dataa muille ekosysteemitoimijoille sekä hyödynnettäessä muiden ekosysteemitoimijoiden tuottamaa dataa.

2.1.1. Data hyödykkeenä ja tuotteena

Data voidaan ajatella toisaalta kunnan lakisääteisten perustehtävien toteuttamisessa syntyvänä ja hyödynnettävänä rakenteisena tietona sekä taloudellisena hyödykkeenä, jolla on hyödyntämisestä syntyvää arvoa ja hallinnasta koituvia kustannuksia. Tuotteistettua dataa myyvätkin nykyään erilaiset dataa tuottavat yksityisen sektorin toimijat ja datan välittäjät. Tällöin ostaja hankkii datan myyjältä, kuten minkä tahansa muun tuotteen. Datan välittäminen on usein ajateltu dataa keräävien ja koostavien toimijoiden, kuten mielipidetutkimuksia tuottavien yritysten ja sosiaalisen median toimijoiden liiketoiminnaksi. Kuitenkin myös julkisen sektorin toimijat myyvät esimerkiksi kaavapiirustuksia ja tilastotietoa, mutta tarjoavat yleiseen käyttöön muuta dataa pääasiassa veloituksetta. Dataa voidaan tuotteistaa ja jakaa, vaikka sen hyödyntämisestä ei perittäisi maksua.

Data ei ole perinteisten tuotteiden kaltainen hyödyke muun muassa siitä syystä, ettei tietyllä datalla ole yleensä varsinaista kilpailevaa tuotetta. Datan arvo riippuukin pitkälti käyttötarpeesta ja kasvaa, mitä useammin tai mitä enemmän sitä hyödynnetään. Arvo riippuu myös siitä, minkälaisesta datasta on kyse ja miten se auttaa kehittämään ymmärrystä, palveluja ja tuotteita. Julkinen data ei välttämättä vaadi tarkkaa ja kustannuksia aiheuttavaa hallintaa.

Datan tuotteistamisella tarkoitetaan jaettavan datahyödykkeen paketoimista tuotteeksi, jolla on sisällön lisäksi muun muassa ennalta sovittu muoto, käyttöohjeet sekä käyttöehdot, ja se tuodaan hyödyntäjien saataville tiettyjä kanavia pitkin. Datatuotetta ja sen julkaisemista käsitellään tarkemmin myöhemmissä luvuissa.

2.1.2. Datan jakaminen ja vaihtaminen

Datan jakaminen organisaatioyksiköiden, kuten kuntien, toimialojen, virastojen tai yksityisen sektorin kanssa ja välillä, on ollut jo pitkään tavanomaista toimintaa. Datan vaihtoa on kuitenkin suunniteltu ja toteutettu yleensä vähemmän järjestelmällisesti ja usein kertaluontoisesti, jolloin mahdollisuus ekosysteemien ja kattavampien arvovirtojen kehittymiseen hankaloituu. Tällaisten pistemäisten ja siiloutuneiden ratkaisujen toteutuksissa joudutaan usein suunnittelemaan ja neuvottelemaan, mitä dataa jaetaan tai mitä dataa tarvitaan, datan siirto, mahdolliset muunnokset ja tietoturva lähes kokonaan erikseen muista vastaavista toteutuksista. Siilomainen rakenne on työläs sekä käyttöönoton aikana että ylläpidossa, eivätkä ratkaisun hyödyt ja opit yleensä siirry muiden organisaatioyksiköiden käyttöön. Myöskään luotuja osaratkaisuja ei kyetä uudelleenkäyttämään vastaavissa toteutuksissa. Varsin usein tällaisten kertaluontoisten toteutusten arkkitehtuuri perustuu tiedostosiirtoihin, eikä niille suunnitella tai toteuteta yhteisiä ratkaisuja hyödyntäviä integraatiomalleja taikka järjestelmällistä ylläpitoa. Lisäksi tällaisten ratkaisujen tietoturvaa ja datan omistajuutta voi olla hankalaa hallita tehokkaasti, varsinkin kun hallittavia tietovirtoja on useampia, asiantuntijaresursseja vähän sekä tekniset ratkaisut hajanaisia ja toisistaan poikkeavia. Datan vaihdon toteutuksia voidaankin tehostaa määrittämällä yhteisiä standardoituja ratkaisukomponentteja, jotka mahdollistavat toteutusten nopeamman suunnittelun ja käyttöönoton sekä paremman läpinäkyvyyden ja valvonnan.

Datan jakamista on standardoitu ja rakenteellistettu viime vuosikymmeninä alkaen yhteisistä tietoformaateista, EDI-sanomista ja master datan eli ydintiedon hallinnan kehittymisestä aina teollisuusstandardoitujen tapahtumasanomien keräämiseen asioiden internetistä (Internet of Things eli IoT) sekä lähilukijoista. Dataekosysteemityö jatkaakin osaltaan tätä kehitystä standardisoimalla dataresurssien (ja muiden ekosysteemiresurssien) omistajuuteen ja käyttöehtoihin liittyvän sopimisen sekä sopimukset eli tiedonvaihdon hallinnan.

Useiden eri toimijoiden ja tietolähteiden dataa on mahdollista koostaa ja yhdistellä usealla eri tavalla. Yksinkertaisimmillaan dataa voidaan yhdistellä ja integroida useista eri lähteistä, mutta tällöin tietolähteiden datat tulee olla kuvattu samalla tavalla ja samanlaisissa rakenteissa, jotta data saadaan koostettua yhtenäiseksi kokonaisuudeksi. Tästä syystä useilla toimialoilla ja yhteisöissä on otettu käyttöön yhteisiä tietomalleja ja datan metatietoja kuvaavia rakenteita, joiden kautta tietoa voidaan yhdistellä. Yhdistely voi tapahtua rakentamalla tietolähteiden tietomallit samoilla formaateilla tai konvertoimalla tietoja samoihin rakenteisiin yhdistelyvaiheessa. Tässäkin tapauksessa joudutaan yhdistelyjä ja tietovirtojen suorittamista toteuttavat integraatiot usein luomaan tapauskohtaisesti ja yhdisteltyjä tietoja ylläpitämään keskitetyssä lähteessä esimerkiksi tietovarastossa, josta niitä jaetaan eteenpäin. Tällainen tiedon tallennuksen keskittäminen voi kuitenkin aiheuttaa tietoturva- ja suojariskejä, eikä alkuperäinen datan omistaja kykene välttämättä hallitsemaan, mihin dataa jaetaan ja miten sitä saa käyttää. Useimmat kunnat ovat tässä tilanteessa.

Modernit ekosysteemiarkkitehtuurit perustuvat data-avaruuksiin, joissa dataa säilytetään tietolähteessä, esimerkiksi jonkin järjestelmän tietokannassa ja jaetaan sieltä suoraan datan tarvitsijoille vain silloin, kun dataa tarvitaan. Datan jakaminen tarjoajien ja vastaanottajien välillä perustuu tällöin luottamukseen ja yhteisiin sääntöihin ja standardeihin, jotka määrittävät, miten dataa talletetaan ja jaetaan. Data-avaruuksiin perustuva ekosysteemityö jatkaakin osaltaan datan hallinnan kehitystä standardisoimalla datan (ja muiden ekosysteemiresurssien) omistajuuteen ja käyttöehtoihin liittyvän sopimisen sekä sopimukset eli tiedonvaihdon hallinnan. Datan tarjoajien ja vastaanottajien lisäksi data-avaruuksiin perustuvassa ekosysteemissä voi toimia myös datan välittäjätahoja. Tämän dokumentin kuvaama ekosysteemiarkkitehtuuri perustuukin pääosin data-avaruusarkkitehtuuriin.

Ekosysteemityössä on vielä syytä erottaa datan vaihto, *engl. data exchange*, ja datan jakaminen, *engl. data sharing*. Datan vaihdolla tarkoitetaan datojen siirtelyä vertikaalisten arvoketjujen tukemiseksi,

luomiseksi ja optimoimiseksi esimerkiksi julkishallinnon sisällä. Käytännön esimerkkinä tällaisesta datan siirtelystä on kunnan asiakastietojen kerääminen eri palveluista vastaavista yksiköistä kuntatasolle ja vastaavasti näiden tietojen keräämisestä kunnilta ministeriöille ja laitoksille, jotka tekevät saatujen datojen perusteella tilastointia sekä vertailua ja päättävät resurssien kohdistamistarpeita valtakunnallisesti. Tällaisissa tapauksissa datan keräämisestä ja jakelusta on voitu säätää laeilla. Lisäksi tietojen siirtämistä on usein sujuvoitettu yhteisten toimintatapojen sekä tiedonsiirto- ja esitysmuotostandardien avulla.

Datan jakamisella puolestaan tarkoitetaan sekä vertikaalisesti että horisontaalisesti tapahtuvaa datan siirtelyä ja yhteistyötä ekosysteemin sisällä yhteisten arvontuottotavoitteiden saavuttamiseksi. Tällaista on esimerkiksi palveluketjun sisällä tapahtuva tiedonsiirto, kuten kunnan palveluyksikön, kaavoituksen, tilasuunnittelun ja palveluverkkosuunnittelun välinen yhteistyö palvelutilojen sijoittamisesta kuntalaisten kannalta edulliseen sijaintiin. Datan jakamista on yleensä standardoitu datan vaihtoa vähemmän, ja toimintatavat on sovittu tapauskohtaisesti yleensä organisaatioyksikön kuten kunnan sisällä. Dataekosysteemien tarjoamat yhteiset palvelut ja standardit mahdollistaisivat kuitenkin myös datan jakamisen järjestelmällisemmin esimerkiksi kuntien välillä yhteisten palveluketjujen kehittämiseksi sekä ratkaisujen jakamiseksi yhdessä kohdattuihin haasteisiin.

Datan ja muiden resurssien ollessa hajautettu useille eri tietolähteille, toimijoille ja järjestelmille, on datan ja resurssien luotettavan jakamisen edellytys tuoda sekä resursseja että toimijoita kuvaavat tiedot yhteisesti nähtäville. Yleisesti ottaen dataekosysteemeissä datatuotteita ja muita resursseja kuvaavat metatiedot ovat federoitu ja julkaistu ekosysteemin yhteiseen katalogiin, josta ne ovat haettavissa ekosysteemin eri toimijoille. Datatuotteilla tuleekin olla kuvaukset itsestään, joita voidaan jakaa datan ja siihen liittyvien käyttöehtojen löytämisen helpottamiseksi. Lisäksi dataekosysteemeissä on hyvä olla federoidu identiteetinhallinta, jonka avulla ekosysteemitoimijat voivat nähdä toistensa tietoja, muun muassa luottamustasoja (jos sellaisia käytetään), mikä edesauttaa välttämään datan päätymistä vääriin käsiin. Federointipalveluja kuvataan tarkemmin luvussa 3.4.

3. Dataekosysteemin arkkitehtuurin suunnittelu

Dataekosysteemien toimivuuden perusta syntyy yhteisistä pelisäännöistä ja niiden noudattamisesta. Tämä edellyttää sopimuksellista määrittelyä ekosysteemin toimijoiden yhteistyölle, tiedon hallinnalle ekosysteemissä ja datan hallinnan teknisille ratkaisuille. Tässä luvussa esitellään tarkemmin dataekosysteemien keskeiset osakokonaisuudet, joiden avulla kyetään rakentamaan ekosysteemille arkkitehtuuri, joka mahdollistaa tiedonvaihdon ekosysteemin sisällä luotettavasti, läpinäkyvästi ja turvallisesti. Nämä osakokonaisuudet koostuvat:

- Toimintamallista, joka ohjaa ekosysteemin organisointia sekä prosesseja.
- Tiedon ja tietomallien tuotteistamisesta sekä julkaisemisesta luotettavaa jakamista varten.
- Teknisistä määrityksistä dataekosysteemin keskeisille järjestelmä- ja integraatiopalveluille.

Ekosysteemin arkkitehtuurin voi rakentaa implementoimalla kaikki seuraavissa luvuissa esitellyt ekosysteemiarkkitehtuurin komponentit, joskaan tämä ei ole välttämätöntä. Kunnat voivatkin halutessaan implementoida ja kokeilla ekosysteemiarkkitehtuurin keskeisiä elementtejä erilaisten käyttötapausten avulla. Tässäkin tapauksessa on käytön laajetessa hyvä sopia jo etukäteen ekosysteemin yhteisistä toimintatavoista, standardeista ja teknisistä ratkaisuista, jotta tiedon omistajuuden ja jakamisen hallinta on läpinäkyvämpää ja luotettavampaa.

Luvussa esiteltävät kuvaukset perustuvat pitkälti IDS-RAM, Gaia-X ja OASC MIMs -määrityksiin, joista on poimittu mukaan keskeiset elementit ja joihin viitataan siten, että lukija tietää mistä tiettyjä tässä

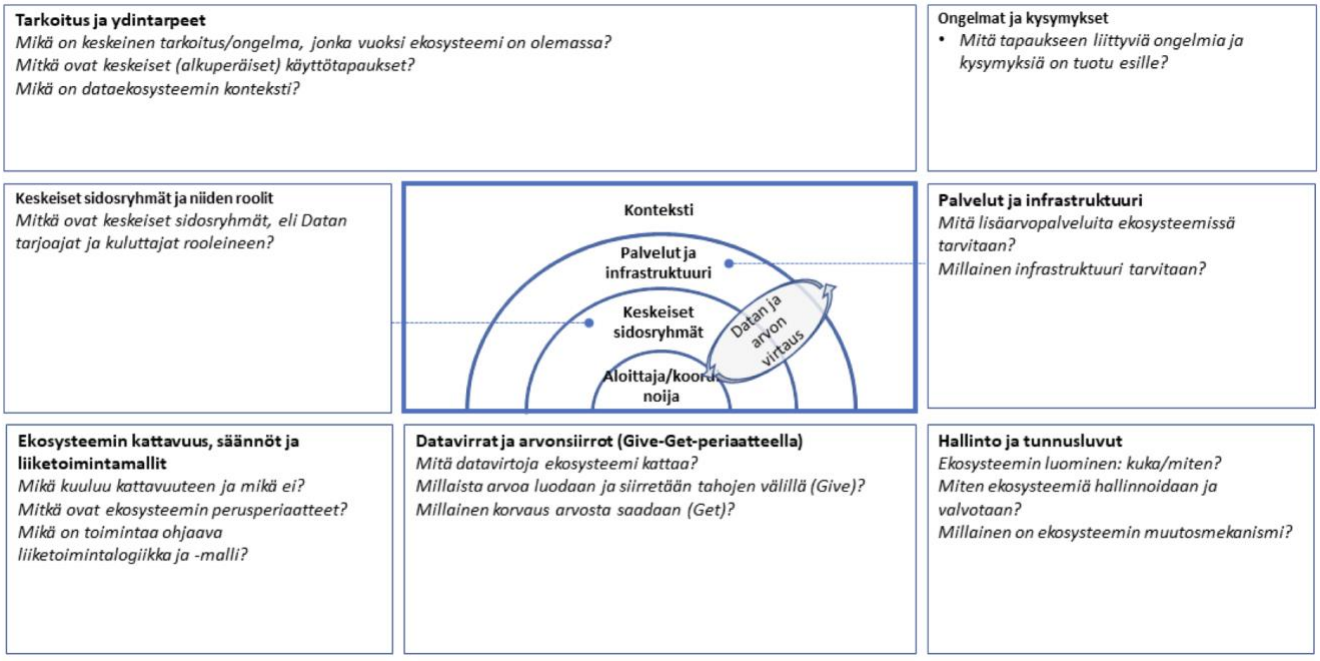
dokumentissa tarkentamattomia asioita voi lukea. Lisäksi kuvausten pohjana on hyödynnetty Sitran julkaisemaa Reilun datatalouden sääntökirjaa ja julkishallinnon API-linjauksia.

Tässä ohjeistuksessa ei keskitytä kuvaamaan tarkasti, miten kunta luo itse dataekosysteemejä vaan syvennytään siihen, miten kunta voi tuottaa dataa ekosysteemiin ja hyödyntää eri dataekosysteemeistä saatavaa dataa. Seuraavan luvun alussa on kuitenkin kevyt ohjeistus dataekosysteemin suunnittelusta, mikä ohjaa alkuun niissä tapauksissa, joissa kunta haluaa perustaa sisäisiä data-alustoja tai -ekosysteemejä, tai uuden dataekosysteemin muiden toimijoiden kanssa.

3.1. Dataekosysteemin suunnittelun aloitus

Dataekosysteemin suunnittelu kannattaa aloittaa määrittelemällä tarkoitus ja raamit dataekosysteemin olemassaololle. Suunnittelussa voidaan hyödyntää esimerkiksi Reilun datatalouden sääntökirjaa (Sitra). Sen tarkoituksena on tarjota helposti saavutettava opas dataverkoston perustamiseen sekä datanjakosopimusten yleisten ehtojen määrittelyyn. Sääntökirjamalli auttaa kuntia ja muita organisaatioita muodostamaan uusia dataekosysteemejä ja -verkostoja, laatimaan niille yhteisiä sääntökirjoja ja edistämään yleistä reilua datataloutta.

Alla oleva Reilun datatalouden sääntökirjassa esitelty ekosysteemimallikaavio avustaa kuntaa vastaamaan ekosysteemin tarkoitusta ja hallintaa koskeviin kysymyksiin sekä auttaa alkuun ekosysteemityössä. Ekosysteemimallia hyödynnetään ekosysteemin suunnittelu- ja rakentamistyössä sekä sen hallinnassa ja viestimisessä.



Kuva 3 - Reilun datatalouden sääntökirjan mukainen ekosysteemikuvaus

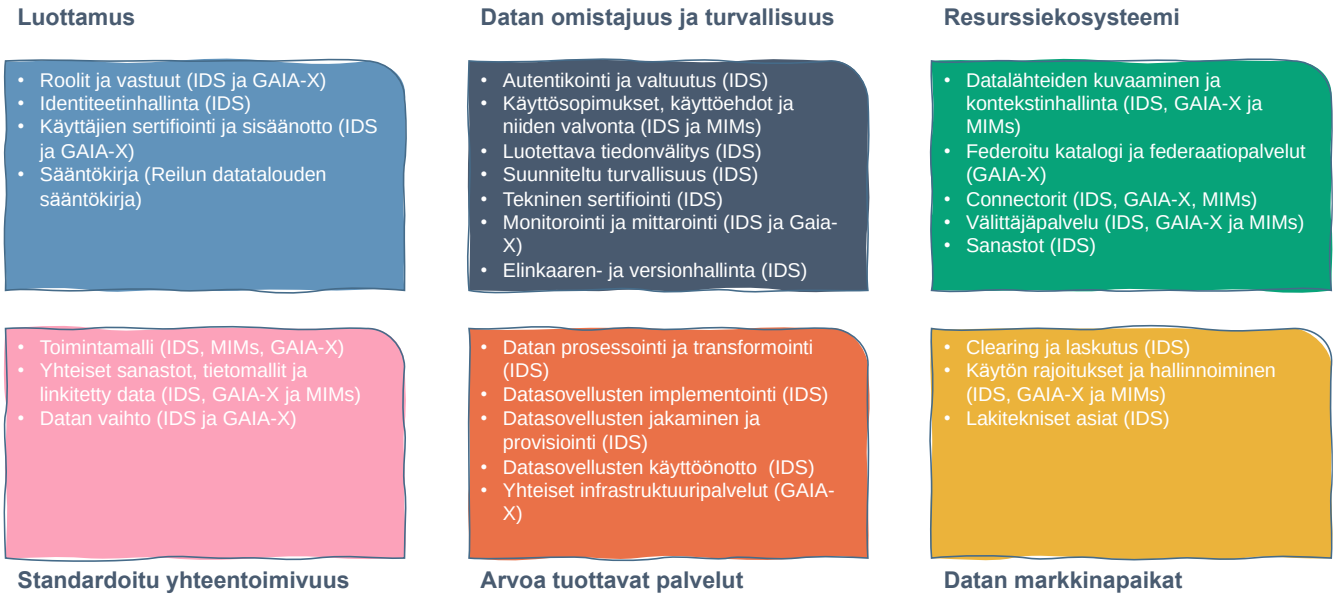
Ekosysteemimalli kannattaa laatia ensin niissäkin tapauksissa, joissa kehittäminen aloitetaan yksittäisistä käyttötapauksista, jos voidaan ennakoida käyttötapausten määrän, siirtovolyymien ja ekosysteemin toimijoiden määrän kasvavan.

Tässä dokumentissa ei mennä syvemmälle dataverkoston sääntökirjamallin osioihin, mutta myöhemmin esiteltyjä komponentteja voidaan tarkentaa sääntökirjamallissa esitettyjen tukikysymysten avulla. Dataekosysteemeille suositellaan luomaan myös oma sääntökirja, jossa tarkennetaan, kuinka ekosysteemi toimii ja mitä sen jäseniltä sekä teknisiltä ratkaisuilta odotetaan. Sääntökirjalla on mahdollista edistää toiminnan läpinäkyvyyttä ja sitä kautta reiluutta, yhteentoimivuutta sekä ekosysteemin jäsenten välistä luottamusta.

Seuraavassa luvussa esitellään lyhyesti dataekosysteemin ja yhteentoimivuuden keskeiset toiminnalliset elementit, joiden käyttöä ohjeistetaan tarkemmin luvun 3 myöhemmissä aliluvuissa. Komponenttien tuntemus auttaa ekosysteemin toiminnan määrittämisessä.

3.2. Dataekosysteemin ja yhteentoimivuuden keskeiset toiminnalliset elementit

Tässä luvussa käydään läpi dataekosysteemin mahdollistavat keskeiset elementit, jotka löytyvät alla esitetystä kaaviosta. Kaavio on kuvattu IDS:n tekemällä jaottelulla data-avaruusarkkitehtuurin mukaiselle toiminnallisille vaatimuksille ja sitä on täydennetty Gaia-X:stä sekä OASC:n MIM:eistä tunnistetuilla muilla elementeillä. Ei-toiminnalliset vaatimukset, eli suoraan ekosysteemin toimintaan liittymättömät vaatimukset, kuten skaalautuminen määritellään ekosysteemikohtaisesti sen käyttötarkoitusten perusteella. Ei-toiminnallisia vaatimuksia käsitellään tarkemmin myöhemmissä luvuissa.



Kuva 4 - Dataekosysteemin toiminnalliset komponentit (taydennetty IDS:n toiminnallisista vaatimuksista)

Kaaviossa esiteltyjen elementtien perään on lisätty vielä viitearkkitehtuurit, joita on hyödynnetty elementtien kuvauksissa ja joista löytää lisätietoja dataekosysteemien suunnitteluun.

3.2.1. Luottamus

Luottamus on yksi keskeisimmistä vaatimuksista ekosysteemin toiminnalle ja sitä hallitaan yleensä seuraavien neljän elementin avulla:

Vaatimuselementti	Kuvaus
Roolit ja vastuut	Jokaisella ekosysteemin toimintaan liittyvällä toimijalla on rooli tai useita rooleja, joihin liittyy tiettyjä oikeuksia ja vastuutehtäviä.
Identiteetinhallinta	Jokaisella ekosysteemiin liittyvällä dataa tai palveluita tarjoavalla taikka hyödyntävällä connectorikomponentilla tulee olla yksilöivä identiteetti ja yleensä voimassa oleva sertifiointi tai sopimus. Ekosysteemin toimijoiden tulee myös kyetä tunnistamaan ja verifioimaan ekosysteemin connectorit.
Käyttäjien sertifiointi ja sisäänotto	Jokainen ekosysteemiroolissa toimiva taho sertifioidaan ja käyttösopimukset luodaan läpinäkyvästi yhteistä prosessia noudattaen. Tällä varmistetaan ekosysteemin osallistujien luotettavuus.
Sääntökirja	Yhteisesti noudatettu sääntökirja selkeyttää ekosysteemin toimintatapoja ja tuo läpinäkyvyyttä toimintaan, lisäten luottamusta osallistujille.

3.2.2. Datan omistajuus ja turvallisuus

Datan omistajuus ja turvallisuus ovat keskeisiä haasteita datan jakamisessa erilaisissa ekosysteemeissä. Näitä haasteita voidaan välttää seuraavilla seitsemällä vaatimuksella:

Vaatimuselementti	Kuvaus
Autentikointi ja valtuutus	Jokaisella ekosysteemiin liitetyllä connectorilla tulee olla voimassa oleva sertifikaatti tai käyttösopimus, jonka muut ekosysteemitoimijat voivat varmistaa. Tällä luodaan turvallisuutta datan jakamiselle ja vaihdolle. On suositeltavaa, että ekosysteemin datan tai muun resurssin jakaja pystyy varmentamaan vastaanottajan turvallisuustoimenpiteet ja identiteetin. Sensitiivisen datan tai muun resurssin suojauksessa on syytä hyödyntää päästä päähän valtuutusta, jolloin datan jakaja voi varmistua, että resurssi päätyy sovitulle kohteelle (endpointille). Tällöin myös kommunikointi endpointien välillä on syytä kryptata. Vaatimukset suojaukselle voidaan määritellä tapaus- tai ekosysteemikohtaisesti.
Käyttösopimukset, käyttöehdot ja niiden valvonta	Jaettaville datatuotteille ja muille resursseille tulee voida asettaa käyttöehdot ja niiden käytölle käyttösopimukset. Käyttöehdot saattavat sisältää rajoituksia datan hyödyntämiselle, kuten datan säilytysajoille tai datan siirtämiselle kolmansille osapuolille. Sopimusten ja käyttöehtojen noudattamista on syytä myös valvoa ja niiden rikkomisesta tulla seurauksia, kuten käyttöoikeuden menetys. Osa valvonnasta voidaan toteuttaa automaattisesti clearing- ja lokipalveluiden avulla.
Luotettava tiedonvälitys	Connectorien, datasovelluskauppojen ja välittäjien tulisi kyetä näkemään hyödyntäjien identiteettitiedot ja sertifioinnit ja datan käyttäjästä tulisi kyetä varmistumaan. Kommunikoinnin luotettavuutta voidaan vielä lisätä suojausvaatimuksilla ja automaattisella käytönvalvonnalla.
Suunniteltu turvallisuus	Kaikki datat ja connectorit eivät välttämättä tarvitse samantasoista suojaa. Connectorien ja ekosysteemiresurssien sekä tiedonsiirron suojaus kannattaakin mitoittaa sopivaksi sekä dataa luovuttavassa että vastaanottavassa päässä.
Tekninen sertifiointi	Keskeisimmille ekosysteemin teknisille komponenteille suositellaan laadittavaksi tekniset sertifiointi- ja varmentamistavat. On syytä varmistua erityisesti tiedon välittämisestä ja korkeansuojausluokan datan jakamisesta, siirtämisestä ja vastaanottamisesta vastaavien connectorien luotettavuudesta. Tämä voi vaatia keskitettyä sertifiointia, joka suorittaa

	teknisen tarkastuksen tarvittaessa connectorin tai palvelun fyysisessä sijainnissa. Sertifiointiprosesseja kuvataan tarkemmin toimintamalliluvussa.
Monitorointi ja mittarointi	Monitorointi ja mittarointi ovat hyödyllisiä palveluita ekosysteemin toimivuuden ja käytön valvonnan kannalta. Monitoroinnilla tarkoitetaan ekosysteemikomponenttien tilan valvontaa ja sen mahdollistamia hälytyksiä. Jos jokin komponentti, kuten connectori vikaantuu, mahdollistaa monitorointi automaattiset hälytykset connectorin hyödyntäjille. Monitorointi auttaa myös lokien hakemisessa ja ekosysteemin toiminnan tilastoinnissa. Monitorointioikeuksia hallitaan käyttösopimuksin. Mittaroinnilla tarkoitetaan monitoroinnin tapaista palvelua, jolla seurataan erilaisia indikaattoreita sekä käyttöstatistiikkaa. Mittarointi tukee erityisesti connectorien ja muiden palveluiden käyttölaskutusta.
Elinkaaren- ja versiohallinta	Elinkaaren- ja versionhallinta ekosysteemien resursseille on ensiarvoisen tärkeää. Ekosysteemiresurssien versiot sekä elinkaaren tila tulee kyetä tunnistamaan ja muutoksista kommunikoimaan, jotta niihin kyetään varautumaan hyödyntävässä päässä.

3.2.3. Resurssiekosysteemi

Dataekosysteemissä tai missä tahansa muussa digitaalisten resurssien ekosysteemissä on tärkeää, että jaettavaa dataa on mahdollista kuvailla ja se on löydettävissä sekä tulkittavissa, jotta sitä kyetään hyödyntämään.

Vaatimuselementti	Kuvaus
Datan- ja muiden resurssien lähteiden kuvaaminen ja kontekstinhallinta	Dataekosysteemin toimijoiden tulisi kyetä kuvaamaan jaettava resurssinsa sekä julkaista, ylläpitää ja versioida resurssia kuvaava metadata, jotta itse data voidaan tuotteistaa. Datan tuotteistusta on kuvattu tarkemmin luvussa 3.3. Kontekstinhallinnalla tarkoitetaan tässä edellä mainitun metadatan hallintaa siten, että jaettavia resursseja kuvaavat tiedot ovat löydettävissä ja haettavissa. Kontekstinhallinta on yleensä liitetty ekosysteemin federoituun katalogiin. Se voi olla rajapintapalvelu, jonka kautta katalogi on saavutettavissa tai oma palvelunsa (esimerkiksi sivusto tai GitHub-kirjasto). Kontekstinhallintaa on määritelty tarkemmin OASC:in MIM1:ssä.
Federoitu katalogi ja federaatiopalvelut	Federoitu resurssikatalogi koostaa useiden datatuotteiden ja muiden resurssien kuvaukset ekosysteemitoimijoiden nähtäville. Katalogeihin voidaan koostaa yhden tai useamman resurssien jakajan jakamia connectoreita kirjastoiksi, joiden avulla niiden hyödyntäjät kykenevät löytämään tarvitsemansa resurssit ja niitä tarjoavat connectorit. Federoitu katalogi on myös osa ekosysteemipalvelujen hallintaa siten, että sen avulla voidaan toteuttaa palvelunhallintaa käyttösopimuksien- ja ehtojen mukaisesti, esimerkiksi rajoittamalla resurssien näkyvyyttä vain sopimusten osapuolten välille. Katalogeja kuvataan tarkemmin luvussa 3.3.
Connectorit	Connectori on datan vaihdon keskeinen tekninen komponentti, joka vastaa datan siirtämisprosessin suorittamisesta tai reitittämisestä osallistujaroolin sisäisiin tietovarantoihin tai niistä ulospäin. Connectori voi muodostua yhdestä tai useammasta data endpointista, eli se voi koostaa

	dataa yhdestä tai useammasta lähteestä sille asetetun tietomallin, kyselyiden ja datan transformaatioprosessien mukaisesti. Connectorien tiedot julkaistaan yleensä federoidussa katalogissa välittäjäpalvelun avulla. Julkaistavia tietoja ovat muun muassa rajapintakuvaus, autentikointimekanismit, julkaistava data ja datan käyttöehdot. Yksittäinen connectori voidaan julkaista useammassa ekosysteemissa eri välityspalvelujen kautta, mikäli käyttöehdot ja ekosysteemien sääntökirjat sen mahdollistavat. Connectoreita tarkastellaan tarkemmin luvussa 3.4.
Välittäjäpalvelu	Jokaisen ekosysteemin connectorin ylläpitäjän on kyettävä julkaisemaan data- ja metadatarajapintansa, ja jokaisen connectorin on kyettävä lähettämään metadataansa yhden tai useamman välittäjäpalvelun kautta. Ekosysteemitoimijoiden tulee olla mahdollista etsiä ja selata jaettujen resurssien metadatoja välittäjäpalvelun kautta julkaistusta federoidusta katalogista tai muusta metadatakirjastosta omien käyttösopimustensa rajoissa. On myös suositeltavaa luotettavuuden ja läpinäkyvyyden vuoksi, että välittäjäpalvelun avulla sen käyttäjät voisivat nähdä välittäjäpalveluun rekisteröidyt muut käyttäjät. Välittäjäpalvelua tarkastellaan tarkemmin luvussa 3.4.
Sanastot	Connectorin julkaisija voi hyödyntää connectorin metadatan luomiseen ja rakenteistamiseen erilaisia olemassa olevia, itse luotuja tai yhteistyössä luotuja sanastoja. Sanastot mahdollistavat nopeamman tiedon vaihdon ekosysteemin sisällä, jos ne on standardoitu. Sanastopalveluja käsitellään tarkemmin luvussa 3.4.

3.2.4. Standardoitu yhteentoimivuus

Standardoitu datan vaihto ja jakaminen on dataekosysteemien keskeinen tavoite. Yhteentoimivuutta voidaan edistää ekosysteemeissä sekä toimintaan, tietoon että järjestelmiin liittyvällä standardoinnilla.

Vaatimuselementti	Kuvaus
Toimintamalli	Toimintamallilla pyritään yhtenäistämään ekosysteemin toimintatapoja sekä roolien oikeuksia ja vastuita. Toimintamallia kuvataan tarkemmin luvussa 3.2.
Connectorien toiminta	Connectorien toiminta tulisi olla konfiguroitavissa ja niiden sisäiset työnkulut määritettävissä. Lisäksi connectorien käyttäjät pitää pystyä tunnistamaan ja käyttöä hallitsemaan. Connectorien toiminta määritellään käyttöehdoilla ja connectorit ja niihin liittyvät datan säilytysratkaisut tulisi kyetä suojaamaan salasanoin. Connectorien käytöstä tulee ylläpitää lokia, jotta käyttöä kyetään valvomaan ja ongelmiin puuttumaan. Connectorien yhteentoimivuudessa on myös huomioitava, minne ja miten ne on implementoitu. Esimerkiksi suoraan datan lähdejärjestelmiin tai alustapalveluihin rakennetut connectorit voivat hankaloittaa teknistä yhteentoimivuutta tiedonsiirtoformaattien, tietoon pääsyn, tiedon jakamiseen ja transformointiin liittyvien kapasiteetin skaalautuvuuden, autentikoinnin yms. teknisten rajoitusten kautta. Nämä rajoitukset on syytä pitää mielessä connectoria tai sen hyödyntämistä suunniteltaessa. Connectoreita on kuvattu tarkemmin luvussa 3.4.

Yhteiset käsitteistöt, tietomallit ja linkitetty data	Ekosysteemeissä hyödynnetään usein yhteisiä käsitteistöjä, tietomalleja ja linkitettyä dataa semanttisen ja rakenteellisen yhteentoimivuuden varmistamiseksi. Näitä kuvataan tarkemmin luvussa 3.3.
Datan vaihto	Connectorien tulee kyetä saamaan dataa lähdejärjestelmistä taikka viemään sitä kohdejärjestelmiin joko push- tai pull-periaatteella. Keskeistä on suunnitella datan jakaminen siten, että tietovirran mahdollistavien teknisten komponenttien yhteentoimivuus kyetään varmistamaan. Dataa voidaan jakaa lähdetietovarannosta rajapinnan kautta tai työntämällä sitä suoraan muille ekosysteemitoimijoille push-periaatteella. Jotta tämä olisi mahdollista, tulee connectorit kyetä yksilöimään. Kaikista lähdejärjestelmistä tai niihin luoduista connectoreista ei ole mahdollista saada dataa tai muita resursseja aina suoraan halutussa formaatissa tai halutuilla metodeilla. Tällöin on syytä käyttää erillisiä datan siirto ja konversiopalveluita. Näitä kuvataan tarkemmin seuraavassa luvussa.

3.2.5. Arvoa tuottavat palvelut

Jaettava data ei usein ole tarjolla suoraan sen hyödyntäjän haluamassa formaatissa. Tällöin dataa saatetaan joutua muokkaamaan ennen datan siirtoa tai sen jälkeen. Ekosysteemissä tulisikin olla työkaluja datan muokkaamiseen haluttuun muotoon. Dataa muokkaavat palvelut nopeuttavat datan saamista aiottuun käyttöön ja tuottavat näin lisäarvoa tarjolla olevalle datalle.

Vaatimuselementti	Kuvaus
Datan prosessointi ja transformointi	Datan prosessointipalvelut ja -sovellukset toteuttavat yksittäisiä ja tarkasti rajattuja prosessointitoiminnallisuuksia, jotka prosessoivat dataa ennustettavasti ja palauttavat sen samassa formaatissa. Datan transformointipalvelut ja -sovellukset puolestaan kykenevät transformoimaan dataa formaatista toiseen datan hyödyntäjien tarpeita vastaavasti, ilman, että datasisältö itsessään muuttuu oleellisesti tai osa datasta katoaa.
Datasovellusten implementointi	Datasovelluksia tulisi kyetä kuvailemaan metadatalla (muun muassa toiminnot, rajapinnat, mahdollinen käytön tai hankkimisen hinnoittelu, lisenssit ja käyttöehdot). Datasovelluksista tulisi ilmoittaa ekosysteemitoimijoille ainakin rajapinnat, riippuvuudet ja pääsynhallintavaatimukset.
Datasovellusten jakaminen ja provisiointi	Datasovellukset tulisi sertifioida ekosysteemissä samalla tavalla, kuin muut resurssit, varsinkin jos niiden käyttö mahdollistetaan useille ekosysteemitoimijoille. Datasovelluksia voidaan julkaista ja provisioida esimerkiksi yhteisten sovelluskauppojen avulla ja ne tulisi olla löydettävissä katalogeista. Tässä kannattaa huomioida turvallisuus tavanomaista tarkemmin, esimerkiksi vahvan tunnistautumisen avulla pääkäyttäjille, jotta voidaan välttää riskiä datan väärinkäytöltä ja päätymiseltä ei-haluttuihin käsiin.
Datasovellusten käyttöönotto	Turvallisuuden lisäämiseksi datasovelluksia on usein syytä jakaa erillisten connectoripalvelujen kautta, jotka asentavat sovelluksen ja poistavat asennuksen tarvittaessa. Käyttöönottoprosessi suositellaan rakentamaan sellaiseksi, että se tarjoaa tukea sovelluksen löytämiseen, asennukseen ja sen hallintaan (esimerkiksi päivitykset).

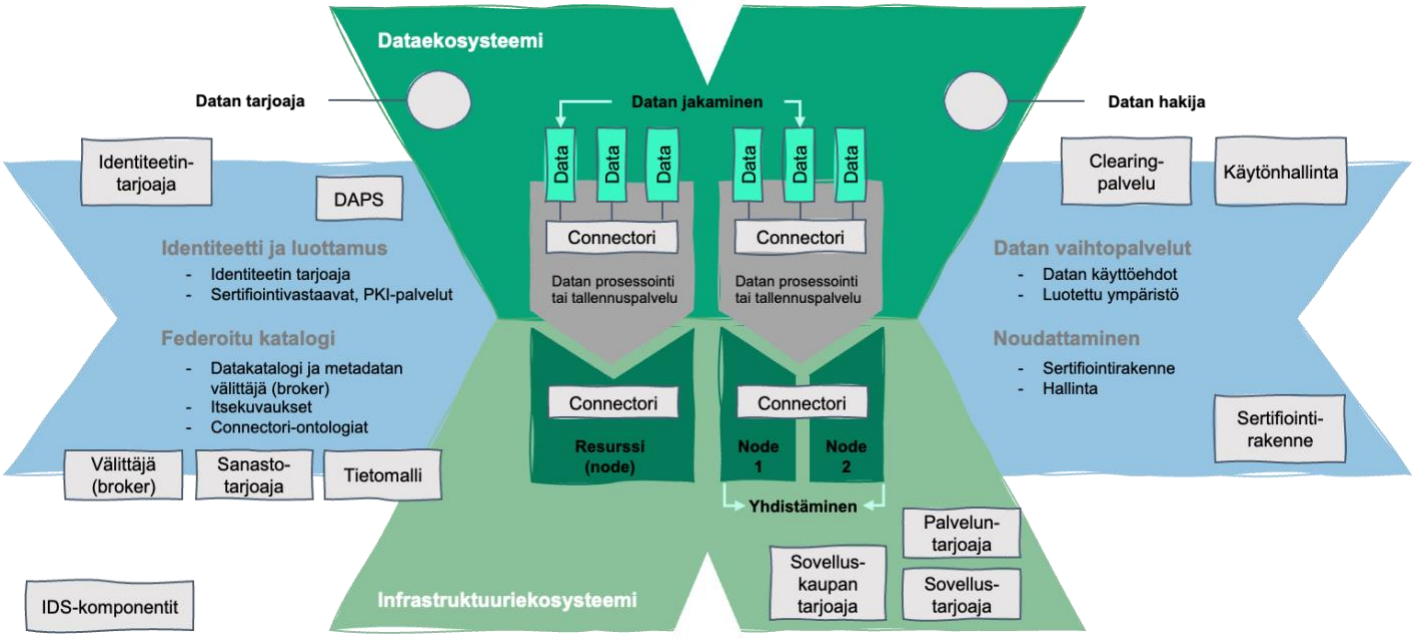
Yhteiset infrastruktuuripalvelut	Ekosysteemeissä voi olla sovellusten lisäksi myös muita jaettuja resursseja, kuten laskentakapasiteettia tai tallennustilaa, joka on samalla tavoin tuotteistettu, kuin muutkin resurssit. Mikäli ekosysteemin palveluja ylläpidetään ja julkaistaan jonkun tietyn toimijan ylläpitämässä ympäristössä, on syytä huomioida ympäristöön kohdistuvat skaalautumisvaatimukset jo ekosysteemiä suunniteltaessa.

3.2.6. Datan markkinapaikat

Datan markkinapaikat vastaavat tarpeeseen valvoa ja hallinnoida dataresurssien käyttöä tarkemmin. Ekosysteemin datoille on syytä määrittää markkinapaikka ja siihen kuuluvat komponentit erityisesti niissä tapauksissa, joissa dataresursseja koskevista transaktioista odotetaan rahallista korvausta.

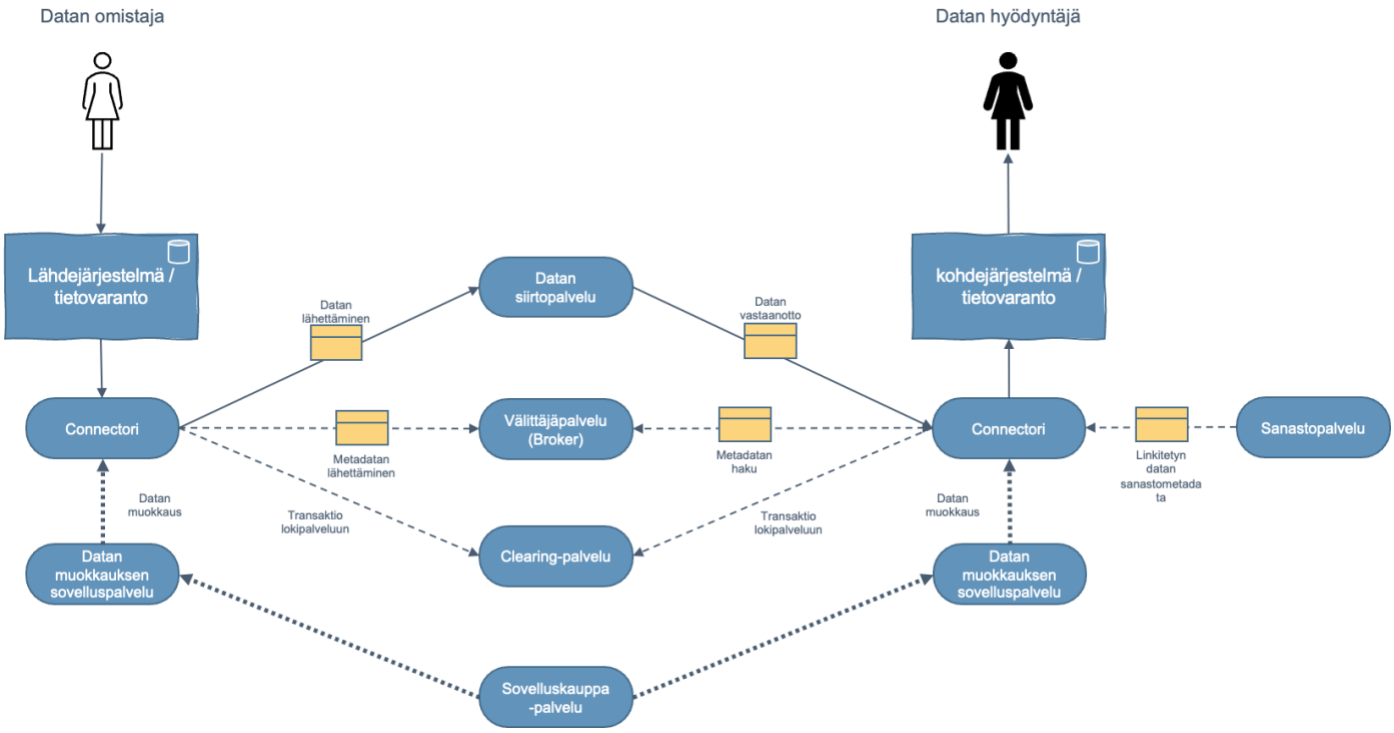
Vaatimuselementti	Kuvaus
Clearing ja laskutus	Mikäli resurssin omistaja luo jakamalleen resurssille hinnoittelumallin ja hinnan, on resurssin käyttöä ja käyttökertoja syytä seurata. Tällöin ekosysteemissä on suositeltavaa ottaa käyttöön clearingpalvelu, joka lokittaa ja validoi transaktioiden suorittamisen. Transaktiolokia voidaan hyödyntää laskutuksen perustana. Laskutusperusteet itsessään on kuvattava resurssin käyttöehtoihin.
Käytön rajoitukset ja hallinnoiminen	Käytön rajoitukset ja käytönhallinta liittyy datan tuottajien, jakajien, hankkijoiden, välittäjien ja loppukäyttäjien välisiin käyttösopimuksiin sekä ekosysteemin datan markkinapaikkojen yhteisiin sääntöihin.
Lakitekniset asiat	Mikäli datasta käydään kauppaa, tulee kaupankäynnille laatia lainmukaiset sopimukset ja ehdot (sekä muun muassa verotukselliset tiedot), joiden avulla kaupankäyntiä voidaan automatisoida. Tämä vaatii myös sopimusten standardointia, jotta niistä ei tarvitse neuvotella transaktiokohtaisesti.

Alla esitetty kaavio tarkentaa, miten IDS:n komponentit sijoittuvat Gaia-X:n ekosysteemimallissa:



Kuva 5 - IDS:n komponentit Gaia-X:n viitekehyksessä

Seuraava kaavio puolestaan kuvaa, miten datan siirto toimii käytännössä ekosysteemin sisällä.



Kuva 6 - Dataekosysteemin keskeiset komponentit datan siirtoon (IDS-RAM)

3.3. Ekosysteemin hallintamalli

Tässä kappaleessa kuvataan ekosysteemiin liittymisen hallintamallia, eli datatuotteen tuottajan ja hyödyntäjän välistä suhdetta sekä oikeuksia ja vastuita. Datatuotteen hyödyntäjänä toimimisen prosessi on suoraviivaisempi, joten tässä luvussa keskitytään hallintamallin kuvaukseen, jolla **kunta tai organisaatio voi toimia datatuotteen tuottajan roolissa**. Luvussa kuvatut kyvykkyydet voivat löytyä kunnan sisäisistä resursseista tai ne voidaan hankkia palveluna eli konsultointina.

Hallintamalli perustuu oleellisin osin Gaia-X ja IDS-viitekehyksiin:

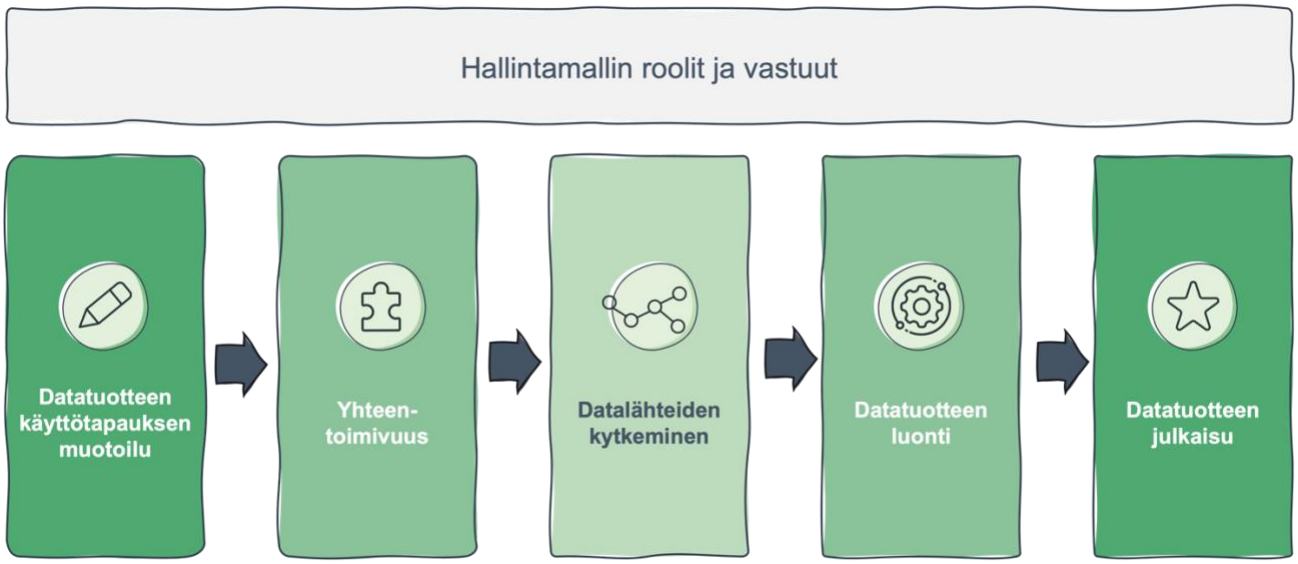
Gaia-X (https://gaia-x.eu/):

- tiedon tuottajan ja hyödyntäjän välinen suhde
- ekosysteemiroolien oikeudet ja vastuut
- identiteetin- ja luottamuksenhallinta
- luottamusviitekehys
- palveluluokat
- federaatio, jakaminen ja hajauttaminen

IDS (https://internationaldataspaces.org/):

- data governance
- data hyödykkeenä
- datan omistajuus
- datan laadunhallinta
- datan jäljitettävyyden hallinta

Alla oleva kaavio kuvaa datatuotteen tuottamisen hallintamallin vaiheet:



Kuva 7 – Datatuotteen tuottamisen hallintamallin vaiheet

3.3.1. Hallintamallin roolit ja vastuut

Jotta organisaatio tai kunta voi toimia datatuotteen tuottajana tai -kuluttajana ekosysteemissä, seuraavat roolit täytyy olla määriteltynä. Käytännössä yksi henkilö voi hoitaa useita rooleja samanaikaisesti, ja roolien täyttäminen tapahtuu osallistuvien henkilöiden osaamisen eikä työkuvauksen pohjalta. Roolit liittyvät oleellisesti datatuotteeseen liittyvän palvelun elinkaaren hallintaan suunnittelusta kehitykseen, toteutukseen, ylläpitoon ja lopulta käytöstä poistoon. Roolien ymmärtäminen ja hyödyntäminen myös ohjaa ja yksinkertaistaa organisaatioiden välistä kommunikaatiota, sillä vastaavat roolit tulee olla määriteltyinä kaikkien eri ekosysteemin osapuolten organisaatioissa.

Roolit esitellään tässä luvussa lyhyesti ja roolien vastuita tullaan tarkentamaan tulevissa luvuissa.

Kuva 8 - Hallintamallin roolit

Projektipäällikkö

Datatuotteen kehitysprojektissa on oltava mukana projektipäällikkö. Projektipäällikön tehtävänä on pitää huolta, että yhdessä sovitut asiat tulevat hoidetuksi ja palavereissa on oikeat asiantuntijat paikalla. Projektipäällikkö voi olla tekninen henkilö tai substanssiosaaja. Tyypillisesti organisaatiossa tai kunnassa projektipäällikön tehtävissä korostuu sisäinen viestintätyö ja organisaatiorakenteiden tunteminen.

Kehityspäällikkö / palvelun omistaja

Organisaatiossa tai kunnassa on määrätty vastuuhenkilö, joka omistaa datatuotteen kehityshankkeen. Tyypillisesti tällä henkilöllä on vastuu kehitysbudjetista, mahdollisesta kilpailutuksesta ja sopimusteknisistä asioita. Kunnan tapauksessa kyseessä voi olla substanssiosaaja kuten yksikön-, tila- tai energiapäällikkö. Kehityshankkeen päätyttyä palvelun omistajuus voi siirtyä esimerkiksi kehityspäälliköltä palvelupäällikölle. On tärkeää, että datatuotteella ja sen tarjoamalla palvelulla on nimetty omistaja, jonka tulee huolehtia omistajuuden siirrosta, mikäli hänen roolinsa ja vastuunsa organisaatiossa muuttuu tai mikäli vastuuhenkilö poistuu organisaatiosta.

Tekninen päällikkö

Projektin tekninen päällikkö. Vaatii ylätason tuntemusta tietojärjestelmistä ja tietoa siitä, mitä järjestelmiä sekä rajapintoja ollaan käyttämässä datan lähteinä. Vastaa ja koordinoi tiedon tuottamisen järjestämisestä yhdessä yksittäisten järjestelmäomistajien kanssa tiedon virtautuksesta rajapinnoille. Tekninen päällikkö voi olla sama henkilö kuin projektipäällikkö. Rooliin sopiva osaaminen voi löytyä tietojärjestelmätoimittajalta tai ICT-palveluntarjoajalta.

Yhteentoimivuusasiantuntija

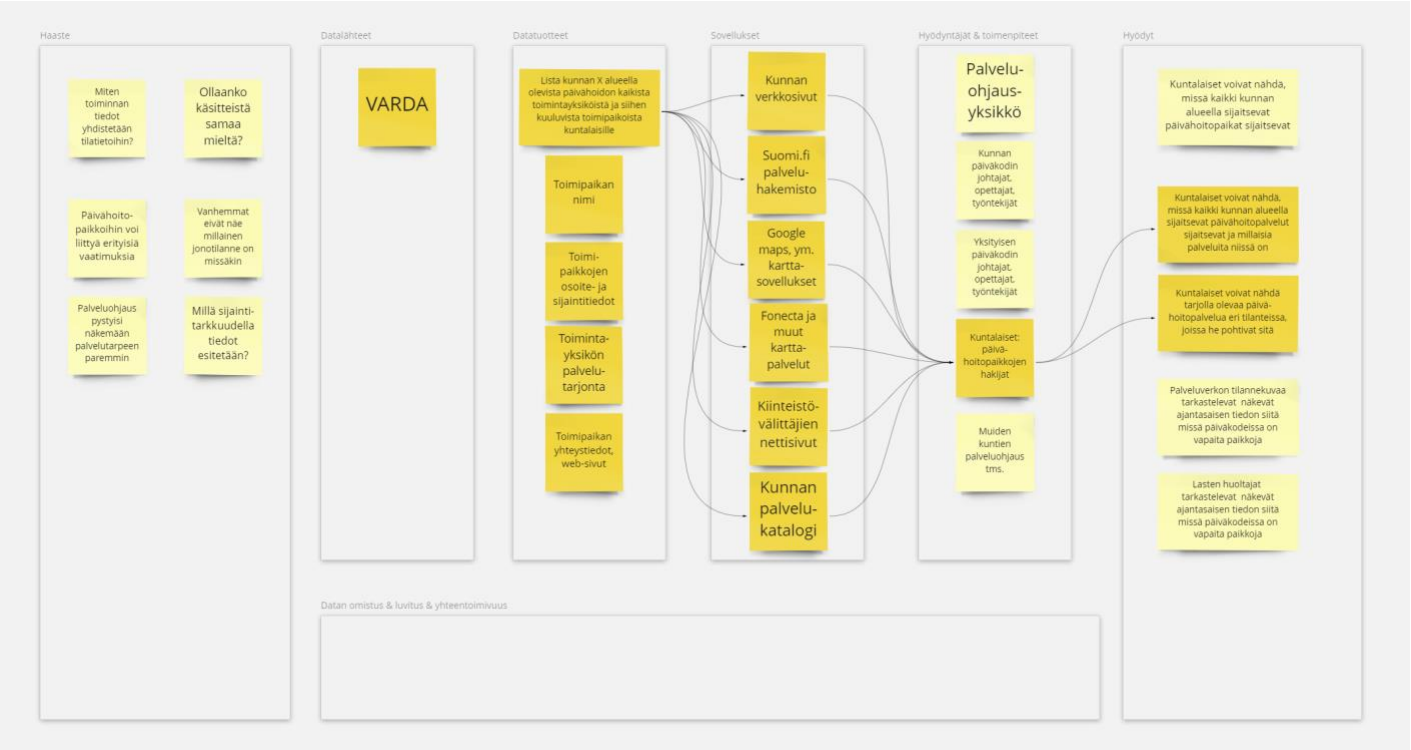
Yhteentoimivuusasiantuntija on tietojärjestelmän (tai -järjestelmien) tietosisällöstä vastaava henkilö. Datatuotteen tuottamishankkeessa henkilö on avainasemassa. Henkilön tulee tietää datatuotteen tietomalli ja sen sisältö kenttä kentältä. Yhteentoimivuusasiantuntija voi työskennellä monenlaisissa rooleissa, kuten esimerkiksi järjestelmänomistajan tai asiantuntijan roolissa. Vastaa yhdessä teknisen päällikön kanssa vaaditun tiedon tuottamisesta rajapinnoille.

3.3.2. Käyttötapauksen muotoilu (design)

Kuten minkä tahansa tuotteen tai hyödykkeen kanssa, datatuotteen määrittely lähtee haasteesta tai tarpeesta, jonka datatuote tulee ratkaisemaan. Palvelumuotoilun menetelmin voidaan pureutua syvemmin tarvelähtöisyyteen ja etsiä vastauksia kysymyksiin kuten:

- Mitä liiketoiminnan tai kunnan haastetta ollaan ratkaisemassa?
- Miten suunniteltu datatuote tulisi ratkaisemaan haasteen (esimerkiksi kustannussäästö, tehokkuuden kasvattaminen automatisoinnin kautta, läpinäkyvyyden lisääminen)?
- Mitä tietolähteitä tarvitaan haasteen ratkaisemiseen?
- Yhdistelemällä em. tietolähteitä, minkälaisia datatuotteita tarvitaan?
- Mistä järjestelmästä ja käyttöliittymästä datatuotteita käytetään (esimerkiksi ERP, Power BI tai vastaava)?
- Kuka datatuotetta tulee käyttämään arjessa ja miten tuote tai palvelu helpottaa kyseisen henkilön operatiivista työtä?
- Kuka omistaa oikeudet tarvittavaan dataan ja miten varmistetaan datan yhteentoimivuus?

Huolellisesti määritelty käyttötapaus ohjaa kehitystyötä alusta loppuun ja pitää tavoitteen kirkkaana. Käyttötapauksen määrittelyn onnistuminen onkin myös tärkein tekijä datatuotteen tuoman arvon toteutumisessa; Mikäli käyttötapaus ei ole selkeänä alusta asti, voidaan palvelun kehityksessä ajautua harhapolulle. Tällöin riskinä on, että edetään tekninen kärki edellä, jolloin tarvelähtöinen ratkaisu hämärtyy ja päädytään toteuttamaan ratkaisu, joka ei vastaa alkuperäiseen tarpeeseen. Tyypillisesti käyttötapaus kirkastuu, kun datatuotteen kehitysprojekti etenee, jolloin kanvas-tyyppinen työpajatyöskentelytapa on osoittautunut oivaksi tavaksi pitää tarvelähtöisyys kehityshankkeen ytimessä.



Kuva 9 - Esimerkki käyttötapauksen muotoilusta kanvas-työpajatyöskentelyn avulla

Yhteenvetona käyttötapauksen määrittelyn tulisi kattaa seuraavat kolme osa-aluetta:

1. Käyttötapauksen tarvelähtöisyyden kuvaus

- ratkaistava haaste
- mittaristo (esimerkiksi rajattu KPI), jolla tuotteen onnistuminen arvioidaan
- arvolupaus

2. Käyttäjän kuvaus

- käyttäjäryhmät (rakennettu segmenttien ja roolien kautta)
- käyttöoikeuksien määrittely käyttäjäryhmille ja rooleille

3. Käyttökontekstin kuvaus

- liiketoiminta / kuntatoimintakonteksti
- tekninen käyttökonteksti

3.3.3. Datan yhteentoimivuus (interoperability)

Kun organisaatio tai kunta optimoi nykyisiä tuotteitaan ja palveluitaan tai innovoi kokonaan uusia, se tehdään usein hyödyntämällä dataa. Jotta dataa voidaan hyödyntää, sitä pitää usein yhdistellä muiden järjestelmien datan kanssa. Mikäli tietojärjestelmät on suunniteltu ja kehitetty yhteensopiviksi, ei yhteentoimimattomuuden ongelmaa ole. Yleisesti tilanne on kuitenkin toinen, ja erilaisiin järjestelmiin tallennettu data ei ole keskenään yhteensopivaa, koska tiedon rakenteet ja kuvaustapa eivät ole vertailukelpoisia. Toisin sanoen, erimuotoisia datoja ei voida yhdistää toisiinsa.

Yhteensopimattomuuden ongelma voidaan ratkaista käyttämällä yhteentoimivuuden menetelmää, joka tunnetaan yleisesti datan muunnostauluna. Sillä tarkoitetaan integroitavien datalähteiden tietokenttien välille muodostettavaa ristiintaulukointia, jossa samaa asiaa tarkoittavien tietokenttien välille tehdään yhteys.

Yleinen tapa saavuttaa yhteensopivuus on sitoutua tiettyyn standardiin. Kun kaikki yhteistyöverkoston toimijat käyttävät yhteistä standardia tiedon kuvailuun ja sen rakenteelliseen jäsentelyyn, tiedon jakaminen ja yhdistely onnistuu helposti yli organisaatiorajojen ja muiden siilojen. Kun halutaan synnyttää laajoja yhteistyöverkostoja, ongelmaksi muodostuu yhteisestä standardista sopiminen. Organisaatioilla on jo valmiiksi käytössään eri tietomalleja, eikä yhteiseen näkemykseen pääseminen ole käytännössä mahdollista, sillä se edellyttäisi osalle toimijoista liian isoa ja työlästä muutosta. Tietomallien standardit toimivat siis siten, että ne vahvistavat ja helpottavat standardia käyttävien organisaatioiden välistä yhteistyötä, mutta toisaalta vaikeuttavat muita standardeja käyttävien organisaatioiden toimintaa kyseisessä verkostossa. Tietomallistandardi toimii siis myös uusien yritysten markkinoille tulon estäjänä.

Platform of Trust -palvelualusta ja universaali tietomalli

Platform of Trust on ottanut yhteentoimivuuden kentässä hyvin erityisen markkina-aseman. Kun muut toimijat yleisesti pyrkivät standardoimaan oman verkostonsa tietomalleja, Platform of Trust tarjoaa menetelmän, jonka avulla eri tietomallistandardit saadaan toimimaan yhteen. Erityistä markkina-asemaamme voidaan avata seuraavalla analogialla: "*Jos tietomallistandardit nähdään erilaisten sähköverkkojen pistorasioiden standerdeina, Platform of Trust tarjoaa eräänlaista matka-adapteria, jonka avulla yhteensopimattomat pistokkeet ja pistorasiat saadaan muunnettua yhteentoimivaan muotoon*", Panu Pitkänen, Platform of Trust -palvelualustan yhteentoimivuuspäällikkö kertoo. Platform of Trust lienee ainoa toimija, joka on keskittynyt nimeomaan tietomallistandardien väliseen yhteentoimivuuteen. Lähtökohtaisesti Platform of Trust tukee siis kaikkia tietomallistandardeja, emmekä ole vielä tähän päivään mennessä kohdanneet niin erikoista tietomallia, etteikö sen ja universaalin tietomallimme välille olisi saatu muodostettua toimivaa muunnostaulua.

3.3.4. Datalähteiden kytkeminen

Datatuotteen tarvitsema tieto voidaan tuottaa ja yhdistellä monin eri tavoin. Tässä luvussa käydään läpi esimerkkiprosessi, miten käyttötapauksesta johdetut datalähteet voidaan yhdistää palvelualustaan ja harmonisoida yhteentoimivaan muotoon. Tiedon virtautuksen arkkitehtuuriin pureudutaan syvemmin luvussa 4.

Datalähteen harmonisointi yhteentoimivaan muotoon sisältää seuraavat vaiheet, jotka kuvataan jokainen omassa aliluvussaan. Jotta prosessi voi edetä seuraavaan vaiheeseen, edellisessä vaiheessa pitää saavuttaa sille asetettu valmiuden määritelmä. Perusperiaate on, että integraatiotyöstä käynnistetään aliprojekti, jolla on suunnitelma ja vastuuhenkilöt.

Integraatioprojektin esitiedot

Integraatioprojektin onnistumisen edellytyksenä on tehokas esitietojen kerääminen yhdessä datatuotteen tuottajan kanssa. Esitietovaatimukset katsotaan riittäviksi, kun niihin on kerätty seuraavat tiedot:

- tieto integraation maksajasta
- käyttötapauksen määrittely

- tieto siitä, käsitelläänkö käyttötapauksessa henkilötietoa vai ei (GDPR-yhteensopivuus)
- projektin vastuuhenkilöiden määrittely
- teknisen starttipalaverin ajankohdan ja vastuuhenkilöiden sopiminen

Tekninen starttipalaveri

Tekniseen starttipalaveriin osallistuvat ainakin seuraavat roolit sekä datatuotteen tuottajan että palvelualustan puolelta:

- tekniset päälliköt
- palvelun omistaja
- yhteentoimivuusasiantuntijat
- projektipäälliköt

Teknisen starttipalaverin lopputuloksena syntyvät seuraavat tuotokset:

- integraatiosuunnitelma
- datalähteen pääsytiedot
- datanäyte, datan kentät ja niiden selitykset

Connector-kehitystyö

Integraation connector-komponentin kehitystyöstä vastaa kehittäjäkumppani, jonka palvelualusta tai datatuotteen tuottaja on tehnyt sopimuksen kyseisestä integraatioprojektista. Työ katsotaan valmiiksi, kun valmis konfiguraatiotiedosto on toimitettu testaukseen.

Muunnostaulu

Muunnostaulu muodostetaan integroitavan datalähteen tietomallin ja palvelualustan ontologian välille. Työ tehdään teknisessä starttipalaverissa selvitettyjen datakenttien ja niiden selitysten perusteella. Työn suorittaa palvelualustan yhteentoimivuusasiantuntija, joka kysyy tarvittaessa lisätietoa asiakkaan yhteentoimivuusasiantuntijalta. Työ katsotaan valmiiksi, kun valmis muunnostaulu on toimitettu kehittäjäkumppanille.

Datatuotteen skeeman luonti

Datatuotteen skeeman luonnista vastaa palvelualusta. Työssä käytetään tyypillisesti Data Product Schema Configurator -työkalua. Työ katsotaan valmiiksi, kun valmis JSON Schema on toimitettu integraation kehittäjäkumppanille.

Testaus & käyttöönotto

Testaus on pitkälti automatisoitu vaihe, joka katsotaan valmiiksi, kun tarvittavat testit on läpäisty. Kun tarvittavat testit on läpäisty, palvelualusta vastaa integraation käyttöönotosta tuotantoympäristössä.

3.3.5. Datatuotteen luonti

Teknisesti Platform of Trust –palvelualustalla on jokaisen datatuotteen taustalla toimiva connector-komponentti. Kun tarvittavaan tietolähteeseen ja tietosisältöön on onnistuneesti kehitetty connector-komponentti, on se otettava käyttöön.

Connector-komponentin käyttöönotto

Saadakseen datatuotteesta toimivan, on connector-komponentti otettava käyttöön eli luotava konfiguraatio. Tästä vaiheesta vastaa connector-komponentin kehitystyötä koordinoinut tekninen päällikkö, joka kerää ja tuottaa seuraavat tiedot:

- konfiguraatiotiedosto
- datatuotteen tuotekoodi
- connectorin instanssi
- käytettävä ympäristö (test, sandbox, production)

Konfiguraatiotiedosto

Jokainen konfiguraatiotiedosto sisältää tarvittavat asiakaskohtaiset tiedot datalähteeseen kytkeytymiseen. Riippuen datalähteen tyypistä, tällaisia tietoja voi olla esimerkiksi rajapinta-avaimet tai SFTP-palvelimen käyttäjätunnukset. Konfiguraatiotiedosto näyttää tältä:

```
{
  "template": "congrid",
  "static": {
    "api": "api-sandbox",
    "version": "v2",
    "resource": "notes",
    "token": "<congrid-api-token>",
    "type": "Note",
    "outputArray": "note",
    "idProperty": "projectId",
    "contextValue": "https://standards.oftrust.net/v2/Context/DataProductOutput/Note/"
  },
  "dynamic": {
    "authConfig.path": "targetObject.idLocal",
    "dataPropertyMappings": "type"
  },
  "request": {
    "ids": [
      "<project-id>"
    ]
  }
}
```

Esimerkkejä konfiguraatiotiedostoista saa Platform of Trust:n sekä kehittäjäkumppaneiden Github versionhallinnasta (https://github.com/PlatformOfTrust/multi-connector/tree/master/config/examples).

Ohessa on esimerkki Congrid-rajapintaan kehitetyn connectorin konfiguraatiotiedostosta, joka poimii työturvallisuushavainto-dokumentteja (https://github.com/PlatformOfTrust/multi-connector/blob/master/config/examples/congrid-note.json). Kyseinen esimerkki on rajapinta-avainta vaille valmis otettavaksi käyttöön. Näin ollen eri asiakkaille saadaan nopeasti luotua Congridin rajapinnasta työturvallisuushavaintoja hakevia datatuotteita heidän omilla rajapinta-avaimillansa.

Datatuotteen tuotekoodi

Datatuotteella on oltava uniikki julkinen tuotekoodi järjestelmässä, joka toimii myös tunnisteena datatuotteen kulutusvaiheessa. Samaa tuotekoodia täytyy käyttää datatuotteen metatietojen luomisvaiheessa.

Connector-komponentin instanssi

Nykyisessä palvelualustan arkkitehtuurissa, connector-komponentit jakautuvat kehittäjäkumppaneiden instansseihin, joiden lähdekoodi säilytetään eri versiohallinnassa. Connectorin käyttöönottovaiheessa on ilmoitettava oikea instanssi, jotta datatuote toimisi oikein. Instanssi määritetään myös datatuotteen metatietojen luomisvaiheessa.

Käytettävä ympäristö

Palvelualustalla on käytössä kolme eri ympäristöä: test, sandbox ja production. Datalähteiden järjestelmät saattavat tarjota erilaisia ympäristöjä ja datatuote voi olla konfiguroitu käyttämään datalähteen sandbox-versiota eikä tuotantoa, joten datatuotteita alustalla luodaan erilaisiin ympäristöihin eri käyttötarkoituksiin. Haluttu palvelualustan ympäristö riippuen käyttötarkoituksesta on ilmoitettava käyttöönottovaiheessa. Kulutusvaiheessa on käytettävä vastaavan ympäristön Products-rajapintaa.

Datatuotteen metatietojen luominen

Datatuotteesta on luotava entiteetti metatietoineen Products-rajapintaan, joka käytännössä tarkoittaa yhtä rajapintakutsua, jonka palvelualustan tekninen päällikkö suorittaa. Oleellisimmat kentät kutsussa olevissa metatiedoissa datatuotteen teknisen toiminnan näkökulmasta ovat tuotekoodi ja connectorin sekä connectorin julkisen avaimen osoite. Tuotekoodin on vastattava samaa arvoa, mikä toimitettiin connectorin käyttöönottovaiheessa. Connectorin sekä julkisen avaimen osoite seuraavat samantyyppistä kaavaa (`https://api-external-<ympäristö>.oftrust.net/<instanssi>/translator/v1/fetch`) ja vaativat muokkausta riippuen, mihin instanssiin ja ympäristöön connector oli otettu käyttöön. Lisää tietoa rajapintakutsuista löytyy oheisesta rajapintakuvauksesta (https://docs.oftrust.net/#Product).

3.3.6. Datatuotteen julkaisu

Kun datatuote on valmis käytettäväksi, tulee se vielä julkaista haluttuihin kanaviin. Julkaisu voidaan jakaa kolmeen julkaisutapaan: julkiseen, kaupalliseen ja sisäiseen julkaisuun. Julkaisun ensimmäinen vaihe onkin julkaisutavan määritteleminen.

Julkaisutavan määrittely pohjautuu datatuotteen käyttötapaukseen ja tuotteen sisältämään tietoon. Julkaisuvalinta voidaan tehdä kolmen vaihtoehdon välillä:

1. Julkinen datatuote - avoimin vaihtoehto, jossa datatuote julkaistaan kaikkien toimijoiden käyttöön.
2. Kaupallinen datatuote - astetta rajoitetumpi vaihtoehto, jossa pääsy datatuotteeseen rajoitetaan maksaville asiakkaille.
3. Julkaisematon/sisäinen datatuote - rajatuin vaihtoehto. Tällaisen datatuotteen olemassaolo ei ole näkyvillä kuin niille toimijoille, joille tuotteen tekijä tuotteesta kertoo.

On tärkeää ymmärtää, että julkaisun avoimuus ei perustu tuotteen käytön avoimuuteen. Tuotteen käyttötarkoitus ja käytön rajaus toteutetaan datatuotteeseen liitettävän sopimuksen kautta. Samasta datatuotteesta voidaankin luoda myös julkinen, kaupallinen ja sisäinen versio yhdenaikaisesti erilaisilla sopimuksilla. Esimerkiksi kunta tai kaupunki voi määrittää datatuotteen olevan julkisesti saatavilla tutkimus- ja koulutuskäyttöön, kaupallisesti saatavilla ympäristöä tukevaan/vihreään liiketoimintaan, ja sisäisesti saatavilla kaikkiin käyttötarkoituksiin.

Julkaisutavan määrittelyä seuraa **julkaisukanavan valinta**. Julkaisukanava voi olla esimerkiksi valtion tietokanta, jokin yksityinen tietokanta tai jonkin palvelualustan tarjoama markkinapaikka. Jokaisen kanavan sisäinen julkaisuprosessi on yksilöllinen, mutta yleisesti ottaen julkaisua varten on vähintään varmistettava, että:

- Julkaisijalla on oikeus julkaista datatuote, eli julkaisija on datatuotteen omistaja tai omaa omistajan luvan julkaista.
- Datatuote ei sisällä arkaluontoista sisältöä, joka estää julkaisun.

- Kanava tarjoaa tarvittavan toiminnallisuuden toteuttaa ja monitoroida data tuotteen sopimuksessa määriteltyjen ehtojen toteutumisen.
- Tuote noudattaa kanavan asettamia ehtoja.

Julkaisukanavasta riippuen tuotteen julkaisu voi olla manuaalinen tai automatisoitu prosessi tai jotain siltä väliltä. Usein julkaisu voi sisältää myös katselmoinnin tai vaatia julkaisukanavan hyväksynnän. Datatuotteen tai palvelun julkaisun yhteydessä voidaan katsoa kehitysvaiheen päättyvän ja palvelun operatiivisen- tai ylläpitovaiheen alkavan. Tässä kohtaa palvelun omistajuus voi siirtyä kehityspäälliköltä esimerkiksi palvelupäällikölle. Ylläpitovaiheessa palveluun on tärkeää määrittää:

- palvelutasosopimus (SLA)
- tekninen tukikanava (asiakaspalvelukeskus puhelimella tai sähköpostilla)
- hälytysjärjestelmä ja mahdolliset automaattiset häiriöviestit
- palauttaminen vikatilanteissa
- palvelun huoltoikkunat sekä mahdollinen päivityssykli

3.3.7. Dataekosysteemin prosessit

Dataekosysteemiin liittyminen ja sertifioinnit

Datan virtautuksen etiketin muodostumiseen tähtäävää työtä on aloitettu muun muassa Gaia-X -järjestön toimesta. Etiketin/konseptin määrittely auttaa toimijoita digitaalisen infrastruktuurin ymmärrettävyyden ja luotettavuuden lisäämisessä. Etiketin kehitys tukeutuu kolmeen johtavaan teemaan:

1. Läpinäkyvyys: luottamusta tukevat rakenteet.
2. Suvereniteetti: käytettävyyttä, ohjattavuutta ja omistusoikeutta tukevat rakenteet.
3. Yhteentoimivuus: valinnanvapauden toteutumista tukevat rakenteet.

Nämä teemat auttavat datan virtautuksen maturiteetin kehityksessä ja tukevat käyttäjiä valistuneiden tarvelähtöisten päätösten tekemisessä luotetun tietoinfrastruktuurin osalta. Etiketin tarkoitus on auttaa valitsemaan käyttötapausten tarvitsema luottamustaso ilman pitkiä ja vaikeita tarkastuksia. Vaatimustenmukaisuus ja kriteeristö auttavat ja helpottavat hallitsemaan ja todentamaan eri toimijoiden rooleja ja kyvykkyyksiä. Esimerkiksi mikä on tarkoituksenmukaisen turvallista, suojattua tai luotettua.

Keskitetty palvelu tiedonvälityksestä sopimiseen

Tiedonsiirron ehdoista sopiminen on ensiarvoisen tärkeää, kun halutaan virtauttaa tietoa siten, että noudatetaan lakia ja kunnioitetaan eri osapuolten oikeuksia suhteessa välitettävään tietoon. Tiedon julkaisijan on voitava asettaa tiedonsiirrolle ehtoja perustuen lakiin ja turvatakseen tiedonsiirtoon liittyvien osapuolten, kuten kaupunkien, itsensä ja muiden ihmisten oikeuksien ja velvollisuuksien toteutuminen. Kun tiedonsiirrossa käsitellään henkilötietoa, pitää siihen saada kaikilta asianomaisilta henkilöiltä. Tällaisia yksityishenkilöiltä kysyttäviä lupia kutsutaan suostumuksiksi. Luvitus ja siihen kuuluvat suostumukset ovat oleellinen osa tiedonsiirron ehdoista, jotka pitää määritellä ennen kuin tietoa voidaan julkaista jaettavaksi. Jotta tieto voi liikkua, pitää ehtojen toteutua. Luvituksen osalta tämä tarkoittaa sitä, että tarvittavat luvat ja suostumukset pitää olla saatuna. Jotta luvitusta ja tiedonsiirron muita ehtoja voidaan hallita käyttäjäystävällisesti, tarvitaan siihen keskitetty palvelu. Keskitetyssä palvelussa voidaan määritellä yhteyksiä henkilöiden roolien, organisaatioiden, tietolähteiden, tuotteistettujen tietovirtojen ja tiedonsiirtoa määrittelevien sopimusten välille.

Esimerkki tiedonvälityksen sopimuksenhallintatyökalusta

Platform of Trust on kehittänyt tiedonvälityksestä sopimiseen graafisen hallintatyökalun, jossa on valmiita sopimuslausekkeita ja –pohjia. Contract Console – työkalu vähentää merkittävästi lakihenkilöiltä vaadittavaa työtä tiedonsiirron ehdoista sopimisessa. Kun lakihenkilö on muodostanut työkaluun valmiita sopimuslausekkeita ja niihin perustuvia sopimuspohjia, voivat organisaation muut henkilöt, kuten tuotepäälliköt, muodostaa niiden pohjalta tuotteistuksia. Tiedonsiirron määrittely onnistuu työkalun avulla siten, että muutoksia rajapintaan ei tarvita, joka taas tuo merkittäviä säästöjä integraatiokustannuksiin. Kun API on kerran integroitu, voidaan työkalusta käsin määritellä joustavasti mitä tietoa siirretään missäkin käyttötapauksessa.

Juridisen sopimisen lisäksi työkalu nopeuttaa ja helpottaa siis merkittävästi tiedonsiirron käytännön järjestelyjä. Jokaista käyttötapausta varten voidaan muodostaa oma datatuotteensa ja siihen liittyvä tiedonvälityksen sopimus. Työkalussa on käytössä rooliperusteinen käyttäjähallinta, mikä helpottaa organisaation työntekijöiden oikeuksien hallintaa uusien työntekijöiden ja muuttuvien vastuiden kohdalla. Contract Console on systeeminen menetelmä tiedonjakamisen määrittelyyn, ja se mahdollistaa kaikkien eri tiedonsiirron kannalta tärkeiden asioiden joustavan yhdistelyn. Työkalulla muodostettavat sopimukset ovat juridisesti päteviä, kun ne allekirjoitetaan osapuolina olevien organisaatioiden nimenkirjoitusoikeudellisten henkilöiden toimesta. Digitaaliseen allekirjoitukseen voidaan käyttää Vastuu Groupin SignSpace-palvelua.

Tiedon virtautuksen tukisovellusten provisiointi ja hyödyntäminen

Tiedon virtautuksen tukisovellukset (myöhemmin datasovellukset) ovat connectoreihin liitettäviä dataa prosessoivia tai transformoivia sovelluksia. Datasovelluksia on erilaisia yksittäisistä ja yksinkertaisen toiminnon suorittavista kontteihin paketoiduista mikropalveluista aina kompleksisiin analytiikkasovelluksiin.

Datasovelluksia voidaan tarjota käyttöön samaan tapaan kuin datatuotteita, eli tuotteistettuina palveluina. Tällöin datasovellukset yleensä rekisteröidään identiteetin- ja luottamuksenhallinnan palveluihin, tai suoraan jonkinlaiseen katalogiin, joka voi olla federoitu useiden toimijoiden tarjoamista palveluista. Katalogi voi sijaita sovelluskaupassa eli alustalla, josta datapalveluja voidaan löytää ja ladata. Esimerkkinä tällaisesta alustasta voi olla ITSM-järjestelmät, joista yleensä löytyy riittävät toiminnallisuudet julkaista, tarjota ja luvittaa ladattavia sovelluksia.

Ekosysteemiarkkitehtuureissa on usein syytä hyödyntää sertifiointia eri toimijoille sekä teknisille ratkaisuille. Tällöin myös datapalvelut tulee rekisteröidä. Datapalvelun tarjoaja tarkistaa ensin, vaatiiko ekosysteemi datasovellusten sertifiointia. Mikäli sertifiointi vaaditaan, lähettää datapalvelun tarjoaja sertifioinnista vastaavalle taholle sertifiointipyynnön, joka sisältää itsekuvauksen julkaistavaksi halutusta datapalvelusta.

Sertifioinnista vastaava taho tarkistaa, että sillä on tarvitsemansa tiedot julkaistavaksi halutusta datapalvelusta, jonka jälkeen se suorittaa sertifiointiprosessin. Jos sertifiointi voidaan tehdä, myöntää sertifioinnista vastaava taho datasovellukselle sertifikaatin, minkä jälkeen datasovelluksen tarjoaja voi julkaista sovelluksen sovelluskauppaan.

Datasovellusten hyödyntäminen tapahtuu hakemalla sopivaa datasovellusta sovelluskaupasta itsekuvausten tarjoamilla tiedoilla. Tämän jälkeen hyödyntäjä pyytää sovellusta käyttöön samaan tapaan kuin datatuotetta. Datasovelluksen toimittaminen riippuu pitkälti sen tarjoamiseen hyödynnettävästä sovelluskauppa-alustasta.

Tarkemmat kuvaukset datasovellusten hyödyntämisestä connectoreissa löytyy luvusta 3.5. ja IDS-referenssiarkkitehtuurimallin luvusta 3.3.3.

Tietoarkkitehtuurin suunnittelu

Tiedon virtautuksen tietoarkkitehtuuri ohjeistaa ekosysteemitoimijoita ja resursseja kuvaavien tietojen hallinnassa ja käytössä sekä tiedon tuotteistamisessa ja tietotuotteiden hallinnassa. Lisäksi se tukee semanttista ja rakenteellista yhteentoimivuutta tarjoamalla ohjeistusta olemassa olevien käsitteistöjen sekä tietomallien hyödyntämiseen, uusien käsitteistöjen ja tietomallien luomiseen ja näiden hallintaan.

Dataekosysteemin tietomallia tarkastellaan tässä pääosin IDS:n jaottelun mukaan, mikä mahdollistaa datatuotteen tai muun digitaalisen resurssin, kuten datasovelluksen kuvailemisen, julkaisemisen ja löytämisen. On syytä huomata, että ekosysteemissä voi olla myös muita tuotteistamattomia resursseja (mukaan lukien dataa), joita ei kuvata ekosysteemin federoidussa katalogissa vaan yksityisesti tietyn tarjoajan ja hyödyntäjän välillä

3.3.8. Itsekuvaus ja sanastot

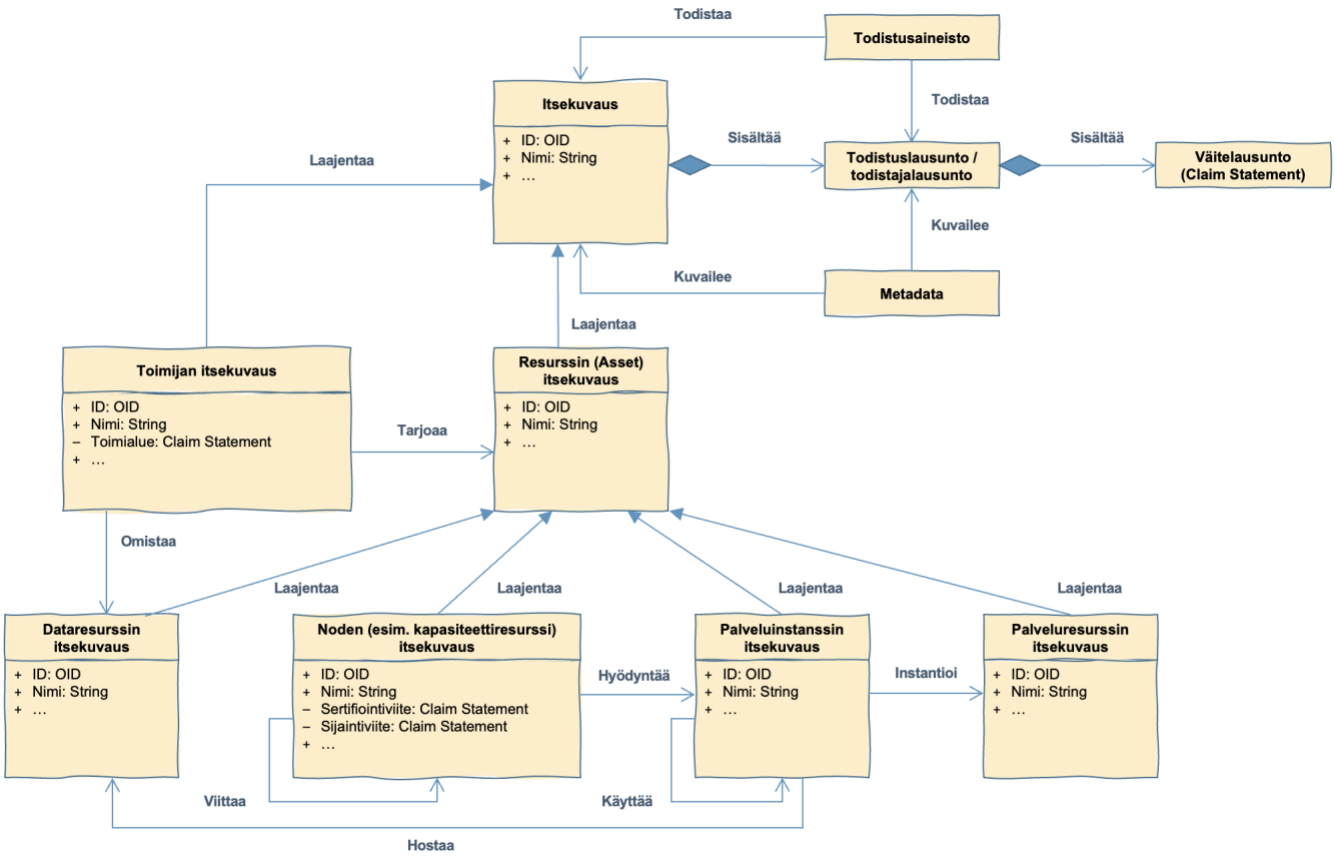
Dataekosysteemin keskeinen hallittava tieto on sen resurssien ja toimijoiden itsekuvaukset. Itsekuvaukset tarkoittavat resurssia tai toimijaa kuvailevia metatietoja, joiden tarkoitus on antaa taustatietoja itse kohteesta. Itsekuvaukset mahdollistavat ekosysteemitoimijoiden ja -resurssien, kuten datatuotteiden tuomisen muiden toimijoiden nähtäville siten, että niiden hyödynnettävyydestä ja luotettavuudesta voidaan tehdä päätöksiä. Itsekuvaukset erilaisten verifiointipalvelujen tukemana mahdollistavat ekosysteemitoimijoita tekemään päätöksiä resurssien tarjoamisesta ja hyödyntämisestä. Itsekuvauksia voidaan hyödyntää muun muassa

- Resurssien, kuten datatuotteiden löytämiseen.
- Työkaluavusteiseen datatuotteiden ja palveluinstanssien arviointiin, valintaan ja integrointiin.
- Noudattamisen hallinnan, jatkuvan validoinnin ja luottamuksen valvonnan toteuttamiseen yhdessä käyttöehtojen ja -sopimusten kanssa.
- Resursseja ja käyttäjiä koskevien sopimusten neuvottelemiseen.

Itsekuvausten olisi hyvä olla saatavilla standardoidusti ja koneluettavassa muodossa, jotta niitä kyetään hyödyntämään eri prosesseissa automatisoidusti.

Gaia-X kuvaa lisää toiminnallisia vaatimuksia itsekuvausta koskien Gaia-X:n teknisen arkkitehtuurin luvussa 2.4.

Itsekuvauksia tulee kyetä linkittämään keskenään ja muihin tietoihin, kuten todistuksiin tai luottamusverkkoihin, jotta ekosysteemin resurssien käyttöä kyetään paremmin hallitsemaan ja käyttösopimuksien syntymistä automatisoimaan. Alla on kuvattu Gaia-X:n näkökulma itsekuvauksien välisiin suhteisiin ja vähimmäisvaatimukset itsekuvausten tietosisällöille, joita suositellaan vähimmäiskuvauksiksi myös kuntien dataekosysteemityössä.



Kuva 10 - Esimerkki itsekuvauksen rakenteellisesta hierarkiamallista Gaia-X:n mallia mukaille

Jotta itsekuvauksia, mutta myös jaettavaa dataa, kyettäisiin hyödyntämään mahdollisimman laajasti, suositellaan ekosysteemejä suunnitellessa kiinnittämään huomiota erityisesti yhteisiin standardiformaatteihin ja ontologioihin, joita hyödynnetään muun muassa itsekuvausten ja tiedonsiirron yhdenmukaistamiseen. Yhteiset formaatit ja ontologiset käsitteistöt auttavat ekosysteemitoimijoita ymmärtämään ekosysteemikomponenttien sisällön ja tarkoituksen.

Semanttisen yhteentoimivuuden varmistamiseksi olisikin hyvä luoda itsekuvauksille myös deklaratiivisen tason tasolla käsitemalli, jossa määritellään tietosisältöjen objektit ja niiden ominaisuudet. Näiden kuvaukset voidaan linkittää olemassa oleviin yleisesti käytettyihin käsitteistöihin tai sanastoskeemoihin. Tämän mahdollistamiseksi itsekuvausten käsite- ja tietomalleissa suositellaan hyödyntämään linkitetyn datan standardeja, tärkeimpinä

- linkitetyn datan RDF-kuvauskieltä (JSON-LD, RDF/XML tai TURTLE -serialisoituna)
- OWL-ontologiakieltä
- SPARQL-kyselykieltä
- SHACL-graafikuvauskieltä
- soveltuvia sanastoja (RDF Schema, Dublin Core, SKOS, jne.) käyttötapauksesta riippuen

Edellä mainituilla standardeilla kuvatut käsite- ja tietomallit on mahdollista konvertoida muihin standardeihin muun muassa Yhteentoimivuusalustalla, esimerkiksi jos datatuotetta halutaan hyödyntää useissa ekosysteemeissä, joissa käytetään eri standardeja. Lähtökohtaisesti semanttisesti yhteentoimivaan rakenteiseen tiedonsiirtoon sekä validointiin toimijoiden välillä tulisi käyttää edellä mainittuja standardeja.

3.3.9. Datatuote ja datatuoteaihio

Datatuote on digitaalinen resurssi, jolla on yksilöivä tunniste ja itsekuvaus, ja jota voidaan jakaa muille dataekosysteemin toimijoille. Datatuotteita voidaan kuitenkin hyödyntää tiedon jakamisessa myös ekosysteemien ulkopuolella. Ne tuodaan hyödyntäjän saataville connectoripalveluiden, federoitujen katalogien sekä valittäjäpalveluiden avulla. Myös muut digitaaliset resurssit, kuten ohjelmistot ja infraresurssit (vrt. Gaia-X:n nodet) voidaan jakaa samoilla keinoilla.

Datatuote koostuu yhdestä tai useammasta datasetistä sekä näiden metatiedoista, jotka tuodaan datan hyödyntäjien saataville. Se voi verkkoresurssiparadigmaan perustuen koostua sisällöltään muun muassa dokumenteista, kuvatiedostoista, sensoridatasta, arkistoista, media-streameista yms. digitaalisista arvoa tuottavista resursseista (kts. https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm#tab_5_1).

IDS on määrittänyt digitaalisille resursseille kuusi näkökulmaa, joiden avulla datatuotetta voidaan tarkastella. Nämä näkökulmat avaavat, minkälaisia tietoja datatuotteesta tulisi kerätä, jotta niistä voidaan muodostaa datatuotteelle itsekuvaus.

Kuva 11 - IDS:n näkökulmat datan tuotteistukseen

Tässä mallissa sisältö ja sen esitysmuoto voidaan purkaa tai kääntää konseptiksi, joka viittaa sisältöä koskevaan kontekstiin, joka tekee sisällöstä relevantin ja arvokkaan mahdolliselle datan hyödyntäjälle. Tiedonsiirto puolestaan tarkentaa, kuinka dataa vaihdetaan, kun taas hyödykeosio viittaa datan vaihdon perusteisiin. Luottamusyhteisö on puhtaasti ekosysteemin tai muun luottamusverkon rakenteeseen ja datan vaihtamiseen liittyvää tietoa, joka kertoo ketkä toimijat ja mitkä connectorit voivat käsitellä dataa. Gaia-X suosittelee datatuotteen itsekuvaukseen sisällytettävän ainakin tiedot datan omistajasta, käyttöehdoista, tarkemmat tiedot datan alkuperästä, teknisistä määrityksistä ja sisältöön liittyvät kuvaukset. IDS:n esittämä malli tarkentaakin näitä kuvauksia ja soveltuu hyödynnettäväksi myös Gaia-X:n periaattein rakennetussa dataekosysteemissä. Datatuotteen kuvaus käy myös Julkishallinnon API-periaatteiden mukaiseksi rajapintakuvaukseksi. Ajantasaisin versio API-periaatteista tulee kuitenkin tarkistaa mahdollisten muutosten vuoksi datatuotteen suunnittelun aikana.

Datatuotteen määrittelyprosessia kuvataan tarkemmin luvuissa 3.1. ja 4. Tässä luvussa esitellään ylätasolla jokainen edellä esitetyistä datatuotteen näkökulmista. Tarkemmat tiedot näkökulmista löytyy IDS-referenssiarkkitehtuurimallin luvusta 3.4.3.

Datatuotteen sisältö

Datatuotteen sisältö (yllä esitetyssä kaaviossa resurssi) huomioi:

- dokumentin sisältötyypin (esimerkiksi video- tai dokumenttitiedosto)
- sisällön esitystavan, kuten formaatin ja tietorakenteen (esimerkiksi MIME-tyyppi, skeema ja sen esittävä skeemadokumentti, paketointitapa)
- mahdollisen sisältöesimerkin
- instanssitiedon, eli dokumentin koosta ja luontiajankohdasta kertovat tiedot (esimerkiksi versiotieto, luontipäivä ja tiedostojen koot)

Datatuotteen sisältöä suunniteltaessa suositellaan yhteentoimivuuden varmistamiseksi ja tiedon jakamisen sujuvuuden vuoksi harkitsemaan, tulisiko ekosysteemissä hyödyntää yhteisiä käsitteistöjä, ontologioita ja mahdollisesti tietomalleja. Esimerkkejä yhteisistä ontologioista ovat muun muassa opetus- ja koulutussanasto (OKSA), taloustieto-ontologia Universal Business Language (UBL/BusDox) ja EU:n yhteentoimivuuden sanasto. Esimerkkejä yhteisistä tietomalleista ovat puolestaan Julkishallinnon tietokomponenttikirjasto ja EU:n CoreVoc/Isa2Core-tietokomponenttikirjasto.

Yhteiset ontologiat ja tietomallit voidaan kuvata ja ottaa käyttöön yhteisille alustoille, esimerkiksi yhteentoimivuusalustan sanastot.suomi.fi ja tietomallit.suomi.fi -työkaluilla, joista ne ovat jaettavissa kaikille ekosysteemitoimijoille. Sanastojen, tietomallien ja datan määrittämistä ja linkittämistä kuvataan tarkemmin seuraavissa luvuissa.

Datatuotteen sisältöelementtejä ja käsitemalleja on kuvattu tarkemmin IDS referenssiarkkitehtuurimallin luvussa 3.4.3.5.

Datatuotteen konteksti

Datatuotteen konteksti kertoo tarkemmin datatuotteen sisällön ajallisesta ja paikkaan liittyvästä kontekstista, kuten:

- Ajanjakso, jonka datatuotteen sisältö kattaa.
- Aika ja paikka, jossa sisältö on kerätty.
- Reaalimaailman entiteetit, johon datatuote liittyy ja joita voidaan käyttää hakusanoina (esimerkiksi maakoodi: ISO 3166 FI ja aikakausi: keskiaika).

Näiden tietojen esittäminen auttaa tiedon etsijää arvioimaan, onko datatuote hänelle relevantti ja vastaako se hänen tietotarpeitaan.

Datatuotteen kontekstielementtejä ja käsitemalleja on kuvattu tarkemmin IDS referenssiarkkitehtuurimallin luvussa 3.4.3.6.

Datatuotteen konsepti

Datatuotteen kontekstitiedot avaavat tarkemmin datatuotteen tarkoitusta siihen liitettävällä annotaatiolla. Konsepti koostaa keskeisiä “teemoja” ja avainsanoja datatuotteen sisällöstä, kontekstista sekä kommunikaatiotiedosta datatuotteen hakua helpottamaan, kuten:

- Havaittava kohde, johon data liittyy.
- Minkälaisessa kohteessa havainto on tehty.
- Jonkin parametrin, esim. ajanhetken merkitys (esimerkiksi aineiston aikasarjan aloitus tai lopetus).

Näitä tietoja koostetaan vapaasti valittaviksi avainsanoiksi ja termiluokitteluiksi (esimerkiksi literaalit ja sanastoviittaukset). Jos termiluokitteluja ei haluta tai voida käyttää, voidaan antaa myös tyyppiluokittelu esimerkiksi viittauksena johonkin tietoformaattiin kuten asiakirjaformaatteihin.

Datatuotteen konseptielementtejä ja käsitemalleja on kuvattu tarkemmin IDS referenssiarkkitehtuurimallin luvussa 3.4.3.7.

Datatuotteen kommunikaatiotieto

Datatuotteen kommunikaatiotieto kattaa yksityiskohdat datatuotteen tiedonsiirtotavoista, kuten:

- Datatuotteen hakijalta vaadittu syöte.
- Tuetut kommunikaatioprotokollat.
- Miltä validi sanoma näyttää?
- Mihin endpointin osoitteeseen kyselyt tulee lähettää?

Kommunikaatiotieto on jaettu kolmeen luokkaan: operaatiotieto, viestitieto ja endpointin tieto. Operaatiotieto kuvaa datatuotteiden connectorirajapintojen mahdollistamat operaatiot, joita voidaan koostaa palvelurajapinnoiksi. Lisäksi se tarkentaa, mitä parametrejä operaatioihin tulee syöttää ja operaatiotyypit. Yleisesti ottaen suositellaan hyödyntämään moderneja rajapinta-arkkitehtuureja, kuten REST:iä, jotka mahdollistavat laajan joukon operaatioita ja ovat yleisesti hyödynnettyjä.

Sanoma kuvaa siirrettävän sisältöpaketin metadatan, jonka avulla viesti voidaan tarvittaessa jäljittää (esimerkiksi osoitteet ja transaktio-ID). Lisäksi sen tulisi esittää, missä yhteydessä dataa siirretään (esimerkiksi sisältötyyppi ja käyttösopimus). Metadatan sisällyttäminen joko komplementaarisena datana itse sisällön siirtoon tai erillisenä metadataosiona riippuu pitkälti tiedon siirtämisessä käytetystä tekniikasta. Tiedon virtauttamiseen liittyy erilaisia resurssin pyytämiseen, vastaukseen ja ilmoituksiin liittyviä sanomia.

Endpoint-tieto kertoo vielä rajapintaosoitteen resurssi-endpointeille sekä palvelu-endpointeille, sekä käytettävät yhteysprotokollat (esimerkiksi HTTPS) ja osoiteskeeman (esimerkiksi HTTP(S) URI).

Datatuotteen kommunikaatiotiedon elementtejä ja käsitemalleja on kuvattu tarkemmin IDS referenssiarkkitehtuurimallin luvussa 3.4.3.8.

Datatuotteen hyödyketieto

Datatuotteen hyödyketieto tarkentaa, miten laadukasta datan voi olettaa olevan ja mitä rajoituksia sekä mahdollisia maksuja datatuotteen hyödyntämiseen liittyy.

Hyödyketieto vastaa pääosin seuraaviin kysymyksiin datatuotteesta:

- Onko datatuotteen data luotettavasta lähteestä?
- Mikä on tarjotun datatuotteen laatutaso?
- Mitä rajoituksia datatuotteen hyödyntämiseen liittyy?
- Kuinka paljon datatuotteen hyödyntäminen maksaa?

Datan alkuperä on erityisen tärkeä tieto, joka kertoo tarkemmin, miten data on syntynyt ja miten sitä on muokattu (ja milloin). Alkuperätiedon tulisi kertoa muun muassa onko data syntynyt jossakin hankkeessa ja kuka on rahoittanut datan tuottamisen. Alkuperätietoja tarvitaankin arvioimaan, onko tieto sopivanlaista hyödyntämiseen liittyvään käyttötapaukseen.

Laatutieto tarkentaa kuinka kokonaista, validia, tarkkaa, ajantasaista, saatavissa olevaa ynnä muita laatuun liittyviä tietoja sekä kuinka laatua on mitattu, kuka laatua on tarkastanut ja esimerkiksi muiden kuin tarjoajan arvioita datatuotteen laadusta.

Käyttöehtotiedot tarkentavat, mitä sääntöjä, oikeuksia, lupia ja kieltoja datatuotteeseen liittyy, ketkä ovat datan siirron osapuolet, mitä he voivat tehdä ja mistä he ovat vastuussa ynnä muita kohteen käyttöön liittyviä tietoja.

Mikäli datatuotteen hyödyntäminen on hinnoiteltua, tulee siitä ilmoittaa hintatiedossa. Samalla tulee kertoa, miten datatuote on hinnoiteltu.

Datatuotteen hyödyketietoja ja käsitemalleja on kuvattu tarkemmin IDS referenssiarkkitehtuurimallin luvussa 3.4.3.9.

Datatuotteen luottamusyhteisötieto

Datatuotteen luottamusyhteisötieto on data-avaruuspohjaisten ekosysteemien avaintietojoukko, jolla mahdollistetaan datan vaihtaminen ja jakaminen luotettavasti sekä turvallisesti, mutta vaarantamatta datan omistajan sekä tarjoajan omistajuutta dataan. Se auttaa vastaamaan seuraaviin kysymyksiin:

- Mitä tiedetään aiotun datatransaktion osapuolesta?
- Onko osapuolen lähettävä tai vastaanottava connectori teknisesti turvallinen ja luotettava?
- Voidaanko yllä esitettyjen kysymysten vastaukset todistaa esimerkiksi sertifioinneilla?
- Mitä rajoituksia transaktiossa hankittavan datan käyttöön kohdistuu?

Kommunikaatiotiedon osapuolitieto kuvaa viittauksin datatuotteen osapuolet, jotka voivat olla esimerkiksi yksittäisiä henkilöitä, organisaatioita tai niiden yksiköitä. Osapuolet vaihtavat tietoa sille tarkoitetun ohjelmistokomponentin eli connectorin avulla. Sekä osapuolia että connectoreita voidaan sertifioida ja sertifiointitieto liittää käytönhallintaan käyttösopimuksen muodossa. Osapuoli-, connectori-, sertifiointi- ja käyttösopimustietojen määrittäminen mahdollistaa myös datatuotteiden näkyvyyden hallinnan federoidussa katalogissa vain aiotuille osapuolille.

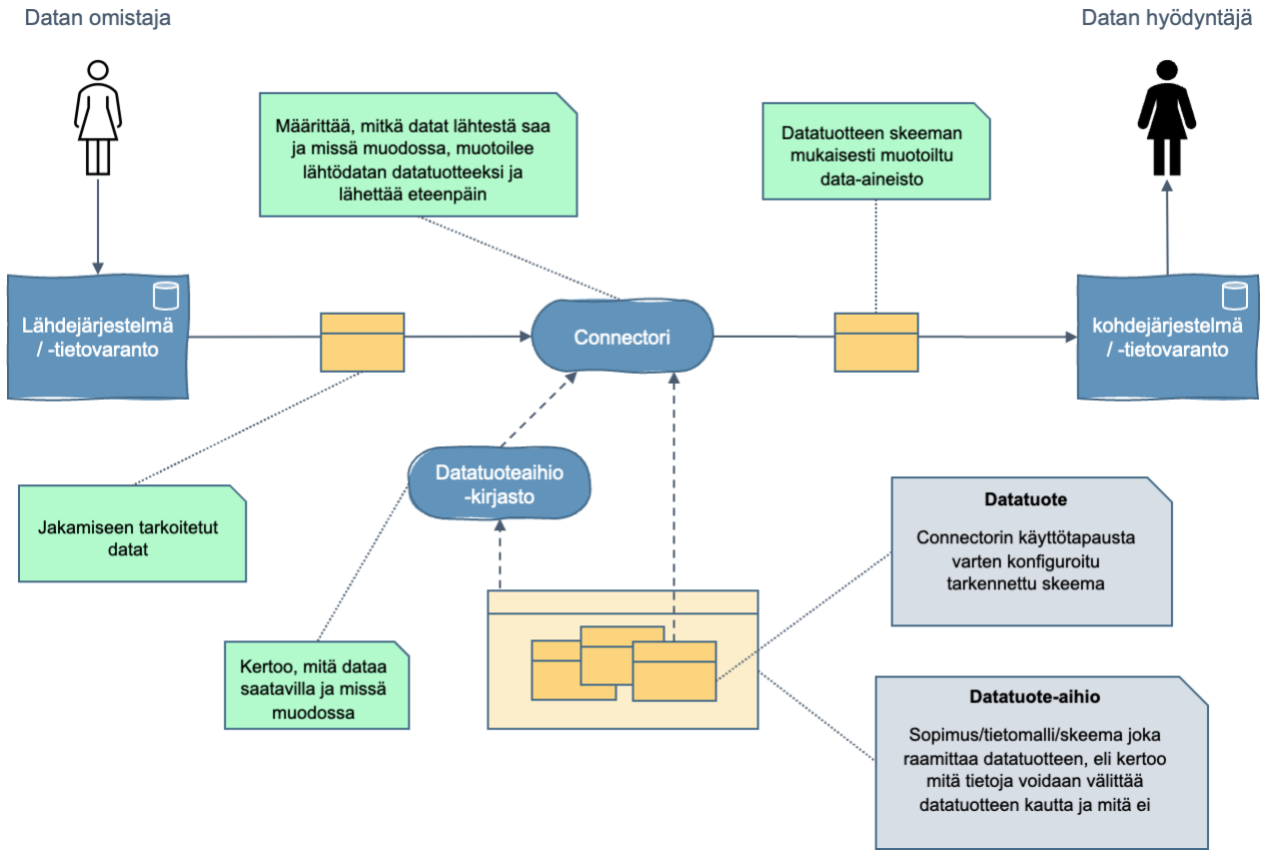
Datatuotteen luottamusyhteisötiedon elementtejä ja käsitemalleja on kuvattu tarkemmin IDS referenssiarkkitehtuurimallin luvussa 3.4.3.10

Datatuoteaihio

Datatuoteaihio on datatuotetta koskeva sopimus, tietomalli tai skeema, joka kertoo, millä tiedoilla datatuotetta voidaan laajentaa. Datatuoteaihioita voidaan hyödyntää tapauksissa, joissa datatuotteen käyttötapauksen tietotarpeen odotetaan laajenevan, jotain tarvittua tietoa ei ole datatuotteen luontihetkellä saatavilla tai samasta tiedosta halutaan luoda hieman erilaisia datatuotteita eri hyödyntäjiä varten.

Datatuoteaihiolla ei ole omaa connectoria, mutta se voidaan tuoda federoituun katalogiin tiedoksi siitä, mitä tietoja voisi olla saatavilla. Vaihtoehtoinen tapa palvella eri toimijoiden tietotarpeita samalla datatuotteella on provisioida datatuotteen attribuutit hyödyntäjäkohtaisesti, eli tarjota eri toimijoille mahdollisuus nähdä ja hakea vain se osa datatuotteesta, jonka he tarvitsevat.

Alla esitetty kaavio kuvaa tarkemmin datatuotteen ja datatuoteaihion väliset suhteet sekä roolin tiedonsiirrossa.



Kuva 12 - Datatuoteaihion hyödyntäminen

3.3.10. Miten otan tietomallit käyttöön – esimerkki DigiPAVen tietomallit

Ennen siirtymistä potentiaalisten implementaatiopolkujen kuvaamiseen, on olennaista muistuttaa tietomallien roolista hankkeessa sekä niiden hankkeen puitteissa saavutetusta valmiusasteesta. DigiPAVen tietomallien premissinä on ollut mahdollistaa hankkeen toimintamallin sekä viitearkkitehtuurin asettamien tavoitteiden täyttäminen. Mallien toteutuksessa on pyritty huomioimaan hankkeen sisäisten tarpeiden lisäksi aiemmin tuotetut määrittelyt ja hyödyntämään kansallista yhteentoimivuusmenetelmää, ylikansallista EIF-kehikkoa (European Interoperability Framework) ja kansainvälisiä semanttisesti yhteentoimivan tiedon standardeja.

Kuten muidenkin osaprojektien tuotosten kohdalla, myös tietomallien varsinainen validointi tapahtuu niiden käyttöönoton myötä. Tämä tarkoittaa väistämättä sitä, että mallien sisältöä jouduttaneen laajentamaan tai korjaamaan todellisen toimintaympäristön tiedoilla tapahtuvassa pilotoinnissa havaittavien poikkeusten ja rajatapausten myötä. Mallien määrittelyssä tämä on pyritty huomioimaan tekemällä niistä helposti refaktoroitavia.

Otettaessa malleja käyttöön, täytyy niitä implementoivan organisaation tiedostaa, että käyttöönottoa ei tule toteuttaa perinteisen vesiputousmallin mukaisesti, eikä mallien skeemoihin tule suhtautua staattisina monoliitteina. Mallien käyttöönoton vanavedessä organisaation täytyy väistämättä sitoutua myös niiden muutostenhallintaan sekä kansalliseen yhteentoimivuustyöhön. Yhteentoimivuusalustalla on käynnistynyt tämän hankkeen loppupuolella merkittävä kehitystyö sekä sen käytettävyyden että teknisen pohjan ja mallintamisparadigman selkiyttämiseksi. Vuorovaikutuksessa sidosryhmien kanssa tehtävä kehitystyö tulee kasvattamaan alustan käyttöä ja edesauttamaan sekä harmonisoitujen

julkisen hallinnon että tietoaluekohtaisten rakenteiden päivittämistä ja validointia. Tästä kehityksestä tietokomponenttikirjastoa implementoivan organisaation on hyvä olla perillä.

Hankkeessa on luotu yhteentoimivuusmenetelmän pohjalta Yhteentoimivuusalustalle kahdenlaisia malleja: tietokomponenttikirjasto (myöhemmin TKK) sekä siitä johdettuja soveltamisprofiileja (myöhemmin SP). Alustan ja yhteentoimivuusmenetelmän mukaisesti mallien skeema on kuvattu linkitettynä datana RDF-kielellä. Käyttöönoton näkökulmasta keskeinen tiedostettava seikka on se, että tietokomponenttikirjasto kuvaa palveluverkkosuunnittelun *ydintietoa* loogisena mallina, ja soveltamisprofiilit ovat siitä johdettuja tiettyjä käyttötapauksia palvelevia näkymiä. Rinnastuksena relaatiomalliin SP:t ovat analogisia laskennassa toteutuneisiin näkymiin (materialized views). TKK itsessään ei sanele ydintietojen hallintaan käytettävää fyysisen tason toteutusta, ainoana vaatimuksena on siinä kuvattujen loogisten rakenteiden säilyminen toteutuksessa, jotta SP:issa kuvatut sanomat voidaan toteuttaa.

Käyttöönottoon on lukuisia polkuja, joiden välillä tehtävä valinta riippuu käyttöön ottavalla organisaatiolla ensisijaisesti seuraavista tekijöistä:

- Ydintietovarantojen tämänhetkinen (sisäinen) yhteentoimivuuden taso.
- Sisäisen ja ulkoisten toimijoiden kanssa tehtävän tiedonvaihdon yhteentoimivuudelle asetettava tavoitetaso.

Molemmat tekijät ovat riippuvaisia tietojärjestelmien ja niiden rajapintojen, organisaation olennaisten osien toimintamallien, strategian ja toimintakulttuurin niin kutsutusta teknisestä velasta (aiemmasta sitoutumisesta muutosjoustamattomiin ratkaisuihin), siiloutumisesta johtuvasta potentiaalisesta tilannekuvan hajanaisuudesta, sekä päällekkäisestä tiedontuotannosta että toiminnasta.

TKK:ta voidaan lähtötilanteesta ja asetettavasta tavoitetilasta riippuen käyttää eri tavoin:

1) **Rajattu toteutus palveluverkkosuunnittelun tukemista varten:** käytössä olevista tietojärjestelmistä rakennetaan rajapintojen kautta ETL (Extract, Transform, Load) -prosessilla putki, jolla palveluverkkosuunnittelun tietoalueen tarvitsemat tiedot saadaan käyttöön TKK:n kuvaamaan loogiseen muotoon. Tässä ratkaisussa TKK toimii itse johdettuna näkymänä suhteessa ulkoisiin ydintietovarantoihin. Potentiaaliset kehitysresurssit kohdistuvat automatisoituun ETL-muunnokseen sekä TKK:n näkökulmasta puuttuvien ydintietojen täydentävien tuotantoprosessien pystyttämiseen. Uuden tuotannon osalta tulee erikseen tehdä päätös, ottaako TKK ydintietovarannon roolin kyseisten tietojen osalta, vai hallinnoidaanko niitä ulkoisesti ja tuodaan TKK:n saataville muiden tietojen tapaan rajapinnan kautta.
2) **Kattava toteutus koko organisaation ydintiedon yhteentoimivuuden kehittämiseen:** organisaatio tunnistaa vähintäänkin keskeisten TKK:ssa kuvattujen ydintietojen tietueiden (toimija, palvelu, sijainti jne.) osalta käytössään olevien järjestelmien skeemojen vastaavat rakenteet ja pyrkii normalisoimaan niitä siten, että DigiPAVessa kuvatut palveluverkkosuunnittelun tietotarpeet ovat johdettavissa ydintiedon loogisesta rakenteesta. Tässä dokumentissa annetaan vahva suositus siitä, että tällaista ydintietoja koskevaa harmonisointia ei tehdä itsenäisesti jokaisessa organisaatiossa, vaan ohjatusti kansallisen yhteentoimivuustyön tasolla. Tällöin harmonisoinnissa toteutuvat todennäköisesti sekä standardien että parhaiden käytänteiden (engl. *best practices*) mukaiset ratkaisut.

Fyysisellä tasolla TKK:n käyttöönotto organisaatiossa voidaan myös toteuttaa useammalla tavalla:

1) **Natiivi RDF-graafitietokanta:** tässä ratkaisussa TKK:n kuvaamaa tietoa hallitaan suoraan RDF-muodossa, ja sitä käytetään ensisijaisesti SPARQL-endpointin kyselyiden kautta. Etuna ratkaisussa on se, että tieto säilyy linkitettynä datana niin sanomissa kuin itse kannassakin, ja SPARQL tarjoaa joustavan ja ilmaisuvoimaisen rajapinnan tiedon semanttiseen louhintaan.

2) **RDF-rajapinta relaatiokannalla:** tässä ratkaisussa fokus on semanttisen rajapinnan määrittämisessä siten, että tietoa saadaan ulos ensisijaisesti valmiiksi määriteltyjen sanomien (SP:t) mukaisessa muodossa linkitettynä datana (esim. JSON-LD). Tieto on itse kannassa tallennettuna TKK:n loogista rakennetta vastaavaan relaatioskeemaan. Tämä toteutus vaatii sekä itse kannan skeeman että siitä johdettujen linkitetyn datan kuvauksien tuottavien näkymien ylläpitoa.

Molemmissa ratkaisuissa olennaisessa osassa on osapuolten kommunikaatiossa siirtyvän tiedon semanttisen rakenteen yhdenmukaisuus, sillä lähtökohtaisesti yhteentoimivuudessa on kyse juuri tästä. Se, miten tietoa hallitaan teknisesti tietojärjestelmissä, on epäolennaista suhteessa siihen keskeiseen päämäärään, että tieto saadaan käyttöön semanttisesti yhteentoimivassa muodossa.

Ratkaisuissa olennaisen eron muodostaa se seikka, että vaihtoehdossa 1 tiedon rakenne (RDF) ei muutu kantaan tallennettaessa, sen sisällöllinen validointi on kuvattu samalla kielellä (RDF-pohjainen SHACL), ja kaikki tiedon manipulointi tapahtuu SPARQL-kyselyiden kautta (joissa transaktiot on sidottu SHACL-rajoitteisiin). Vaihtoehdossa 2 joudutaan ylläpitämään kahden eri paradigman mukaista validointia (SQL DDL-skeemaa ja SHACL-rajoitteita). Keskeinen haaste tässä tapauksessa on varmistaa, että SQL-kyselyllä kantaan kirjoitettaessa ei sallita sisältöä, joka olisi SHACL-validoinnin kannalta virheellistä.

3.3.11. Käsitemallinnus

Tietomallintaminen perustuu käsitteistön varaan rakennetulle käsitemallille. Tällä ei tarkoiteta muun muassa UML-mallintamisen piirissä tyypillistä vapaamuotoista niin kutsuttua tussitauluesitystä mallinnettavia luokkia vastaavista käsitteistä ja niiden suhteista vaan jo lähtökohtaisesti koneluettavaa mallia niistä käsitteistä, joiden varassa hankkeen määrittely lepää.

Käsitemallin tehtävänä on kuvata yksiselitteisesti, minkä käsitteiden kautta hahmotetaan olennaisia osia reaalimaailmasta ja miten kyseiset käsitteet kytkeytyvät toisiinsa. Ensisijaisesti tällä pyritään siihen, että kaikki toimijat jakavat täsmälleen saman käsityksen niistä asioista, joita on määritelty. Käsitteisiin viitataan termien avulla, ja käsitteiden merkitystä avataan määritelmällä, joka on ihmisluettava kuvaus käsitteen tarkoituksesta. Käsitteitä voidaan kuvata myös käsitepiirteiden eli niin kutsuttujen *fasettien* kautta, jotka kuvaavat, erilaisia tapoja luokitella käsitettä vastaavia kohteita. Esimerkiksi termillä *palvelun osallinen* voidaan kuvata niitä toimijoita, jotka ovat osallisena jossain palvelussa, ja nämä voidaan jakaa esimerkiksi osallisen roolin (asiakas, palvelunantaja), osallisen tyypin (luonnollinen henkilö, oikeushenkilö) tai monen muun eri fasetin mukaan. Käsitteiden välille muodostuu myös erityyppisiä suhteita, esimerkiksi *oppijan* hierarkkinen yläluokka on *asiakas*, sillä kaikki oppijat ovat (koulutuspalvelun) asiakkaita.

Tässä vaiheessa on olennaista pohjustaa käsitemalleja (erityisesti linkitetyn datan näkökulmasta) täsmentämällä, että koneluettavat käsitemallit ovat *ontologioita,* siinä missä koneluettavat loogiset mallitkin. Ontologialla tarkoitetaan erityisesti linkitetyn datan kontekstissa formalisoitua kuvausta jaetusta käsitteistöstä - käytännössä koneluettavaa tietorakennetta, josta ontologian pohjana käytetyn kielen ilmaisuvoimasta riippuen voidaan tehdä enemmän tai vähemmän sofistikoitunutta konepäättelyä.

Tyypillisesti käsitemallien haasteena on ollut siiloutuneisuus ja koneluettavuuden puute. Samoja termejä käytetään useissa hieman toisistaan poikkeavilla määritelmillä varustetuista käsitteistä eri malleissa, ja pahimmillaan eri toimijoiden itse määrittelemät käsiterakenteet johtavat siihen, että samankaltaisten käsitteiden varaan määritellyt tietorakenteet sisältävät sisällöltään osittain leikkaavaa mutta silti yhteismitatonta tietoa. Tämä on erityisen näkyvää tietokantaskeemojen lisäksi monissa koodistoissa sekä muissa luokittelujärjestelmissä.

Käsitemalleja voidaan kuitenkin määrittää koneluettavasti linkitettynä datana siten, että useita itsenäisesti määriteltyjä käsitemalleja voidaan yhdistää toisiinsa, analysoida niiden keskinäisiä tai sisäisiä ristiriitoja sekä edesauttaa niiden yhtenäistämistä. Yhteentoimivuusalustan Sanastot-työkalussa sanastoja määritetään RDF-pohjaisen SKOS-sanaston (Simple Knowledge Organization System) varaan. Kyseinen sanasto on kokoelma HTTP URI -muotoisia resursseja, joiden varassa voidaan määritellä ja yhdistää koneluettavia käsitemalleja. SKOS-pohjaisia käsitemalleja voidaan vuorostaan hyödyntää suoraan linkittämällä niitä RDF-pohjaisissa tietomalleissa määriteltyihin resursseihin (luokkiin, assosiaatioihin ja attribuutteihin). SKOS-käsitemallissa käytettäviä rakenteita ovat muun muassa laajempi tai kapeampi käsite, liittyvä käsite sekä vastaava tai lähes vastaava käsite. Sanasto-termin käytön osalta sekaannusta saattaa aiheuttaa se, että englanninkielistä termiä *vocabulary* käytetään yleisesti linkitetyn datan aihepiirissä kuvaamaan SKOS:n kaltaisia varatuista sanoista (tässä tapauksessa varatuista resurssien URI-tunnisteista) koostuvia kokoelmia, jotka mahdollistavat tiedon tyypittämisen ja rakenteiden kuvailun kyseisten varattujen sanojen varassa.

Perinteiseen käsitteellinen-looginen-fyysinen-mallinnustasojen jakoon liittyen on tärkeää huomioida, että käsitemalli ei ole loogisen mallin rakenteen vaihtoehtoinen esitysmuoto, eikä sen tehtävä ole alisteisesti täydentää loogista mallia sitä kuvaavalla metadatalla. Päinvastoin: looginen malli on käsitemallin johdannainen ja alisteinen sille käsitteistöön pohjautuvalle jaottelulle, jonka varassa valittua tietoaluetta koskevaa tietoa halutaan tuottaa ja hyödyntää (kts. JHS 179, liite 7, 4.2.). Näiden kahden mallin välinen linkki ei useinkaan ole yksi yhteen (jokainen käsite kuvautuu sitä vastaavaksi luokaksi käsitesuhteiden mukaisilla luokkien välisillä suhteilla).

Esimerkiksi käsite *henkilöasiakas* voi käsitemallissa olla määritelty käsitteiden *henkilö* ja *asiakas* alikäsitteeksi (jokainen *henkilöasiakas* on sekä *henkilö* että *asiakas*), mutta loogisesta mallista saattaa löytyä vain geneerinen *henkilö*-luokka, ja *henkilöasiakasta* semanttisesti vastaava tietue löytyy vain *henkilö*-luokan instanssidatasta koodiston arvolla *”asiakas”* tyypitettynä. Käsitepuolella käytetty taksonomia (*henkilöasiakas* on sekä *henkilö* että *asiakas*) ei automaattisesti johda siihen, että loogisessa mallissa täytyisi käyttää moninperintää haluttaessa määrittää *henkilöasiakas*-luokka, sillä kyse on perustavanlaatuisesti erityyppisistä skeemoista. TKK:n näkökulmasta henkilö ja rooli ovat käsitteellisesti itsenäisiä luokkia, mutta niiden yhdistettä (*henkilö,* jolla on roolina *asiakas* eli *henkilöasiakas*) ei löydy TKK:sta vaan ainoastaan sen varaan määritellystä instanssitiedosta tilanteessa, jolloin jokin tietty henkilö saa elinkaarensa aikana roolin *asiakas*. Tämä mallien välinen dissonanssi on erityisen vahvaa UML-mallintamisessa, jossa sanastotyön ja käsitemallien määrittelyssä luontainen polyhierarkia sekä saman entiteetin jaottelu useiden eri fasettien mukaisesti ei luonnistu, vaan tietomallirakenne joudutaan määrittämään käytännössä yhden perspektiivin puuhierarkiana.

Loogisen tason TKK ja SP-mallien sisältö (luokat, assosiaatiot, attribuutit) on linkitetty suoraan SKOS-pohjaiseen käsitemalliin. Loogisen mallin puolelta on siis mahdollista tarkastella suoraan niitä käsiterakenteita, joiden pohjalta loogiset rakenteet on johdettu.

3.3.12. Tietomallinnus loogisella ja teknisellä tasolla

Edellisessä luvussa mainitun mukaisesti loogiset mallit pohjautuvat yhteisessä käsitemallissa määritellylle käsitteelliselle tietoalueen ositukselle. Loogisen mallin määrittely on tehty käsitteitä vastaavien loogisten rakenteiden varaan niin pitkälle kuin mahdollista.

Tästä periaatteesta on jouduttu tinkimään vain yhdessä suhteessa: entiteettien välisten assosiaatioiden ominaisuuksien kuvauksessa on jouduttu noudattamaan niin kutsuttua reifikaatiota eli assosiaation ominaisuudet on kuvattu omalla luokallaan. Ratkaisu johtuu siitä, että standardoitu RDF-

mallin spesifikaatio ei salli esimerkiksi kahden entiteetin välisen assosiaation voimassaolon ilmaisua liittämällä kyseistä metatietoa itse assosiaatioon. Vastaavaa rakennetta (yhdistelmätaulu) joudutaan käyttämään myös relaatiomallin tapauksessa. Vielä kandidaattivaiheessa olevat RDF-star sekä SPARQL-star -standardiluonnokset ratkaisevat tämän ongelman, ja niitä voidaan hyödyntää suoraan suosituimmissa kaupallisissa ja avoimen koodin RDF-kannoissa sekä SPARQL-endpointeissa. Star-kandidaatteja ei voitu kuitenkaan hyödyntää tässä hankkeessa, sillä Yhteentoimivuusalustan tuki rajoittuu standardoituihin versioihin.

Yhteentoimivuusalustalle tuotettuja malleja ei niiden tämänhetkisestä (H1 2022) luokkakaaviomaisesta kuvaustavasta huolimatta tule sekoittaa UML-malleihin. Implementoidessa malleja – erityisesti tulkittaessa kuvattua rakennetta DDL:llä relaatioskeemaksi – on keskeistä muistaa linkitetyn datan mallien rakenteen keskeiset pilarit:

- Yhteentoimivuusalustalla näkyvät luokat, assosiaatiot ja attribuutit ovat kaikki itsenäisiä atomisia URI-muotoisia resursseja graafimuotoisessa tietorakenteessa. Luokka ei UML:n tapaan omista sille määriteltyjä assosiaatioita ja attribuutteja, vaan nämä ovat kaikki itsenäisiä subjekti-predikaatti-objekti-kolmikoita (engl. *triple*), joissa predikaattina toimiva resurssi yhdistää binäärisesti aina yhden subjektin yhteen objektiin. Erona attribuutti- ja assosiaatioresursseilla on se, että attribuuttiresurssi on predikaatti, joka yhdistää subjektin johonkin XML-datatype -literaaliarvoon (esimerkiksi xsd:DateTime -tyyppiseen aikaleimaan), kun taas assosiaatioresurssi yhdistää kaksi (URI) resurssia keskenään.
- Pysyvät tunnisteet instanssitiedolle annetaan sitä luotaessa, eikä niiden esittäminen osana skeemaa ole lähtökohtaisesti skeemana käytettävän mallin asia. Tietoa skeeman mukaisesti luovan tahon vastuulla on valita käytettävä nimiavaruus ja sen puitteissa jaettavien tunnisteiden nimeämiskäytäntö (esimerkiksi URL + UUID URN -muotoinen URI). Mikä tahansa tiedon tuottajan luoma URI muuttuu ”skeeman mukaiseksi” vasta ilmoittamalla hallitussa datassa triplenä, että kyseinen resurssi on tyypiltään jokin skeeman resurssi (eli instanssidatan subjektina toimiva resurssi yhdistetään skeeman objektina toimivaan resurssiin RDF-sanaston URI-resurssilla `http://www.w3.org/1999/02/22-rdf-syntax-ns#type`, tai nimiavaruuksien etuliitteitä esimerkiksi .ttl -serialisoinnissa käyttämällä lyhyemmässä muodossa `rdf:type`).
- Lähtökohtaisesti jokainen URI-resurssi resolvoituu (eli esimerkiksi HTTP-pyynnöllä kyseisen resurssin URI palauttaa sitä koskevaa tietoa). Instanssitiedon URI-tunnisteiden jakelu voidaan toteuttaa yhdistettynä esim. REST-rajapintaan polkurakenteella `https://<domain>/<DigiPAVe-TKK:n resurssityyppi>/<instanssitiedon UUID>`.

3.3.13. Metatietojen hallinta

TKK:n pääasiallinen rooli on esittää mallinnettavan kohdealueen sekä siihen liittyvien tietojen tilan kokonaiskuvaa suhteessa aikaan; malli kuvaa tietokokonaisuuden ennakoitua tai suunniteltua, nykyistä tai jo toteutunutta tilaa. Näiden tilojen sisällön kuvaukseen käytettävillä rakenteilla ensisijainen metatieto on niiden elinkaari ja viittaus tiedon edelliseen versioon. Tällä on mahdollistettu helppo tapa selvittää tietokokonaisuuden tila halutulla ajanhetkellä tai välillä, tai porautua jonkin tietyn resurssin osalta sen muutoshistoriaan.

Mallissa ei kuitenkaan ole kuvattu laajemmin provienienssia eli tietojen alkuperä-, omistajuus- tai sijaintihistoriaa syntyhetkestä lähtien. Tähän on olemassa laajalti käytössä olevia ja vakiintuneita metadatan kuvaustapoja, kuten Dublin Core tai PROV-O-ontologia, joita suositellaan käytettävän instanssitiedon resursseja täydentävinä sanastoina.

3.3.14. Koodistojen ja formaattien valinta

Koodistojen rooli malleissa on keskeinen: TKK:n tasolla kuvatut luokat, assosiaatiot ja attribuuttityypit edustavat niitä pysyviä entiteettejä, joita mallinnettavan kohteen kontekstissa on katsottu tarvittavan tilanteesta riippumatta. Näitä täydentämään on laadittu erinäisiä koodistoja, joilla on ulkoistettu itse mallin rakenteesta erilliseen taksonomiaan sellaisia sisältöjä, joiden mallintamisesta kiinteiksi osiksi TKK:ta ei ole katsottu olevan tässä tilanteessa hyötyä, tai joiden on katsottu potentiaalisesti heikentävän mallin muutosjoustavuutta suhteessa tulevaan harmonisointityöhön.

Osa hyödynnetyistä koodistoista on vakiintuneita standardeihin pohjautuvia tai jo käytössä olevia: muun muassa IETF-kielikoodit, kuntaluokitus, rakennusluokitus, julkisten palvelujen luokat, tuottajatyypit, kohderyhmät ja tuotantotavat. Osa on luotu itse TKK:n JHS-tasoisten luokkien (henkilö, tapahtuma, jne.) mukaisesti ehdotuksiksi kansallista harmonisointia silmällä pitäen.

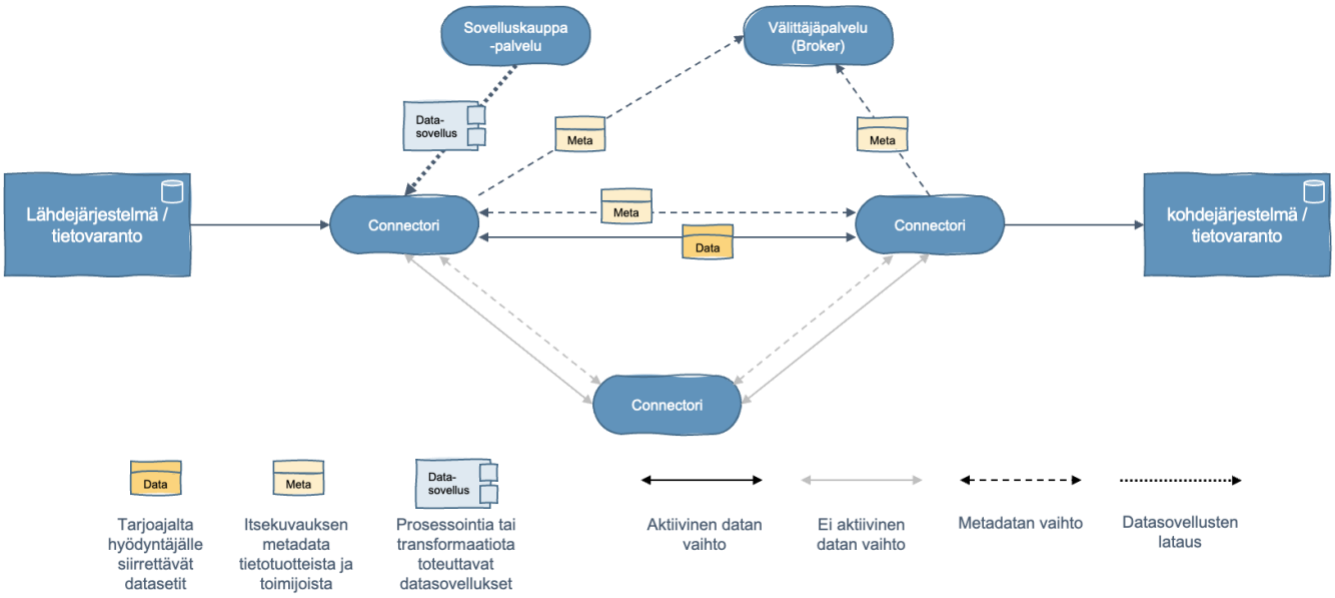
Instanssitietoa käsiteltäessä keskeinen kysymys on tiedon serialisoinnin muoto rajapinnalta kyseltäessä tai sen kautta tietoa siirrettäessä. Ensisijaisesti suositellaan käytettäväksi JSON-LD (JSON Linked Data) -skeemaa, sillä se on sekä laajalti tuettu että helposti parsittavissa tarpeen tullen myös puhtaana JSON-tietorakenteena. Tietoa voidaan ottaa käyttöön myös muissa muodoissa serialisoituina, esimerkiksi puhtaana XML:nä, jolloin elementteihin voidaan liittää SADWSDL (Semantic Annotations for WSDL and XML Schema) -metatietokuvauksena elementtien yhteys tiedon pohjana olevaan ontologiaan.ja hyödynnettävissä asiakasohjelmissa

3.3.15. Datasovellusten väliset rajapinnat

Ekosysteemiä tukemaan voidaan ottaa käyttöön datan manipulointia toteuttavia datasovelluksia ja niitä tarjoava sovelluskauppapalvelu. Tässä tapauksessa datasovelluksilla on hyvä olla itsekuvaukset, jotta ne voidaan tarvittaessa jakaa ja ladata katalogien sekä erillisten sovelluskauppojen (esimerkiksi ITSM-järjestelmät) avulla ekosysteemin toimijoille. Datasovellusten tarjoamat operaatiot kuvataan katalogiin syötteiden, tulosteiden, protokollien ja endpointien avulla.

Datasovelluksia voidaan sisällyttää dataa siirtäviin connectoreihin, josta niitä hyödynnetään, kuten mitä tahansa muuta connectorin sisältämää palvelua.

Alla on kuvattu yksinkertaisesti datan välityksessä hyödynnettävien teknisten komponenttien yhteistoiminta ja interaktiot.



Kuva 13 - Datan jakamisen ja välittämisen komponentit sekä niiden vuorovaikutus

3.3.16. Tietovarantojen määrittäminen

Kappaleen 3.4.3. mukaisesti tietovarantojen määrittäminen tulisi toteuttaa organisaation lähtötilanteen ja asetetun tavoitetason perusteella. Erityisesti siirtymävaiheessa saattaa olla hyödyllistä tukeutua olemassa oleviin lähtötietovarantoihin ja hyödyntää TKK:ta harmonisoidun tiedon tietovarantona, joka voi palvella palveluverkkosuunnittelussa mukana olevia toimialueita ja rooleja sekä mahdollisesti sen kanssa vuorovaikuttavia toimialueita.

Harmonisoitua muotoa voidaan käyttää referenssimallina uudistettaessa muita tietovarantoja, erityisesti haluttaessa varmistaa, että tietyn tietoalueen tietoa tuotetaan ydintietona vain yhteen varantoon, ja että organisaatiossa ei ole kilpailevia ristiriitaisia näkymiä kyseisen tiedon semanttisesta määritelmästä, rakenteesta ja suhteesta muuhun organisaation tuottamaan tietoon.

Tietovarantojen määrittäminen on vain yksi osa laajempaa kokonaisprosessia, jossa organisaatio harmonisoi käsitteistönsä, tietorakenteensa ja prosessinsa. Ensimmäisiä konkreettisia askeleita kyseisellä kehityspolulla ovat:

- Prosesseissa hyödynnettävän tiedon alkulähteen jäljittäminen ja kyseisen lähtötiedon muodon tunnistaminen.
- Prosesseissa tuotettavan tiedon yksikäsitteisen tallennuspaikan ja ydintietomäärittelyn tuottaminen.
- Joko määritelmältään tai sisällöltään päällekkäisten säilöttyjen tietojen tunnistaminen, keskinäinen harmonisointi ja kilpailevien tuotantoprosessien poistaminen.
- Ydintietojen uudelleenmäärittely yhteisen ontologian varaan, niiden normalisointi ja tarjoilu rajapinnoilla ontologian mukaisilla metatiedoilla varustettuna.

Itse tiedon teknistä hallintaratkaisua (relaatiokanta, graafikanta, avain-arvokanta tms.) merkittävästi tärkeämpää on se, missä muodossa tieto on jäsenneltynä loogisissa tietovarannoissa. Olennaista on saada eri käyttötapauksissa tarvittava tieto käyttöön rajapintojen kautta siten, että sen linkitetyn datan metatiedot ovat mukana. Tämä mahdollistaa eri prosesseissa, eri toimialueilla, eri tilanteissa ja eri muodoissa käyttöön otetun tiedon semanttisen yhteentoimivuuden säilyttämisen, sillä jokainen tietue on linkitetty pohjalla määritettyyn ontologiaan ja täten eri muodoissa esitettävä tieto voidaan yhteismitallistaa ja sijoittaa yhden ja saman tietorakenteen kehykseen. Samalla mahdollistetaan myös

tiedon jäljitettävyys, eli sanomien sisältöä ei jouduta tulkitsemaan itse sanoman rakenteesta, vaan semanttinen annotaatio itse sanoman rakenteessa antaa suoran linkin sen kontekstualisointiin.

Edellä mainittujen ratkaisujen käyttöönotto vaatii toteuttavalta organisaatiolta holistista otetta; erityisesti tietoarkkitehdeilta vaaditaan mediaattoriroolia substanssipuolen tarpeiden, teknisten ratkaisujen ja yhteentoimivuuskehikon yhteensovittamisessa. Samoin tiedonhallintaan käytettävien järjestelmien ja rajapintojen toteutuksessa tulee käyttää tarvepohjaista harkintaa paradigman valinnassa, sillä liian tiukka etukäteinen sitoutuminen tiettyihin teknisiin ratkaisuihin voi muodostaa merkittävän esteen tavoiteltavan harmonisoidun tiedonhallinnan käytännön toteutumiselle.

3.4. Tietojärjestelmäarkkitehtuurin suunnittelu

Tässä luvussa kuvataan yleisellä tasolla datatuotteiden ja ekosysteemin hallintaan liittyvät tekniset komponentit, joilla tarkennetaan turvalliseen ja luotettavaan tiedonvaihtoon vaadittavat järjestelmä- ja infrastruktuuripalvelut.

IDS erottelee järjestelmätasolla kolme keskeistä pääelementtiä, jotka mahdollistavat tiedon jakamisen, hakemisen ja siirtämisen dataekosysteemissä:

- Connectori, eli tekninen komponentti, jolla hallitaan datatransaktiota.
- Välittäjäpalvelu, jolla on oma connectori ja se tarjoaa palvelut datatuotteiden ja connectorien rekisteröimiseen, julkaisuun, ylläpitoon ja kyselyihin.
- Datasovelluskauppa, joka tarjoaa keskitetysti erilaisia datankäsittelykomponentteja, joita voidaan liittää osaksi connectoreiden toiminnallisuutta.

Lisäksi on tarkennettu edellä mainittuja tiedon jakamisen komponentteja tukevat järjestelmäpalvelut:

- Identiteetinhallinta, jolla rekisteröidään ja hallitaan ekosysteemitoimijoiden identiteettejä sekä taustatietoja.
- Sanastohubi, johon tallennetaan ja ylläpidetään ekosysteemin sisällä tai toimijakohtaisesti käytettävät deklaratiiviset mallit sanastoista, eli sanastoskeemat.
- Päivitystietovaranto, johon talletetaan connectorien konfiguraatiot ja jonka kautta connectoreihin tehdään päivityksiä.
- Luottamustietovaranto, johon kootaan tiedot luotettavista teknologiakokonaisuuksista.

Edellä mainituista komponenteista voidaan vielä koostaa Gaia-X:n kuvaamat federoidut katalogit, jonne kerätään luettelotiedot connectoreista ja niiden sisältämistä datatuotteista sekä muista resursseista ja ekosysteemissä toimivista identiteeteistä.

Ekosysteemiarkkitehtuuriin voidaan liittää myös välitallennusta, lokihallintaa sekä monitorointia ja mittarointia toteuttavia järjestelmäpalveluja. Lisäksi ekosysteemiin voidaan liittää yhteisesti jaettuja infrastruktuuripalveluja.

Ekosysteemiarkkitehtuurin teknisiä palveluja voidaan koostaa ja hajauttaa erinäisiin tietojärjestelmiin. Esimerkiksi connectorit, katalogipalvelut, luottamuksenhallinta sekä välittäjäpalvelut voidaan koostaa yhteen alustajärjestelmäkokonaisuuteen tai toteuttaa hajautetusti eri järjestelmillä ja alustoilla.

3.4.1. Connectorit

Connectorit ovat tiedon virtautuksen keskeisin tekninen palvelukomponentti, jolla dataa kyetään koostamaan ja jakamaan siihen määritetyn tietomallin mukaisesti. Tässä luvussa esitellään yleisellä tasolla connectoreiden tekninen arkkitehtuuri, connectoreiden ja niiden hallinnan komponentit sekä connectoreiden luomisen periaatteet. Connectoriarkkitehtuuri perustuu tässä IDS-referenssiarkkitehtuurimallin mukaiseen määritykseen, joka puolestaan perustuu mikropalveluihin

sekä sovelluskontteihin, mutta connectoreja voidaan tai joudutaan usein toteuttaa myös muilla ratkaisumalleilla, kuten erilaisten ohjelmistojen omilla rajapintapalveluilla.

Connectorit rakennetaan usein connector-palveluja tarjoavien kaupallisten tuotteiden päälle tai erilaisten pilvialustojen omilla komponenteilla. Jälkimmäisessä tapauksessa hyödynnetään alustapalveluja, kuten virtuaalikoneita ja palvelimia, näiden käyttöjärjestelmiä sekä kontinhallintapalveluja, kuten Dockeria, GitLab:ia, trustme:ta, OpenShift:iä, Kubernetesta tai Amazon ECS:ää. Mikäli aluksi ei ole connector-palveluja tarjoavia työkaluja käytössä eikä niitä päätetä hankkia, voidaan datan avaamiseen hyödyntää mahdollisuuksien mukaan kunnan olemassa olevia integraatiotyökaluja tai jakaa dataa suoraan tietovarantoja hallinnoivien järjestelmien ohjelmointirajapinnoilta. Jälkimmäisissä tapauksissa on suositeltavaa ensin tarkastella, voidaanko datatuotteita muodostaa ja hallita järkevästi.

Connectoreja suunniteltaessa tulee huomioida palvelun laatuvaatimukset (QoS eli Quality of Service), esimerkiksi yhtäaikaisten käyttäjien, päivitysten ja tiedonsiirtokapasiteetin osalta. Connectoripalveluja tukevan alustan tulisikin pystyä skaalautumaan suunnitellun käyttötarpeen mukaisesti, mutta mielellään myös siten, että käyttövolyymien lisääntyessä connectoreille kyetään ohjaamaan lisää kapasiteettia ja resursseja. Myös skaalautuvuuden hallintaan on tällä hetkellä tarjolla useita palveluja, jotka mahdollistavat kapasiteetin dynaamisen skaalauksen tapahtumamäärien mukaisesti. Kapasiteetti- ja skaalautumisvaatimukset on syytä huomioida myös connectorien sisältämiä hakualgoritmejä valitessa, jotta resurssitarpeet pysyvät kohtuullisina.

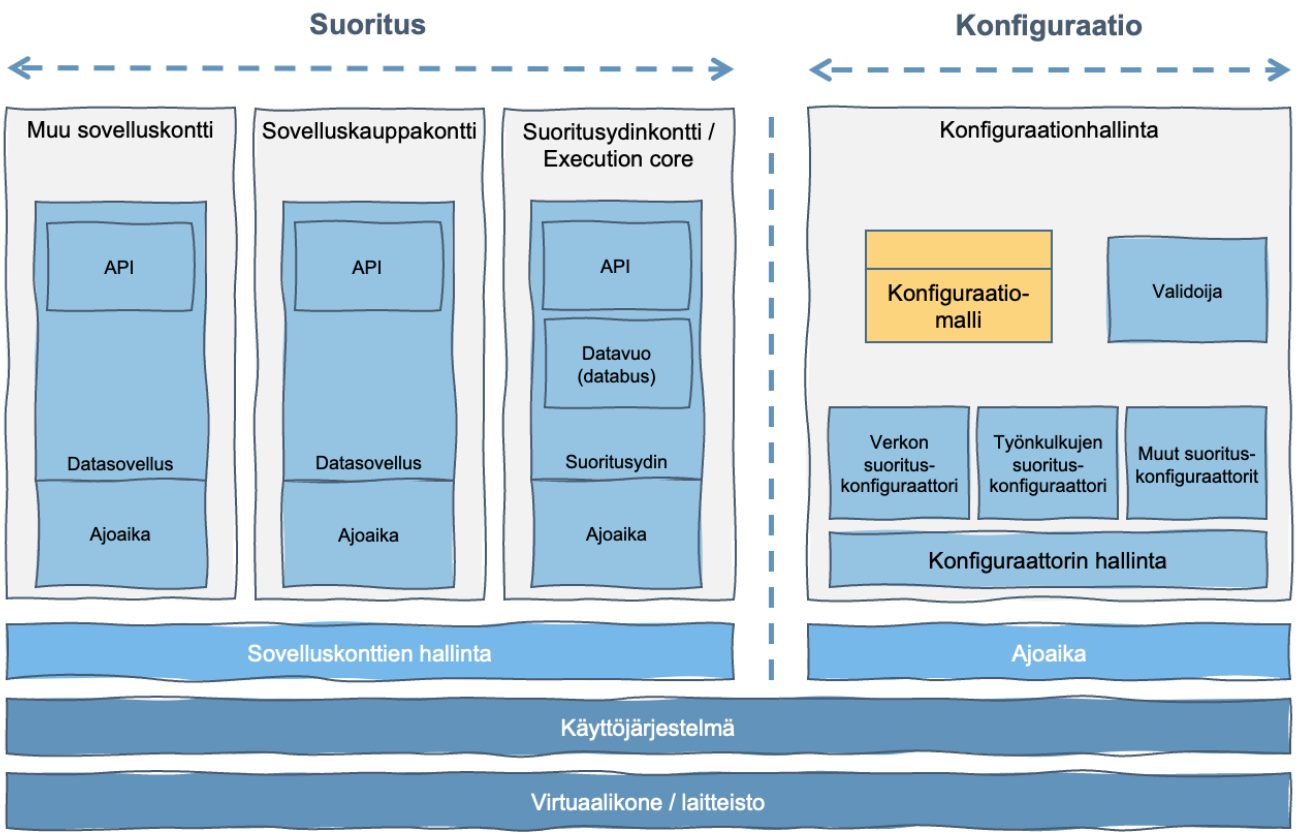
Sovelluskontitukseen perustuva connectoriarkkitehtuuri mahdollistaa datapalveluiden eristämisen ja turvallisen hallinnan konttihallinnan avulla. Datapalvelut yhdistävät niiden järjestelmien API:t, jotka tarjoavat datan tallennuksen, pääsynhallinnan ja prosessoinnin palveluja.

Mikäli jaettu data on luottamuksellista, tulee tietoturvan vuoksi datan käsittely suorittaa mahdollisimman lähellä datalähdettä, josta sitä jaetaan. Tässä voidaan hyödyntää eri käyttötarkoituksiin luotuja connectoreita sisäiselle datan käsittelylle ja datan ulospäin jakamiseen tarkoitetulle connectorille. Datan esikäsittely suoritetaan siis ensin sisäisillä connectoreilla ennen datan jakamista ulospäin.

Connectoreiden erityistapauksia, kuten kehittäjien connectoreita, mobiiliconnectoreita ja connectoreiden miniatyrisointia ei esitellä tässä, mutta niistä voi lukea lisää IDS-referenssiarkkitehtuurin luvusta 3.5.1.2.

Connectorit tulee dokumentoida huolellisesti ja ylläpitää kuvauksia julkishallinnon API-periaatteiden mukaisesti.

Connectorit koostuvat kahdesta vaiheistetusta komponentista: suorituksesta tai ajosta sekä konfiguraatiosta ja näiden alikomponenteista.



Kuva 14 - Connector-komponentin arkkitehtuuri

Ajoaikavaihe

Connectorin ajoaikavaihe sisältää seuraavat komponentit:

Sovelluskontinhallinta / Application Container Management

Sovelluskontinhallinnan käyttö on suositeltavaa silloin, kun connectoreiden datapalvelut sisällytetään sovelluskontteihin. Tällöin datapalvelujen välille ei synny tarpeettomia riippuvuuksia, vaan ne pysyvät uudelleenkäytettävinä silloinkin, kun datatuoteaihiosta on tarpeen luoda useita samankaltaisia, mutta toisistaan jonkin verran poikkeavia datatuotteita. Sovelluskontinhallinta mahdollistaa usein myös datapalvelujen tarkemman hallinnan.

Sovelluskontinhallintaa ei välttämättä tarvita datatuotetta kehitettäessä. Sovelluskontinhallinta ei myöskään ole mahdollista kaikilla teknologiatuotteilla.

Execution Core -kontti

Execution Core -kontti sisältää komponentit datapalvelujen rajapinnoille sekä viestinvälitykseen:

Datareititin / Data Router vastaa konfiguraatioparametrien mukaisesta viestinvälityksestä, eli kuinka dataa lähetetään ja vastaanotetaan datavuon ja datapalvelujen välillä. Datareititttäjä voidaan toteuttaa jollain valmiilla kaupallisella komponentilla tai luoda itse. Se voidaan myös kovakoodata connectoriin tai avata yhteydet suoraan datapalveluun, mutta näitä implementaatiomalleja kannattaa toteuttaa vain tapauksissa, joissa connectorialusta ja datapalvelut ovat rajoittuneita ja yksinkertaisia (esimerkiksi sensorilaitteet, kuten melumittarit). Datatareititin voidaan konfiguroida myös käyttöehtojen hallintaan.

Datavuo / *Databus* vastaa tiedonvälityksestä datapalvelujen ja muiden connectoreiden omien datavoiden välillä, eli se vastaa connectoreiden välisestä dataliikenteestä. Myös datavuo voidaan korvata erilaisilla implementaatioilla. Datavuon tyypin ja tuotteen valinnassa kannattaa huomioida muun muassa datan siirron kapasiteettivaatimukset, mahdolliset kustannukset sekä tarvittavat lisäpalvelut.

Datapalvelu-API / Data Service API

Datapalvelu-API on connectorin julkinen rajapinta, jota kutsutaan datareitittimellä. API:n avulla voidaan ajaa useita erilaisia datapalveluja, joita on sisällytetty connectorin konfiguraatioon. Datapalvelu-API kuvataan konfiguraatiomallin sisältämään kuvaukseen.

Ajoaikaympäristö

Datapalvelun ajoaikaympäristö muodostaa kontin ytimen yhdessä datapalvelun kanssa. Ajoaikaympäristö voidaan toteuttaa useilla eri tavoilla ja teknologioilla, jotka voivat vaihdella konteittain. Yleensä ajoaikaympäristövaihtoehdot riippuvat alustan käyttöjärjestelmästä, joten molemmat tulee valita niiden tarjoamien kyvykkyyksien mukaisesti.

Connectori voi sisältää myös sovelluskauppapalvelusta ladattuja datapalveluja sisältäviä kontteja sekä erillisiä itsekehitettyjä datapalvelukontteja, joita ei välttämättä tarvitse sertifioida ekosysteemissä.

Tarvittavat kontit ja komponentit voivat vaihdella connectorikohtaisesti riippuen datatuotteen käsittelytarpeesta.

Konfiguraatiovaihe

Connectorin konfiguraatiovaihe sisältää seuraavat komponentit:

Konfiguraatiomanageri / Configuration Manager

Konfiguraatiomanageri hallinnoi ja validoi connectoreiden konfiguraatiomalleja connectoreita käyttöönotettaessa. Varsinainen käyttöönotto tapahtuu konfiguraattorinhallinnan delegoimilla ajokonfiguraattoreilla.

Konfiguraattorinhallinta

Konfiguraattorinhallinta lataa ja hallinnoi connectorin ajokonfiguraattoreita. Se vastaa eri tehtävien delegoinnista jokaiselle konfiguraatiomallissa osoitetulle ajokonfiguraattorille.

Ajokonfiguraattori / Execution Configurator

Ajokonfiguraattorit implementoivat konfiguraatiomallissa määriteltyjä aspekteja, ja ovat yleensä sidoksissa komponenttien toteutusteknologiaan. Jokaisella teknologialla, kuten käyttöjärjestelmillä, sovelluskontinhallintakomponenteilla, on omat ajokonfiguraattorinsa. Ajokonfiguraatioiden suorittaminen voidaan toteuttaa esimerkiksi konfiguraatiotiedostoilla tai erillisellä konfiguraatio-API:lla, riippuen valituista toteutusteknologioista.

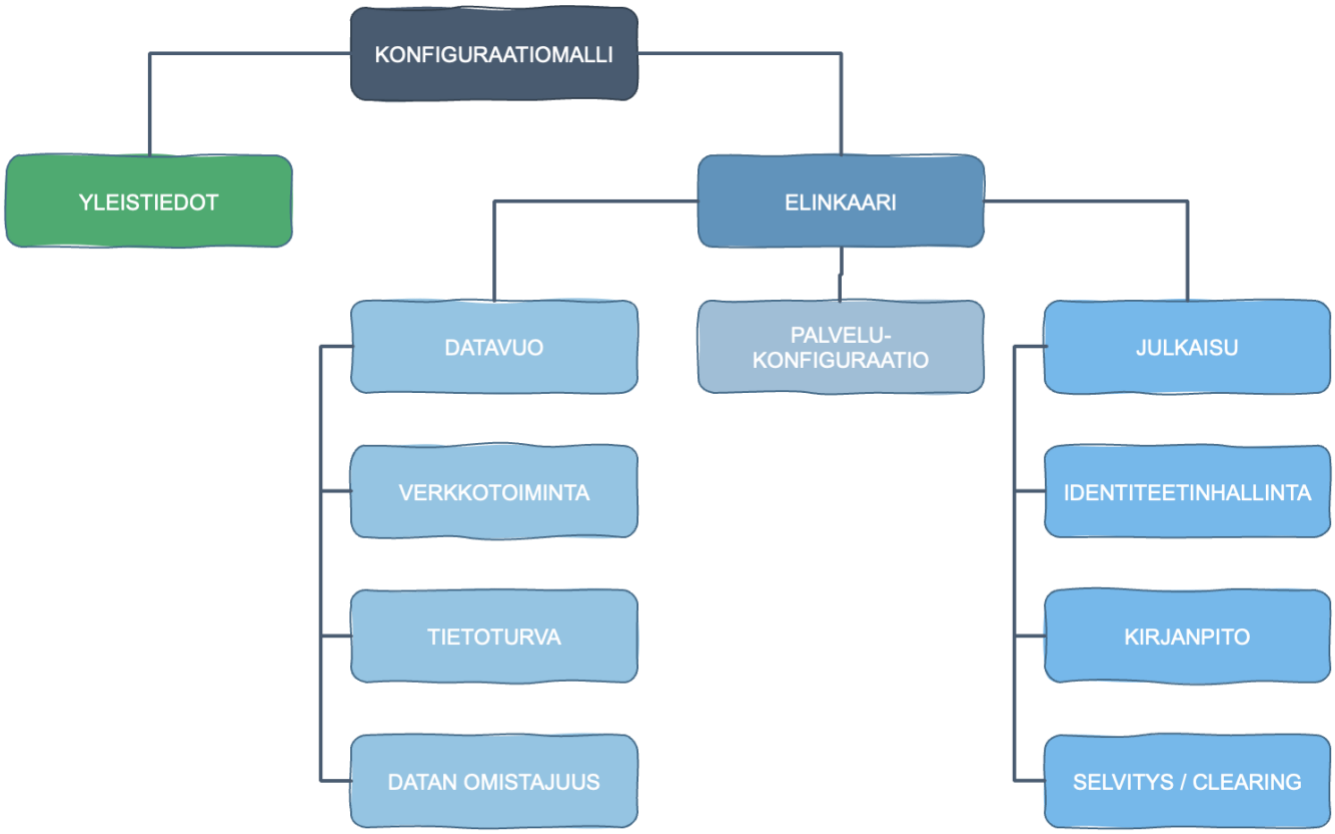
Konfiguraatiomalli / Configuration Model

Konfiguraatiomalli kuvaa konfiguraation, jolla connectori otetaan käyttöön. Connectori voidaan konfiguroida eri statuksilla, kuten kehitys-, testi- ja käyttökonfiguraatio.

Konfiguraatiomalli implementoidaan erillisten ajokonfiguraattoreiden avulla, joilla toteutetaan seuraavat osiot konfiguraatiomallista:

- *Yleistiedot* sisältävät konfiguraatiotyypin sekä connectorin tyypin (esimerkiksi perus- tai kehittäjä-connectori), connectorin version, konfiguraation statuksen ja edellisen muutoksen aikaleiman. Se voi sisältää myös kontaktihenkilön nimen, mutta tässä kannattaa huomioida GDPR ja korvata nimi jollain muulla yhteystiedolla.

- *Elinkaari* sisältää ohjeet datavirralle, palvelukonfiguraatiolle sekä julkaisulle.
    - *Datavirta* määrittelee konfiguraatiot tehtäville ja yhteyksille, jotka datareititin luo datapalvelujen ja datavoiden välille. Datavirta käsittelee yhteyden muodostamista ja tietoturvaa. Tarkempia kuvauksia tietoturvan toteutuksesta IDS-referenssiarkkitehtuurin mukaisesti voi lukea kyseisen dokumentin luvusta 4.1.
        - *Verkkokonfiguraatio* määrittää verkkoparametrit, kuten IP-osoitteet, portit ynnä muut connectorin sisällä sekä yhteyksille muihin connectoreihin. Verkkokonfiguraatiota määrittäessä pitää huomioida myös mahdolliset tarvittavat palomuuriavaukset.
        - *Tietoturvakonfiguraatio* sisältää tiedot tietoturvaparametreista, kuten kuljetuskerroksen suojausprotokollan ja mahdolliset muut kryptografiset ratkaisut. Tietoturvakonfiguraatio riippuu ekosysteemistä, johon liitytään, tai sopimuksesta connectoreiden omistajaosapuolten välillä.
        - *Datan omistajuus tai myöntyminen (compliance)* määrittää säännöt, jotka konfiguraationhallinnan validoija tarkistaa ennen connectorin käyttöönottoa. Myöntymistä hyödynnetään estämään connectorin virheellinen konfigurointi.
    - *Palvelukonfiguraatio* määrittelee, kuinka connectorin komponenttien, kuten datapalvelujen parametrit tulee asettaa. Palvelukonfiguraation *metadataosio* määrittää connectorin komponenttien syötteet ja tulosteet (katso luku 3.3.9.). Datapalvelut voivat myös tuottaa edellä mainittuja metadatakuvauksia, jotka voidaan importoida konfiguraatiomalliin datavirran konfiguroimiseksi.
    - *Julkaisu* määrittelee mitkä connectorin datavirrat ja -palvelut tarjotaan ulkoisille tahoille. Julkaisuosio julkaistaan datavälittäjään (data broker), josta connectorin tiedot ovat muiden luvitettujen ekosysteemitoimijoiden nähtävillä.
    - *Identiteetinhallinta* määrittelee connectoriin integroidun identiteetintarjoajan, jos tällaista käytetään ekosysteemissä. Erilaiset standardoidut verkot, kuten taloustietojen siirtoon tarkoitettu PEPPOL ja erilaiset lohkoketjut sisältävät yleensä identiteetinhallintakomponentteja. Yhteyden muodostamiseksi identiteetinhallintaan saatetaan tarvita datapalveluilta erillisiä kirjastoja, joiden avulla identiteetti kyetään tarkistamaan.
    - Kirjanpito määrittää tarkemmat tiedot datan vaihdon transaktiosta osapuolten välillä, kuten sopimustiedot, hinnoittelumallit ja laskutustiedot. Näitä tietoja hyödynnetään erityisesti silloin, kun dataa hyödyntävä osapuoli maksaa datasta.
    - Clearing kuvaa, ketä clearing-palveluntarjoajaa tulee informoida datatransaktiosta, jotta transaktiot kyetään selvittämään.



Kuva 15 - Konfiguraatiomallin rakenne

Validoija / Validator

Validoija tarkistaa konfiguraatiomallin yhteensopivuuden ekosysteemin sääntöjen kanssa. Erilaisilla ekosysteemeillä voi olla erilaisia validointitarpeita. Jos käyttöönotettavassa konfiguraatiomallissa esiintyy virheitä (esimerkiksi osapuoli ei ole luotettu ekosysteemitoimija), voi validoija keskeyttää konfiguraatioprosessin.

Tietoturvan hallinta

Connectoreille voidaan myöntää staattiset turvallisuustasot yhteydenmuodostusten luvittamiseksi, mutta sekä IDS että Gaia-X mahdollistavat luvituksen aikana tehtävät kutsuvan connectorin turvallisuustarkistukset erillisen turvallisuusprofiilin avulla. Connectorien turvallisuusprofiileista sekä tiedonvaihdon turvallisuudesta on ohjeistettu lisää IDS-referenssiarkkitehtuurin luvussa 4.1. Turvallisuusprofiilit voidaan myös linkittää dynaamiseen attribuuttien hallintapalveluun (DAPS), joka mahdollistaa turvallisuusprofiilin jakamisen datatransaktion osapuolten välillä. DAPS:a kuvataan tarkemmin seuraavassa luvussa.

Connectoriarkkitehtuurin sekä yleisesti ekosysteemiarkkitehtuurin tietoturvaa suunniteltaessa on syytä hyödyntää yleisiä standardeja ja hyviä käytäntöjä. Esimerkiksi connectorien välistä pisteestä pisteesen kryptaustunneloinnin avulla sekä päästä päähän luvitus tulee mahdollistaa uusia ekosysteemejä luodessa – varsinkin, jos connectori tarjotaan näkyväksi kunnan ulkopuolisille tahoille. Useissa ekosysteemeissä on kuitenkin jo valmis turvallisuusarkkitehtuuri, johon tulee mukautua.

IDS-referenssiarkkitehtuuri tarkentaa luvussa 4.1. turvallisuusarkkitehtuurin rakentamisessa huomioitavia seikkoja.

3.4.2. Katalogit ja datan välittäjä

Katalogiin julkaistaan ekosysteemin datatuotteiden, niitä palvelevien connectorien sekä muiden resurssien ja toimijoiden itsekuvaukset, josta ne ovat löydettävissä ekosysteemin muille toimijoille.

Katalogit vaativat hakutoiminteet ja algoritmit datatuotteiden, resurssien ja ekosysteemitoimijoiden hakemiseen erityisesti silloin, kun sinne rekisteröityjen itsekuvausten määrä on suuri.

Katalogiin on hyödyllistä rakentaa myös navigaatiorajapinnat, jotta tietoja kyetään hallitsemaan ja esittämään käyttäjille hyödyllisellä tavalla.

Katalogit voivat sisältää yhteydenmuodostusten ja datasiirtojen ajonaikaisen välittäjäpalvelun, esimerkiksi integraatioalustoja hyödynnettäessä, mutta tätä toiminnallisuutta ei hyödynnetä kaikissa ekosysteemeissä.

Katalogeja voidaan myös federoida siten, että erilliset ja eri sijainneissa toimivat katalogi-instanssit ylläpitävät kukin vain tiettyä joukkoa saatavilla olevista itsekuvauksista. Tämä on hyödyllistä erityisesti silloin, kun itsekuvausten tietoja halutaan ylläpitää toimijakohtaisesti tai katalogeja on useilla eri alustoilla. Tässä tapauksessa katalogi-instanssien tulisi kuitenkin kyetä verifioimaan ja näyttämään ekosysteemin muiden instanssien tiedot ja sisältämät itsekuvaukset. Lisäksi ekosysteemitoimijoiden tulisi kyetä varmentamaan itsekuvaukset niiden alkuperäisestä sijainnista.

Jotta katalogissa tai katalogeissa esitettyjen itsekuvausten näkyvyyttä kyetään hallitsemaan, tulisi niiden kyetä tarjoamaan käyttäjä (toimija) tai käyttäjäryhmäkohtaiset näkymät.

Katalogeja voidaan luoda muun muassa erilaisten integraatiotyökalujen API-kirjastoihin, dedikoiduilla luottamusalustoilla sekä erillisillä katalogipalvelut mahdollistavilla ohjelmistoilla. Alustaa valittaessa on kuitenkin syytä huomioida vaatimukset, datankäyttösopimukset sekä käyttäjienhallinta katalogin datatuotteiden hallintaan, jotta luotettava ja turvallinen datan avaaminen mahdollistuu.

Datan välittäjä (data broker)

Datan välittäjä koostuu connectorista tai connectoreista, joihin sisältyy palvelut datalähteen tai muun resurssin rekisteröimiseksi, julkaisemiseksi, kuvauksen ylläpitämiseksi ja hakemiseksi. IDS määrittelee datan välittäjän connectorin ja palvelut samaan tapaan, kuin muissakin vastaavissa connectoreissa. Datan välittäjän tarkoituksena on kuitenkin tarjota federoitu katalogi ekosysteemiin rekisteröidyistä tietotuotteista, jonka kautta tietotuotteet löydetään. Datan välittäjän hyödyntämisen prosessit riippuvat ekosysteemissä hyödynnettävästä välittäjästä. Datan välittäjän prosesseista on ohjeistettu ja kuvattu tarkemmin IDS-referenssiarkkitehtuurin luvussa 3.3.

3.4.3. Identiteetin- ja luottamuksenhallinta

Identiteetinhallinta on palvelu tai palvelukokonaisuus, jolla rekisteröidään ja hallitaan ekosysteemitoimijoiden identiteettejä ja taustatietoja. Se tarjoaa palvelut ekosysteemitoimijoille identiteettitietojen luomiseen, ylläpitoon, hallintaan, monitorointiin ja validointiin. Identiteetinhallinta on keskeinen osa ekosysteemin turvallisuuden kannalta ja vähentää riskiä luvattomasta dataan pääsystä.

Identiteetinhallinta koostuu:

1. Sertifiointipalvelusta, jolla hallitaan ekosysteemitoimijoiden sertifikaatteja, eli todistuksia siitä, että toimijat ovat luotettavia.
2. Dynaamisesta attribuuttien provisiointipalvelusta, jolla hallitaan ekosysteemitoimijoiden tietoja.
3. Dynaamisesta luottamuksen monitoroinnista, jolla kyetään valvomaan ekosysteemin turvallisuutta ja verkon käyttäytymistä.

Näistä palveluista voi lukea tarkemmin IDS-referenssiarkkitehtuurin luvusta 4.1. ja Gaia-X:n luvuista 3.5. ja 3.6.

Identiteetinhallintapalvelut voidaan kunnassa rakentaa yleensä ainakin osittain jo olemassa olevien identiteetinhallinnan ratkaisujen, kuten Active Directoryn tai vastaavan päälle.

Sertifiointipalvelu

Joissakin dataekosysteemeissä sertifiointi voi sisältää sertifiointitoimijoiden suorittamaa tarkastelua, jossa toimijat, heidän tilansa, datatuotteensa ja tekniset ratkaisunsa tarkistetaan fyysisesti. Tässä dokumentissa keskitytään kuitenkin tarkastelemaan teknisesti suoritettavaa sertifiointia ja sertifikaatin hyödyntämistä. Eri arkkitehtuureihin perustuvilla ekosysteemeillä voi olla erilaisia digitaalisia sertifikaatteja (esimerkiksi IDS:n X.509-sertifikaatti), jotka provisioidaan ja liitetään jokaiseen ekosysteemin komponenttiin automaattista autentikointia sekä luvitusta varten. Provisioinnin suorittaa ekosysteemin sertifioinneista vastaava taho. Kun digitaalinen sertifikaatti on käyttöönotettu komponentissa, voidaan komponentti rekisteröidä DAPS:iin.

Dynaaminen attribuuttien provisiointipalvelu (DAPS)

Dynaaminen attribuuttien provisiointipalvelu eli DAPS (Dynamic Attribute Provisioning Service) on yksi IDS:n ja Gaia-X-arkkitehtuurimallien keskeisimmistä luottamuksenhallintakomponenteista, jota käytetään tässä esimerkkinä. DAPS on ekosysteemitasoinen palvelu, joka antaa digitaaliselle sertifikaatille yksilöivän tunnisteen sekä PKI-infrastruktuuriin perustuvan tietoturvan ekosysteemeissä myös julkisen avaimen. DAPS vaihtaa tietoa DTM:n kanssa komponenttien käyttäytymisestä, kuten haavoittuvuuksista ynnä muista tietoturvaan liittyvistä tiedoista. DAPS:in keskeinen tehtävä on hallita ekosysteemin komponenttien luottamustasoja ja tarvittaessa poistaa digitaaliset sertifikaatit epäluotettaviksi todetuilta komponenteilta.

Dynaaminen luottamuksen monitorointi (DTM)

Dynaaminen luottamuksen monitorointi (DTM) valvoo ekosysteemin toimijoita niiden luottamuksen seuraamiseksi ja luotettavuusluokittelun mahdollistamiseksi. DTM on niin ikään IDS:n ja Gaia-X:n arkkitehtuurimalleissa keskeinen luottamuksen hallinnan komponentti. Se tulee kytkeä jokaiseen ekosysteemin komponenttiin luottamuksen seuraamiseksi. DTM:ää hyödynnetään yhdessä DAPS:n kanssa viestimään datatransaktion osapuolelle toisen osapuolen nykyisestä luottamustasosta.

IDS-referenssiarkkitehtuuriin pitkälti pohjautuva Gaia-X määrittää identiteetinhallinnan samalla tavalla ja kuvaa federoidun identiteetinhallinnan toiminnan alla esitetyn kuvan mukaisesti. Gaia-X:n identiteetinhallinnasta voi lukea tarkemmin Gaia-X:n luvuista 2.6., 3.3. ja 3.4.

3.4.4. Sanastohubi ja ontologiapalvelut

Sanastohubiin tallennetaan ja ylläpidetään ekosysteemin sisällä tai toimijakohtaisesti käytettävät deklaratiiviset mallit sanastoista, eli sanastoskeemat. Sanastohubi muodostuu usein sanastopalvelimista, kuten Yhteentoimivuusalustan sanasto-ontologiapalvelusta, jotka mahdollistavat yhteiset sanastomallit. Sanastohubi voidaan kiinnittää ekosysteemin federoituun katalogiin ja tai ekosysteemille voidaan tarvittaessa luoda oma, esimerkiksi jos kunnalla on käytössään keskitettyjä sanastojenhallintaan kykeneviä järjestelmiä. Sanastohubeja ja niissä olevia sanastoja on hallittava erityisesti versioinnin osalta, jotta yhteentoimivuutta kyetään ylläpitämään.

Sanastohubeja voidaan hyödyntää connectorien ontologioiden tarkentamiseen, mutta myös tiedon transformoimiseen tarkoitettujen käännöstaulujen muodostamiseen yhdessä tietomalli- eli skeemanhallintapalvelujen kanssa. Tietomalleja voidaan luoda erilaisilla ohjelmistoilla, mutta kunnille suositellaan kuitenkin hyödyntämään Yhteentoimivuusalustan tietomallit-palvelua, jolla graafisesti luodut tietomallit voidaan kääntää suoraan haluttuun deklaratiiviseen muotoon, esimerkiksi JSON-LD-rakenteeseen.

3.4.5. Päivitystietovaranto

Päivitystietovarantoon talletetaan connectorien konfiguraatiot ja jonka kautta connectoreihin tehdään päivityksiä. Päivitystietovarannot ovat usein teknologiasidonnaisia ja riippuvat connectorien toteutuksista.

3.4.6. Lähde- ja kohdetietovarannot

Sekä lähde- että kohdetietovarannoista tulee huomioida niiden rajoitteet, erityisesti silloin, kun connectorit ovat osa tietovarannoista vastaavien järjestelmien arkkitehtuuria. Teknisesti vanhemmissa järjestelmissä saattaa olla vain rajoittuneet mahdollisuudet jakaa tietoja halutuissa formaateissa. Lisäksi tietokantojen tietomallit eivät aina sovellu tiedon jakamiseen. Tällöin on usein syytä hyödyntää erilaisia middleware-komponentteja tai mahdollisuuksien mukaan connectorien tukisovelluksia, jotka kykenevät prosessoimaan ja transformoimaan dataa jaettavaan muotoon.

Rakenteellisten puutteiden lisäksi tiedon sisällölliset virheet ja puutteet heikentävät datan laatua ja samalla sen arvoa. Hyödynnettävät tietovarannot kannattaakin valikoida siten, että niiden tieto on laadullisesti mahdollisimman hyvää. Lähdetietovarantojen tietosisällöt suositellaan käytävän läpi ennen connectorien rakentamista, jotta tiedon laadusta voidaan varmistua ja suorittaa tarvittavat korjaavat toimenpiteet.

3.4.7. Tiedon virtautuksen tukisovellusten käyttö

Tiedon virtautuksen tukisovellukset tarjoavat datapalveluita siirrettävän datan transformointiin ja prosessointiin. Datasovellukset voidaan liittää connectoreihin omina sovelluskontteinaan tai hyödyntää esimerkiksi integraatioalustan palvelujen kautta. Datasovellukset voivat olla datan lähettäjän connectorin sisään rakennettuja, kolmannen osapuolen tarjoamia tai datan vastaanottajan sovelluksia, joita voidaan hyödyntää myös datan tuottajaosapuolen connectorissa.

Datasovelluksista voidaan vielä erottaa:

- *Systeemiadapterit*, jotka muodostavat rajapinnat erilaisiin tietovaranto- ja järjestelmäratkaisuihin tiedon jakamiseksi datan tuottajan connectorista. Adapterit voivat sisältää toiminnallisuutta myös datan transformoimiseen lähtöjärjestelmän rakenteesta datatuotteen haluttuun skeemaan ja tarvittavien metadatojen lisäämiseen.
- *Älysovellukset*, jotka tarjoavat datan prosessointi-, transformointi- ja tallennustoiminnallisuuksia datan hyödyntäjäosapuolen connectoreille. Lähetetyn datatuotteen metatietoja voidaan hyödyntää älysovellusten ohjeistamiseen.
- *Muut datasovellukset*, jotka tarjoavat muita kuin edellä mainittuja toiminnallisuuksia datan käsittelyyn ja hyödyntämiseen sekä datan tuottajan että hyödyntäjän puolen connectoreissa. Esimerkkinä muusta datasovelluksesta voidaan käyttää käyttöehtojen valvontaan tarkoitettua sovellusta luotetussa ympäristössä.

Datasovelluksia voidaan tarjota ekosysteemin toimijoille keskitetystä sovelluskaupasta, joka muodostaa oman federoidun kataloginsa, josta ekosysteemin osallistujat tilaavat tarvitsemansa sovellukset. Kunnan sisällä voidaan hyödyntää IT-palveluhallintajärjestelmää (esimerkiksi ServiceNow ja Efecte) datasovellusten provisioimiseen. Usein myös erilaiset integraatio- ja luottamusalustat, joihin connectoreita rakennetaan mahdollistavat datasovellusten tai vastaavan toiminnallisuuden provisioimisen connectoreille.

4. Käyttötapauksen tiedon virtautuksen suunnittelu

Tämä luku tarkentaa vaiheistuksen yksittäisen käyttötapauksen ja sitä varten luotavan datatuotteen sekä tietovirran korkean tason suunnittelulle. Vaiheissa kuvattujen osa-alueiden kuvaaminen ja/tai huomioiminen on tärkeää, jotta kyetään varmistamaan suunnitellun käyttötapauksen toiminnan edellytykset, ja että käyttötapaukselle suunniteltu tekninen arkkitehtuuri vastaa kunnan tai muun dataekosysteemin vaatimuksia. Moni esitellyistä kohdista nivoutuu tiukasti yhteen, jolloin yhden asian tunnistus tai ratkaisu antaa vastauksen useaan kohtaan. Myös otsikoiden alla olevassa sisällössä on syytä muistaa, että kyseessä on korkean tason arkkitehtuurisuunnittelu. Tarkentamalla näitä samoja kuvauksia päästään myös hyvin yksityiskohtaiseen arkkitehtuurikuvaukseen.

Vaiheistus tarjoaakin otsikoineen muistilistan käyttötapauksen suunnittelussa huomioitavista asioista. Esimerkiksi jossain käyttötapauksessa voidaan suojauksen osalta todeta datan siirtämisen HTTPS-yhteyden yli ja perusautentikoinnilla olevan täysin riittävä, eikä siihen tarvita enempää. Tiedonsiirtoprotokollien ja tunnistamistarpeiden huomioiminen suunnittelussa varmistaa kuitenkin sen, että päätökset näistä on tehty tietoisesti, eivätkä ne ole unohtuneet suunnittelusta.

Käyttötapauksen korkean tason arkkitehtuuria suunniteltaessa on myös tärkeää varmistaa, ettei se ole ristiriidassa käyttötapauksen yksityiskohtaisempien vaatimusten tai tunnettujen teknisten käytännön haasteiden kanssa. Siksi mahdollisten teknisten haasteiden tunnistaminen tässä vaiheessa sujuvoittaa huomattavasti käyttötapauksen tarkempaa suunnittelua ja toteutusta, kun yllättävien teknisten esteiden määrä on saatu minimoitua.

4.1. Tietojen tunnistaminen ja määrittely

4.1.1. Käyttökohteen vähimmäisvaatimukset datalle

Suunnitellaan, mitä dataa vähintään tarvitaan, jotta käyttötapauksen tietovaatimukset saadaan täytettyä. Tietokenttätasolle saakka ei yleensä tarvitse vielä tässä kohtaa suunnittelua mennä, mutta tarvittavat tietoluokat sekä keskeiset muut elementit on hyvä tunnistaa jo tässä vaiheessa.

4.1.2. Käyttökohteen lisäarvodatan tunnistus

Tunnistetaan data, joka voisi tuoda käyttötapaukselle lisäarvoa heti tai myöhemmin. Yleensä käyttötapauksen tarvitsemaa dataa tunnistettaessa datalähteillä, huomataan siellä olevan tarjolla myös muita hyödyllisiä tietoja. Hyödylliseksi tunnistetun muun datan jakamisen mahdollistamiseen liittyvät tehtävät voidaan siirtää jatkokehitystehtäviksi, erityisesti silloin, jos niistä voidaan odottaa lisäarvoa tulevaisuudessa. Toisinaan nämä lisäarvoa tuottavat datat halutaan kuitenkin sisällyttää alkuperäisen käyttötapauksen sisältämiin datoihin, jotta ne saadaan nopeammin käyttöön ilman erillistä projektointia. Tämä ns. lisäarvodata voidaan ottaa huomioon datatuotetta suunniteltaessa ja tuoda valmiiksi datatuoteaihioon, josta se on käyttöönotettavissa nopeasti.

4.1.3. Datan välityksessä poistettavat datat

Kuntiin kohdistuvat tietoturva- ja suojavaatimukset sekä -säännökset määrittelevät miten ja mitä dataa voidaan siirtää. Osaa datoista ei myöskään saa yhdistellä, erityisesti tiettyjä kansalaisia koskevia tietoja. Datatuotetta suunniteltaessa tuleekin kiinnittää huomiota sen sisältämiin datoihin kohdistuviin rajoituksiin, jotka ohjaavat datan yhdistelyä, anonymisointitarpeita, tallennuspaikkaa sekä teknistä arkkitehtuuria.

4.1.4. Muut dataan liittyvät asiat, kuten datan elinkaari, datan maksullisuus, luottamuksen huomioiminen ja datan käyttöön liittyvät muut rajoitukset

Datatuotetta suunniteltaessa on syytä suunnitella sen elinkaari jo etukäteen. Suunnittelussa tulee kiinnittää huomioita muun muassa siihen, miten datatuotteen sisällön tai käyttöehtojen muutoksista tiedotetaan sen hyödyntäjille. Datatuotteelle suositellaan suunnittelemaan myös sen elinkaaren loppuvaiheet, millä voidaan välttää tietoturva- ja suojariskejä sekä vanhentuneen tiedon jakamista.

Mikäli datatuote on maksullinen hyödyke, tulee se ilmoittaa hinnoittelumalleineen datatuotteen itsekuvauksessa tai vähintään niin, että hinnoitteluperusteet tulevat hyödyntäjän näkyviin hyödyntämispäätöksen tueksi.

Datatuotteelle tulee myös luoda käyttöehdot ja -sopimuspohja, jotka kertovat, miten tuotetta voi hyödyntää ja mitä käytön rajoituksia sillä on. Tässä vaiheessa kannattaa myös suunnitella, miten varmistetaan hyödyntäjien luotettavuus. Samalla tulee kuvata muutkin itsekuvauksen osa-alueet (katso luku 3.4.).

4.2. Arkkitehtuurikomponenttien tunnistus

4.2.1. Datalähteet

Seuraavaksi suunnitellaan, mistä tietovarannoista sekä järjestelmistä tarvittava data saadaan kerättyä ja koostettua. Datalähteitä voi olla yksi tai useampia ja ne voivat kostua muun muassa siirtotiedostoista, tietokannoista, tietovarastoista tai rajapinnoista. Lisäksi tulee huomioida datan esikäsittelytarpeet ja pyrkiä toteuttamaan tarvittavat käsittelyt mahdollisimman lähellä datalähdettä tai datalähteessä itsessään. Mikäli varsinaiseen tietovarantoon tai sen sisältämiin tietoihin ei tarvitse tehdä muutoksia, voidaan datalähteeksi merkitä vain tietovarannon rajapinta. Tässä vaiheessa kuvataan myös vaatimukset sille, miten datatuotteen tarjoava connectori yhdistetään datalähteisiin ja miten datat tulisi yhdistellä.

4.2.2. Midware-komponentit

Datalähteiden jälkeen tarkennetaan, minkä midware-komponenttien tai alustojen kautta data toimitetaan hyödyntäjälle ja minne connectori rakennetaan. Tässä vaiheessa on hyvä tarkentaa myös syyt, miksi data kierrätetään midware-komponentin kautta. Keskeisiä midware-komponentteja voivat olla erilaiset luottamusalustat, jotka voivat tarjota välityspalveluita sekä connector-palveluita sekä erilaiset datapalvelukomponentit, integraatioalustat ja tiedostosiirtäjät.

Midware-komponentteja suunniteltaessa tulee suunnitella myös, missä palvelussa datatuote julkaistaan ja miten. Tällaisia palveluita voivat olla ekosysteemin alustoina toimivat luottamusratkaisut, kunnan oma rajapintakatalogi tai erillinen ekosysteemin federoitu katalogijärjestelmä. Joissain tapauksissa voidaan datatuotteen tiedot tuoda nähtäville myös verkkosivustojen tai wikien kautta.

4.2.3. Datan kohteet

Datan kohteet tarkentaa, minne dataa ollaan viemässä. Tässä vaiheessa on huomioitava ja tarkennettava datatuotteen hyödyntäjien tekniset mahdollisuudet datan löytämiseen ja hyödyntämiseen, jotta midware-komponenttien tarve voidaan tarkentaa ja data on saatavilla mahdollisimman helposti. Yleensä datatuotteen hyödyntäjä tiedetään jo etukäteen, mutta usein on syytä huomioida myös datatuotteen käytön mahdollinen laajeneminen entuudestaan tuntemattomille hyödyntäjille, sekä suuntaa antava ajatus heidän datatarpeistaan esimerkiksi suorituskykyvaatimuksia varten.

4.3. Tekniset päätökset

4.3.1. Käytettävät protokollat

Arkkitehtuurikomponentteja tunnistettaessa tulee suunnitella myös tavat ja protokollat, joilla data kuljetetaan eri komponenttien välillä. Tietovirtaa suunniteltaessa on päätettävä esimerkiksi siitä, että toimitetaanko datatuote JSON_LD-formaatissa REST-rajapintojen kautta vai onko datan välityksessä käytössä muita protokollia, ja tukevatko suunnitellut komponentit näitä protokollia. Tarkempia connectorien ja muiden komponenttien API-kuvauksia ei tässä vaiheessa suunnittelua vielä tarvita.

4.3.2. Tietoturvavaatimukset

Tietoturvavaatimuksissa määritellään, miten tieto ja tiedon liikkuminen suojataan vaaditulla tasolla. Tässä tulee huomioida erityisesti datan tallennusvaatimukset, esimerkiksi kryptauksien osalta sekä valittujen tiedonsiirtoprotokollien riittävyys sekä mahdolliset tarpeet suojauksen lisäämisestä, esimerkiksi hashit ja PKI.

4.3.3. Suorituskyky- ja kapasiteettivaatimukset

Suorituskyky- ja kapasiteettivaatimuksissa pyritään tarkentamaan arvio liikuteltavan datan volyymista. Ennakointi auttaa määrittämään mahdolliset skaalautuvuustarpeet ja -kyvykkyydet. Nämä vaatimukset tulee huomioida arkkitehtuurikomponentteja sekä siirtoteknologioita valittaessa siten, että suunnitellut tekniset komponentit pystyvät välittämään vaaditut datat vaaditulla nopeudella datatuotteen elinkaaren aikana. Tällä varmistetaan siis osaltaan suunniteltujen teknisten komponenttien sopivuus suunniteltuun käyttötarkoitukseen, mikä voi vähentää uudelleensuunnittelu ja -toteutustarpeita datatuotteen elinkaaren myöhemmissä vaiheissa.

4.3.4. Muut luottamus- ja suojausvaatimukset

Mikäli datatuotteen avulla tullaan siirtämään sensitiivistä ja/tai tietoturvaluokiteltua dataa, on erittäin tärkeää, että tämä seikka tunnistetaan jo suunnittelun aikana. Datan siirrossa käytettävien suojausta tarjoavien tiedonsiirtoprotokollien ja muiden suojausratkaisujen lisäksi tulee huomioida myös luottamusvaatimukset sekä luvitukset datatuotteen datan hyödyntämiseen. Tässä vaiheessa tuleekin suunnitella ja tarkistaa, onko ekosysteemiin määritetty riittävät suojaus- ja luottamusmekanismit, jotka estävät datan päätymistä vääriin käsiin. Tätä varten tulee huomioida ainakin vaadittava autentikointitaso (ja autentikointitasojen riittävyys), sekä suunnitella ja päättää, kuka ylläpitää tarvittavaa luvitusta.

Muita suojauksen kannalta tärkeitä huomioitavia asioita ovat muun muassa tietoturvahyökkäyksen riskit, niiden potentiaaliset laajuudet sekä vaikutukset, ja miten riskejä voidaan vähentää sekä mitä niiden toteutumistilanteessa tehdään. Näitä voidaan sitten peilata valittuihin arkkitehtuurikomponentteihin ja pohtia, tarvitseeko datatuotteelle suunniteltua arkkitehtuuria muuttaa jotenkin turvallisuuden lisäämiseksi.

4.4. Laitteistosuunnitelma

4.4.1. Laitteiston yleiskuvaus

Korkean tason arkkitehtuurikuvassa ei ole tarvetta mennä tarkkoihin laitteistospekseihin, mutta siinä on kuitenkin hyvä tunnistaa, minkälaista laitteistoa vaaditaan, voidaanko hyödyntää jo olemassa olevaa laitteistoa tai mistä sopiva laitteisto saadaan hankittua. Tässä kohdassa tulee myös päättää kuka luovuttaa laitteistoa käyttöön tai hankkii ne.

Laitteiston osalta on hyvä myös tunnistaa skaalautumismahdollisuudet, esimerkiksi kapasiteettia ajatellen, sekä skaalautumisen raja-arvot, kuten enimmäiskapasiteetti. Näin varmistetaan, ettei laitteiston osalta jouduta ongelmiin projektin aikana tai ylläpitovaiheessa.

4.4.2. Laitteiston sijainti

Laitteiston sijainti on joissain tapauksissa merkityksellistä siirrettäessä ja tallennettaessa esimerkiksi sensitiivistä tai tietoturvaluokiteltua dataa. Yleensä sijainti voidaan kirjata yksinkertaisesti, esimerkiksi: ”Yrityksen X tiloissa” tai ”Kunnan Azure-ympäristö”. Tässä tulee kuitenkin huomioida ekosysteemin tieto- ja tekniset vaatimukset muun muassa laitteiston sertifioinnille.

4.5. Ylläpito ja tuotanto

Ylläpidon- ja tuotannon suunnittelun ja huomioinnin tarve on hyvin datatuotekohtaista. Ylläpidon ja tuotannon suunnittelua ei kuitenkaan saa ylenkatsoa. Panostusta tarvitaan kuitenkin aina vähintään sen verran, että ylläpidolle annetaan riittävät valmiudet toteuttaa ylläpito vaaditulla tavalla ja alasajosta tiedetään ainakin ylätasolla, miten se tulee toteuttaa. Tässä tulee huomioida erityisesti kunnan tiedonhallintalain mukainen tiedonhallintamalli sekä mahdolliset tiedonohjaussuunnitelmat.

4.5.1. Laitteiston ja ohjelmiston ylläpito

Laitteiston ja ohjelmistojen ylläpidosta määritellään eri komponenttien ylläpitotavat ja -vastuut sekä kuka vastaa esimerkiksi päivityksistä datatuotteen käyttöönoton jälkeen. Datatuotteen toteuttaja ja ylläpidon hoitaja ovat usein eri toimijoita.

4.5.2. Monitorointi ja hälytysjärjestelmä

Eri datatuotteilla ja niiden kriittisyystasoilla on erilaisia vaatimuksia monitoroinnille ja hälytyksille. Hälytykset suositellaan asettamaan kaikille datatuotteille, jotta datan tai sen siirron ongelmat saadaan havaittua, tunnistettua ja ratkaistua mahdollisimman nopeasti. Hälytykset voivat perustua sekä yhteisesti sovittuihin toimintamalleihin että automaattisiin ilmoituksiin. Manuaaliset hälytykset voivat olla esimerkiksi muotoa: ”Kun datan vastaanottaja havaitsee, ettei data päivity, ilmoittaa hän siitä saman päivän aikana sähköpostitse datatuotteen omistajalle, joka käynnistää häiriöselvityksen”.

On tärkeää miettiä, onko tämä tapa riittävä vai tulisiko integraatioihin ja tietokantaprosesseihin liittää automaattisia hälytyksiä, jotka voivat havaita virheitä nopeasti suuristakin datamassoista.

On erityisen tärkeää tunnistaa jo suunnitteluvaiheessa, mitä toteutuksessa tarvitsee tehdä tätä osuutta varten.

4.5.3. Varmuuskopiot ja palautuminen vikatilanteissa

Laitteistoissa ja ohjelmistoissa ilmenee usein virheitä ja ongelmia ajan kuluessa. On hyvä olla jonkunlainen ajatus siitä, kuinka näissä tilanteissa reagoidaan ja kuka on vastuussa toiminnan palauttamisesta. Aina toimintaa ei kyetä palauttamaan suoraan, vaan joudutaan palaamaan johonkin vanhaan stabiiliin konfiguraatioon vikatilanteen ratkaisemiseksi. Tämän vuoksi datatuotteen tekniselle arkkitehtuurille on syytä määritellä vaatimukset ja suunnitelma, miten menetellä ongelmatilanteista palautumisessa. Toteutukselle tämä voi asettaa esimerkiksi vaatimuksia varmuuskopion asentamisesta, laitteistojen kahdentamisesta ja varajärjestelmien asentamisesta.

4.5.4. Tietotuotteiden alasajo

Datan elinkaaren viimeinen osuus, eli käytöstä poistaminen ja alasajo, jää usein suunnittelematta, mikä voi aiheuttaa tarpeettomia tietoturva- ja suojariskejä. Alasajolle kannattaa luoda proseduuri ja

määrittää vastuut alasajon toteutukselle, jotta datatuotteen jäädessä tarpeettomaksi, se ei jää näkyviin katalogeissa, mahdolliset välitallennukset puhdistetaan ja integraatiot lakkautetaan. Erityistä huomiota tulee kiinnittää tietosuojasäädösten alaisiin tietoihin, jotka tulee poistaa käytöstä, kun niitä ei enää tarvita.

5. Lähde- ja sidosarkkitehtuuriluettelo

- Gaia-X tekninen arkkitehtuuri, kesäkuu 2020, haettu osoitteesta: https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5
- IDS-referenssiarkkitehtuurimalli *(International Data Spaces Reference Architecture Model*), versio 3.0, huhtikuu 2019, haettu osoitteesta: https://internationaldataspaces.org//wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf
- Guidelines for Modeling with NGSI-LD / Industry Specification Group (ISG) Cross Cutting Context Information Management (CIM), ETSI Whitepaper 42, 1. Painos, maaliskuu 2021, haettu osoitteesta: https://www.etsi.org/committee/cim/?jjj=1654000503974https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp_42_NGSI_LD.pdf
- Julkisen hallinnon API-periaatteet, Julkisen hallinnon ICT, Valtiovarainministeriön julkaisuja 2022:12, haettu osoitteesta: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/163864/VM_2022_12.pdf?sequence=1&isAllowed=y
- OASC MIM:it *(Open and Agile Smart Cities Minimum Interoperability Mechanisms),* versio 1.0.1., haettu osoitteesta: https://mims.oascities.org
- Reilun datatalouden sääntökirja, Sitra, versio 1.3. fi, kesäkuu 2021, haettu osoitteesta: https://www.sitra.fi/app/uploads/2021/06/reilun-datatalouden-saantokirja.pdf
- Architectural Styles and the Design of Network-based Software Architectures, luku 5: Representational State Transfer (REST), University of California Donald Bren Institute for Computer Sciences, 2000, haettu osoitteesta: https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm
- Tiedon jakamisen toimintamalli, Avoindata.fi, päivitetty 22.3.2022, haettu osoitteesta: https://www.avoindata.fi/fi/toimintamalli