

Helsingin datastrategia

Helsinki

Sisällysluettelo

Tiivistelmä ja päätösoisuus	3
1. Datastrategian lähtökohdat	6
1.1. Käsitteistö	6
1.2. Datan hyödyntämisen kohderyhmät ja käyttötapaukset	6
1.3. Datan hyödyntämisen nykytila	8
1.4. Datan jakamisen vaikeudet	9
2. Kaupunki data-alustana	9
2.1. Kaupunkia ympäröivä dataekosysteemi	9
2.2. Kaupungin yhteinen data-alusta	11
2.3. API-linjaukset ja muut tukevat rajaresurssit	11
3. Tietovarannot ja ydintiedon hallinta	13
3.1. Tietovarantojen hallinnan nykytila	13
3.2. Tiedonhallintalain velvoittamat toimenpiteet ja tietovarantojen hallinta	13
3.3. FAIR - tietovarantojen kaupunki-yhteinen hallintamalli	14
3.4. Ydintiedon hallinta	15
3.5. Helsinki-profiili (MyData-periaatteet)	16
3.6. Tietovastaavien työryhmä	16
4. Datan luokittelu	17
4.1. Luokittelu tietosisällön mukaan	17
4.2. Luokittelu datan käytön mukaan	19
4.3. Luokittelu datan käsittelysääntöjen näkökulmasta	20
5. Datan jakamisen ja hyödyntämisen prosessi	21
5.1. Tiedon hyödyntämiseen liittyvä lainsäädäntö	21
5.2. Tiedon hyödyntämisen työryhmä	22
5.3. Käyttötapauksen käsittelyn prosessi	23
5.4. Datan hyödyntäminen ja jakaminen tutkimushankkeissa	24
5.5. Eettiset periaatteet ja algoritmirekisteri	24
6. Tavoitearkkitehtuuri ja teknologiat	24
6.1. Tavoitearkkitehtuuri	25
6.2. Data-alustan palvelumalli	26
6.3. Dataputki ja käytetyt teknologiat	26
6.4. Käyttötapauksen toteutus data-alustalla	27
7. Kompetenssit ja organisoituminen	28
Käsitteitä	30

Tiivistelmä ja päätösosuus

Helsingin kaupungin johtoryhmä hyväksyi kaupungin digitalisaatio-ohjelman tavoitteet ja linjaukset maaliskuussa 2019. Ohjelma kuvaa, miten kaupunkistrategian visio maailman toimivin ja parhaiten digitalisaatiota hyödyntävä kaupunki saavutetaan digitalisaation avulla. Data-, tekoäly- ja robotisointikyvykkyyksien kehittäminen on yksi keskeisiä digitalisaatio-ohjelman tavoitteita, jotka mahdollistavat palvelutarvetta ennakoivan kaupungin ihmisten ehdoilla. Data on mahdollistajana kaikille digitalisaatio-ohjelman strategisille tavoitteille (ks. Kuva 1). Digiloikka ei ole mahdollista ilman datan ja tekoälyn tehokasta hyödyntämistä.

Datavetoinen kaupunki kuvataan digitalisaatio-ohjelmassa seuraavasti:

”Kaupunki hyödyntää dataa kaikessa toiminnassaan. Datan avulla tehdään informoituja päätöksiä, ennustetaan eri toimenpiteiden vaikuttavuutta ja automatisoidaan päätöksentekoa. Dataa hyödynnetään myös palveluiden kehittämisessä, sekä sisäisen suoritus- ja uudistuskäytännön kehittämisessä.”

Kuva 1. Kaupunkistrategian visio ja sen mukaiset digitalisaatio-ohjelman strategiset tavoitteet

Kaupunkistrategia

Maailman toimivin ja parhaiten digitalisaatiota hyödyntävä kaupunki

Strategiset tavoitteet

Asiakasnäkökulma

Proaktiivinen ja kohdennettu asiakkaan palveleminen

Kaupunkilainen vaikuttaa siihen, miten hänestä kerättyä dataa hyödynnetään

Automaatio ja itsepalvelukanavat mahdollistavat palveluiden 24/7 saatavuuden ja skaalautumisen

Mahdollistajat

Datavetoinen kaupunki: Johtaminen, päätöksenteko ja palvelukehitys datan avulla

Digitaalinen kaksonen:
Helsinki tuottaa, hyödyntää ja jakaa maailman parasta avointa dataa

Kaupunki hyödyntää alustatalouden mahdollisuuksia, palvelut tuotetaan ekosysteemin avulla

Datastrategialla tavoitellaan seuraavia neljää muutosta ja hyötyjä:

1. Data mahdollistaa asiakkaiden palvelutarpeita ennakoivan kaupungin ihmisen ehdoilla

Kaupunki on olemassa kaupunkilaisia varten. Kaupungin tulee tavoitella jatkuvasti kaupunkilaisten parempaa palvelua. Kaupunkilaisille parempi palvelu näkyy kun kaupunkilaiset saavat yksilöllisiä, kohdennettuja palveluita ennakoivasti silloin kun he niitä tarvitsevat. Palvelujen ennakoiva kohdentaminen kaupunkilaiselle edellyttää analyttisten menetelmien käyttämistä laajasti. Lisäksi ihmisten tulee voida vaikuttaa siihen miten heidän dataansa hyödynnetään MyData-periaatteiden mukaisesti.

2. Kaupungin johtaminen ajantasaisen tiedon pohjalta

Sujuvuuden, laadun ja tehokkuuden saavuttamiseksi kaupungin tulee myös hyödyntää datasta jalostettavaa tietoa entistä laajemmin ja ajantasaisemmin kaupungin toiminnan johtamisessa. Tietoa kaupungin toiminnasta, asukkaista ja kehityksestä syntyy kaupungissa jatkuvasti eri tasoilla. Jo nyt raportit, tilastot ja tutkimukset antavat eri aihepiireistä tutkimuksellista, dataan pohjautuvaa tietoa päättäjille. Dataa ei kuitenkaan nykyisellään kyetä hyödyntämään päätöksenteossa ja operatiivisessa toiminnassa sillä laajuudella, jonka nykyaikaiset tietoaaltaat (data lake), tietovarastot (data warehouse) ja analytiikkamenetelmät mahdollistaisivat. Esimerkiksi kaupunkimallista muodostettu digitaalinen kaksonen mahdollistaisi kaupungin toiminnan tarkastelun ja simuloinnin eri perspektiivien kautta.

3. Kaupungin toiminnan ja resurssien optimointi datan avulla

Ajantasainen data ja edistynyt analytiikka, kuten koneoppiminen, dynaaminen optimointi ja

ennustavat mallit, mahdollistavat myös kaupungin työntekijöiden toiminnan ja resurssien optimoinnin ja ohjaamisen kokonaistaloudellisesti. Ohjelmistorobotiikan ja datan avulla sisäisiä prosesseja voidaan automatisoida ja saada toimimaan saumattomasti yhteen. Erottavana tekijänä edelliseen tiedolla johtamisen muutostavoitteeseen, jossa ylin ja operatiivinen johto operoivat tiedolla, tässä tavoitteessa pyritään resursseja optimoimaan automaattisesti datan ja algoritmien avulla.

4. Elinkeinoelämän vilkastuttaminen ja ulkopuolisten resurssien hyödyntäminen dataa jakamalla

Sisäisen toiminnan lisäksi, kaupungin hallinnassa olevaa dataa pitäisi voida jakaa ja hyödyntää alustamaisesti ulkopuolisen ekosysteemin, kuten yhteisöjen, yliopistojen ja yritysten, käyttöön. Ulkoiset toimijat voivat tehdä tutkimusta ja kehittää palveluja, joihin kaupungin omaa tarjoamaa ei ole olemassa tai se on puutteellista.

Näiden muutosten edistämiseksi, datastrategia asettaa seuraavan datavision:

Helsingin tuottama data on maailman käytettävintä ja käytetyintä kaupunki-dataa vuoteen 2025 mennessä

Kaupungilla on huomattavia mahdollisuuksia parantaa datan hyödyntämistä. Kaupungin tietojärjestelmissä ja ydintiedon (master data) hallinnassa on kehitettävää liittyen datan tehokkaaseen yhteiskäyttöön. Toimialojen ja kaupungin omistamien liikelaitosten ja konserniyhtiöiden välinen datan jakaminen ja yhdistäminen on tällä hetkellä mahdollisuuksiin nähden vähäistä. Tämä johtuu osin toimintakulttuurista ja osin yhteistyötä edistävien kaupunkiyhteisten pelisääntöjen puutteesta. Lisäksi iso osa kaupungin datasta on henkilötietoa, jolloin juridisten oikeutusten arviointi on vaikeaa.

Linjataan seuraavat yhdeksän periaatetta:

- I. Kaupungilla on käyttöoikeus kaikkeen toiminnassaan ja palveluissaan tuottamaansa dataan
- II. Kaupunkiyhteiset ydintiedot, kuten asiakasdata, tallennetaan vain kerran
- III. Kaikki data on käytettävissä API:en kautta ja se on koneluettavaa
- IV. Kaikki data on hyödynnettävissä kaupungin sisäisesti yli toimialarajojen ja liikelaitoksissa ellei lainsäädäntö sitä estä
- V. Kaupunki hyödyntää ulkopuolista dataa ja jakaa omaansa mahdollisimman avoimesti edistääkseen kaupunkiekosysteemin elinvoimaa
- VI. Ennakkotapauskäsittelyllä nopeutetaan ja varmistetaan lainmukainen ja eettinen datan hyödyntäminen ja jakaminen
- VII. Kaupunkiyhteiset data- ja analytiikka-alustat tukevat ja nopeuttavat toimialojen itsenäistä palvelutuotantoa
- VIII. Data- ja analytiikkakyvykkyksiä kehitetään toimialojen tarpeiden ja konkreettisten käyttötapusten kautta
- IX. Asiakas voi MyData-periaatteiden mukaisesti vaikuttaa siihen, miten hänestä kertynyttä dataa hyödynnetään

Ensimmäiset toimenpiteet (suluissa listattu periaatteet, joita toimenpide edistää):

1. Kootaan ja hankitaan kaupunkiyhteiset data- ja analytiikka-alustat eri käyttötarkoituksiin, jotka tarjoavat työkalut datojen yhdistämiseen ja edistyneen analytiikan monipuoliseen hyödyntämiseen mm.: (IV, V, VII)
 - Tietoallas ja analytiikka-alusta, API-hallinta
 - Tietovarasto, raportointi ja johdon työpöytäratkaisut
 - 3D-kaupunkimalli visualisointien, analyysien ja simulaatioiden tekemiseen
2. Perustetaan tarvittavat kaupunkiyhteiset työryhmät, jotka varmistavat osaltaan datastrategian toimeenpanon seurannan ja toimenpiteiden tarkentamisen:
 - Tiedon hyödyntämisen työryhmä: datan hyödyntämiseen liittyvien ennakkotapausten käsittely tietosuojaja- ja muun lainsäädännön mukaisesti (IV, VI)
 - Tietovastaavien työryhmä: edistää datan laatua, yhteentoimivuutta ja ydintietojen hallintaa sekä kaupunkitasoisesti että tietoaalueittain (II)
 - Ryhmät yhdessä ylläpitävät kaupunkiyhteistä datan luokittelumallia
3. Tarjotaan kaupungin asiakkaille palvelut seurata ja hallinnoida läpinäkyvästi henkilökohtaisen datan käsittelyä: (IX)
 - Helsinki-profiili ja suostumusten hallinta MyData-periaatteiden mukaisesti
 - Algoritmirekisteri, joka kuvaa algoritmien logiikan ja eettiset periaatteet
4. Tietojärjestelmien ja palveluiden hankintasopimuksiin sisällytetään seuraavat ehdot:
 - Kaupungilla on käyttöoikeus tuottamaansa ja hankkimaansa dataan (I)
 - Hyödynnettävät tai kehitettävät algoritmit noudattavat eettisiä periaatteita (VI)
 - Hankittavaan tai kehitettävään järjestelmään on aina pääsy rajapinnan kautta (III, IV, V)
5. Kehitetään kaupungin data- ja analytiikkakyvykkyyttä perustamalla kaupunginkanslian strategiasastolle data ja analytiikkatiimi, jonka tehtävänä on tukea kaupunkiorganisaatiota datavision saavuttamisessa (VII, VIII)
 - Rekrytoidaan uusilla vakansseilla omia datatieteilijöitä ja -insinöörejä
 - Tarvittaessa tuetaan tiimiä ulkopuolisilla konsulteilla
 - Koulutetaan kaupungin nykyistä henkilöstöä datan ja analytiikan osaamisessa

Tämän dokumentin loppuosassa avataan edellä kuvattuja tavoitteita, periaatteita ja toimenpiteitä tarkemmin. Datastrategian kannalta keskeinen digitaalinen toimintaympäristö muuttuu koko ajan, joten tarkempia kuvauksia tullaan jatkuvasti päivittämään ja siksi ne rajataan itse päätöksen ulkopuolelle.

1. Datastrategian lähtökohdat

1.1. Käsitteistö

Datastrategian kannalta keskeisiä käsitteitä, kuten data, informaatio ja tieto, tai vastaavasti tekoäly, koneoppiminen ja edistynyt analytiikka, käytetään monesti ristiin. Tässä dokumentissa käytetään näitä käsitteitä seuraavasti (kattavampi käsitteistö löytyy liitteestä)¹.

Data on tietojärjestelmiin tallennettua koneellisesti käsiteltävissä olevaa ”raakatietaa”. Data voi olla joko rakenteista dataa, jolla on metatiedon mukainen semanttinen merkitys, tai rakenteetonta dataa, jolla sitä ei ole. Informaatio on tulkittavissa olevaa rakenteista dataa, josta voidaan jalostaa tietoa. Tiedolla viitataan kollektiivisesti informaatiosta jalostettuun ja sisäistettyyn tietämykseen, ymmärrykseen (asioiden syitä ja yhteyksiä) ja viisauteen (perustuen laaja-alaiseen kokemukseen ja oppimiseen). Tieto sanaa käytetään kuitenkin datan sijasta vakiintuneissa suomenkielisissä yhdyssanoissa, kuten tietovarasto ja tietoaallas, vaikka näissä täsmällisesti ottaen viitataan dataan eikä tietoon.

Tekoäly termillä viitataan yleisesti autonomisiin, oppiviin ja ennustaviin algoritmeihin ja järjestelmiin (ml. robotit). Tekoälykäsitteen alla, koneoppimisella viitataan ratkaisuihin, joissa kone oppii datasta joko opetusdatan avulla (supervised) tai itseoppivasti (unsupervised) ja muuttaa itse toimintaansa annetun tavoitteen suuntaan saamansa palautteen perusteella. Edistynyt analytiikka termiä käytetään viittaamaan yleisemmin perinteistä tilastotiedettä laajempaan joukkoon dataa hyödyntäviä analyysimenetelmiä, kuten dynaaminen optimointi, simulaatiot ja ml. edellä kuvatun kaltaiset tekoälykäsitteen alle kuuluvat algoritmit.

1.2. Datan hyödyntämisen kohderyhmät ja käyttötapaukset

Kaupungin datakyykykkyysiin panostaminen tuottaa monia hyötyjä, kuten tehokasta ja laadullisesti parempaa tietoperustaa päätöksentekoon ja johtamiseen. Toiminnan tehostamisesta koituvat kustannussäästöt ylittävät pitkällä aikavälillä datakyykykkyysien parantamiseen liittyvät kustannukset.

Datasta saadaan enemmän arvoa, kun sitä käytetään mahdollisimman laajasti. Toisaalta, vaivalla kerätty, yhteentoimimaton ja käyttämätömäksi jäävä data on kalleinta dataa. Kaupungin kannalta on tärkeää, että kaupungin toimialojen ja ekosysteemin tuottamaa dataa hyödynnetään mahdollisimman reaaliaikaisesti, laajasti ja tehokkaasti kaupungin prosesseissa ja palvelutuotannossa. Raportoinnissa ja analytiikassa pyritään datan jatkuvaan, skaalattuun käyttöön. Jotta kaupungin tuottamista datavarannoista saadaan maksimaalinen hyöty, pyritään jatkossa lisäämään edistyneen analytiikan, kuten koneoppimisen, hyödyntämistä.

Datan hyödyntämisen pitää tuottaa arvoa laajasti eri kohderyhmille. Alla on kuvattu esimerkein, miten eri kohderyhmät voivat hyötyä datasta.

Kaupungin asiakkaat: Asiakaskokemuksen parantaminen

Kaupungin digitalisaatio-ohjelman yhtenä strategisena tavoitteena ovat proaktiiviset ja kohdennetut palvelut kaupunkilaisille. Esimerkkinä proaktiivisuudesta on käynnissä oleva esiopetuspaikan tarjoamisen pilotti, jossa kaupunkilaisen ei tarvitse anoa palvelua, vaan hänelle tarjotaan esiopetuspaikkaa automaattisesti omalta alueelta lapsen tullessa esiopetusikänsä.

Yksityisellä puolella palveluiden kohdentamiseen käytetään laajasti analyttisiä menetelmiä. Myös kaupunki voi hyödyntää dataa palveluidensa parempaan kohdentamiseen tietyissä raameissa. Monet kohdentamistapaukset perustuvat asiakkaan eksplisiittiseen suostumukseen siitä, että hän haluaa saada suosituksia ja ehdotuksia. Kirjastonkäyttäjälle voitaisiin antaa suosituksia kiinnostavista kirjoista hänen aikaisempaan lainaushistoriaansa perustuen ja italiankielen opiskelijalle tietoa italialaisen ruuan kokkauskursseista.

¹ Käsitteistö mukailee FINTO sanasto- ja ontologiapalvelun tietotermit osion (<https://finto.fi/tt/fi/>) määritelmiä (ks. Tieto ja data [koneluettava tieto] käsitelmallit)

Esimerkkinä asiakaskokemuksen parantamisesta on kasvatuksen ja koulutuksen toimialan hanke, jossa kehitetään oppimisanalytiikan sovelluksia. Hankkeessa tuodaan uusia pedagogisia menetelmiä ja työkaluja muun muassa oppimisen personointiin, oppimisen etenemisen seurantaan, automaattiseen ohjaukseen ja palautteenantoon sekä pedagogiseen johtamiseen. Oppimisen mahdolliset ongelmat voidaan myös tunnistaa varhaisessa vaiheessa ja vähentää siten opintojen keskeyttämistä ja syrjäytymistä.

Palvelujen yksilöllinen kohdentaminen voi liittyä myös haastavampiin kokonaisuuksiin, kuten analyttisen mallinnuksen perusteella laskennallisesti syrjäytymisvaarassa olevien tunnistamiseen ja palveluiden yksilölliseen kohdentamiseen heille. Profiloinnissa ja automaattisessa päätöksenteossa tietosuojakysymysten ratkaisu on erityisen tärkeää.

Kaupungin ja toimialojen johto: Tiedolla johtaminen

Datastrategian suositus on hyödyntää tiedolla johtamisessa mahdollisimman reaaliaikaisia tilannekuvia kaupungin toiminnoista ja ilmiöistä. Käytännössä tilannekuva voi tarkoittaa kaupungille sitä, että tiedetään, millaisia potilaita Helsingin terveysasemilla käy mihinkin aikaan päivästä, kuukaudesta ja vuodesta, kuinka paljon on asiakkaita jotka käyttävät sekä kirjasto- että teatteripalveluita tai miten katutyöt muuttavat liikenteen sujuvuutta.

Tilannekuva tuodaan päättäjille (sekä ylimmälle että operatiiviselle johdolle) jatkuvasti päivittyvillä visualisoinneilla ja raporteilla, joista nähdään helposti avainmittareiden kehitys sekä ennustetut tulevaisuuden suunnat. Tietoa voidaan hyödyntää mm. budjetoinnissa, tehtävien priorisoinnissa, toimenpiteiden vaikuttavuuden arvioinnissa ja resurssialokaatioissa.

Esimerkiksi kaupunkiympäristöön liittyen Helsingissä on rakennettu kolmiulotteinen semanttinen malli kaupungista ja sen päälle realistisen näköinen visualisointi. Kaupunkimalli mahdollistaa erilaisten vaihtoehtojen analysoinnin ja simuloinnin (esim. tuuli-, energia- ja liikennemallit) auttaen hahmottamaan, miten eri toimenpiteet vaikuttavat kaupungin toimintaan käytännön tasolla. Lupaavia hyödyntämisesimerkkejä ovat:

1. Hiilineutraali Helsinki 2035 tavoitteen edistäminen geoenergiapotentiaali-palvelulla sekä lämmitysenergian kulutusennusteella

2. Kaavoitushankkeiden ja tietoaineistojen esittely kaupunkimallilla
3. Palvelukartan aineistojen esittäminen kaupunkimallilla
4. Tuuli-, sade- ja valumavesianalyysit sekä merenpinnan vaihtelun simulointi

Kaupungin palveluja tuottava henkilöstö: Tuotannon optimointi

Kaupunki operoi lukuisia palveluja ja prosesseja, joiden operoinnista myös muodostuu valtaosa kaupungin kuluista. Optimoimalla palvelujen tuotantoa on mahdollista saada aikaan merkittäviä säästöjä, parantaa palvelujen laatua ja helpottaa työntekijöiden työkuormaa.

Kaikissa kaupungin toiminnoissa käytetään dataa palveluiden tuottamiseen. Operatiivisissa järjestelmissä olevan datan on tärkeää olla korkealaatuista: laadukasta, ajantasaista, totuudenmukaista, sisäisesti yhtenäistä, ihmis- ja koneluettavaa sekä yhdistettävissä muihin datoihin. Data tulee olla myös helposti käytettävissä ja päivitettävissä. Tavoite on se, että prosesseja tukevilla käyttöliittymillä ja rajapinnoilla pystytään käsittelemään työvaiheisiin tarvittavia dataa samanaikaisesti siten, että työntekijöiltä ei kulu aikaa operatiivisten tietojen etsiskelyyn ja kokoamiseen useista eri järjestelmistä. Tämä koskee niin kaupungin sisäisten kuin ulkoisten toimijoiden (esim. Digi- ja väestötietoviraston VTJ-tiedot) tarjoamia dataa.

Palveluiden ja prosessien päivittäisen operoinnin lisäksi dataa tullaan hyödyntämään kaupungin palveluiden ja prosessien optimoimiseen ja kehittämiseen. Tätä työtä ei tehdä operatiivisissa järjestelmissä, vaan datat luetaan rajapintojen kautta data-alustalle, jossa prosessien tuottamia dataa voidaan tarkastella massoina. Tällöin ei esimerkiksi katsota yhden asiakkaan tietoja, vaan dataa käytetään ja yhdistetään pidemmältä aikaväliltä mahdollisimman monesta lähteestä.

Hyvä esimerkki on energiatehokkuuden optimointi kaupungissa, joka tukee kaupunkistrategian keskeistä ilmastotavoitetta, jonka mukaan Helsingistä tulee hiilineutraali vuoteen 2035 mennessä. Hiilineutraaliksi päästään vähentämällä kasvihuonekaasupäästöjä vähintään 80 % vuodesta 1990. Keskeisiä keinoja pienentää päästöjä on kestävä liikenteen kehittäminen ja rakennusten energiatehokkuuden parantaminen. Liikenteen sekä energiakäytön optimointiin liittyy oleellisesti datan kerääminen ja analysointi eri tasoilla.

Esimerkiksi kaupungissa toimivat yritykset, kaupunki ja valtio pystyvät pienentämään energiankulutusta, kun käytössä on reaaliaikaiset ja tarkemmalla tasolla olevat tiedot energiankulutuksesta. Tiedon avulla organisaatiot pystyvät tarkemmin kohdistamaan korjausinvestointeja sekä ohjaamaan kulutusta.

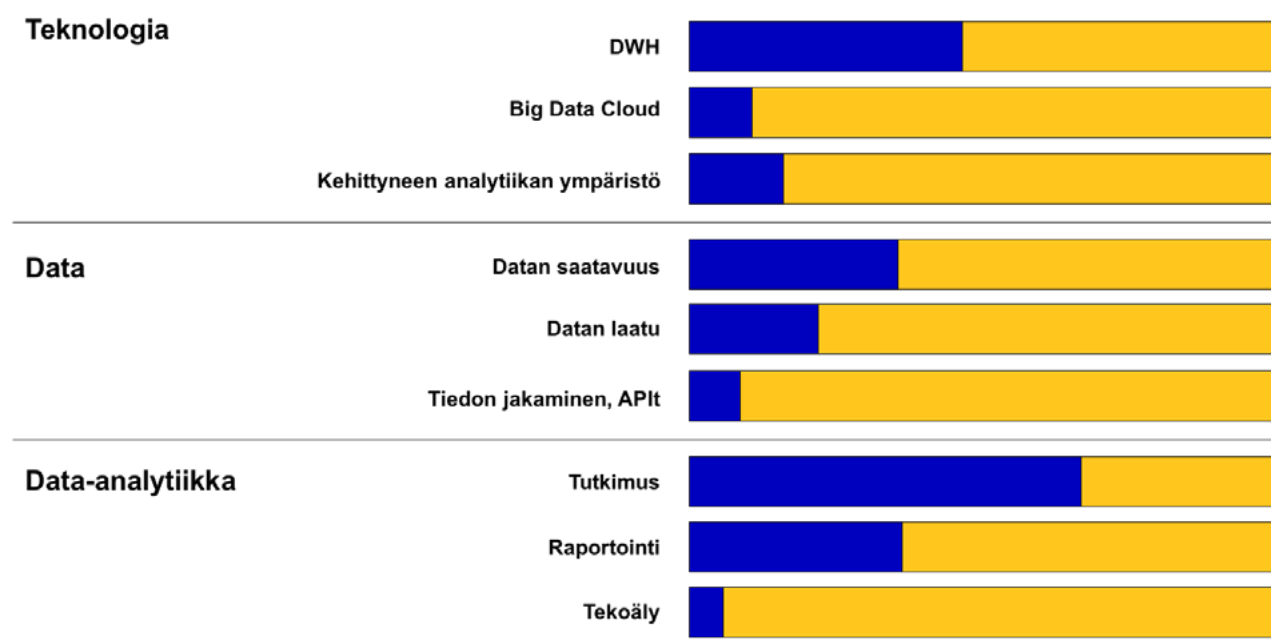
Vastaavasti kaupungin asukkailla on mahdollisuus vaikuttaa oman kotitaloutensa energiankulutukseen, mutta ilman dataa ja analysoitua tietoa energiankulutuksen optimointi on kuluttajalle vaikeaa. Valmiiksi tuotteistetut analyysit esimerkiksi asuinrakennuksen lämmitysjärjestelmän tai käytössä olevien sähkölaitteiden energiatehokkuudesta helpottavat kuluttajan valintoja. Esimerkkinä palvelusta on Helenin Kiinteistövahti, jolla pystytään seuraamaan kiinteistön tasolla energiankulutusta.

1.3. Datan hyödyntämisen nykytila

Helsingin kaupunki on suuri monialakonserni, jonka palvelut koskettavat lähes jokaista kaupunkilaista päivittäin. Palveluista kertyy erilaista dataa, jota kerätään jo tänä päivänä talteen kaupungin tietojärjestelmiin ja tietovarastoihin.

Kaupungilla on sekä vahvuuksia että heikkouksia datan hyödyntämisessä. Kuva 1 esittää yhteenvedon kaupungin datan hyödyntämisen nykytilasta, niin teknologian, datan, kuin data-analytiikan näkökulmista, jotka seuraavaksi käydään kukin läpi.

Kuva 1. Yhteenvedo kaupungin datan hyödyntämisen nykytilasta
(Lähde: datastrategiaprojektin haastattelut ja asiantuntija-analyysi)



Teknologia: Kaupungin omat tietovarannot ovat suurilta osin operatiivisissa järjestelmissä. Keskitettyyn tietovarastoon, HETA, on kerätty mm. kaupunkiyhteiset talous-, HR- ja tilatiedot sekä suoritettietoja SOTE- ja KASKO-toimialoilta. HETA-järjestelmässä ei ole käytössä pilviratkaisuihin natiivisti sisältyviä massadatojen, reaaliaikaisen datan ja koneoppimismenetelmien käsittelyyn soveltuvia työvälineitä. HETAn käyttöä jatketaan toistaiseksi ja selvitetään, miten HETA integroidaan osaksi tulevaisuuden kokonaistavoitearkkitehtuuria.

Kaupungin tutkimus- ja tilastoyksikössä (KTT) ylläpidetään tilasto- ja tutkimusaineistoja sisältävää data-asemaa, jota käytetään datavarantojen hyödyntämiseen ja analysointiin. KTT hallinnoi myös SAS Office Analytics -ohjelmistoa, jossa on monipuoliset datanhallinta- ja analysointiominaisuudet. SAS-ohjelmisto on HETA-tietovaraston tavoin optimoitu rakenteellisen tiedon käsitteelyyn. Lisäksi SAS-ohjelmiston rajoitteena on kyky käsitellä vain oman ympäristön sisällä olevaa tietoa, mikä vaikeuttaa tiedon yhdistämistä ja jakamista.

Datan visualisoinnissa ja raportoinnissa on käytössä SAS Office Analytics. Lisäksi on jonkin verran käytetty SAP BO Dashboardia, Lumiraa ja PowerBI:tä sekä evaluoitu erilaisia muita moderneja työkaluja, kuten Tableau ja SAP Analytics Cloud. Paikkatietoanalytiikkaa käytetään hajanaisesti eri toiminnoissa.

Datastrategian suositus on seurata eri teknologioiden ja analytiikkaympäristöjen kehittymistä ja hyödyntää kuhunkin tilanteeseen parhaiten soveltuvia välineitä.

Data: Dataa on määrällisesti paljon ja suuri osa siitä on toimialojen operatiivisissa järjestelmissä. Lähdejärjestelmien data täyttää monessa tapauksessa alkuperäisen käyttötarpeen laatuvaatimukset, mutta puutteita on mm. yhteentoimivuudessa sekä laadussa ja yhtenäisyydessä alkuperäisestä käyttötarpeesta poikkeavissa käyttötapauksissa. Kaupunkiyhteisessä käytössä olevaa dataa on vähän ja tiedonhallinnan ja tiedon jakamisen prosesseissa on kehitettävää. Rajapintojen kehittäminen on vielä alkuvaiheessa eikä kaupunkiyhteistä harmonisointia ja jakelua datalle tehdä kovinkaan paljon.

Data-analytiikka: Kaupunkitutkimus ja -tilastot -yksiköllä on pitkä historia ja paljon tieto-osaaamista, mikä osaltaan auttaa ymmärtämään, kuinka tärkeää data on kaupungille. Tutkimustoiminta ja vaikuttavuuden arviointi on pääosin kertaluonteista, vaikkakin tiettyjä tutkimuksia toistetaan säännöllisesti. Myöskään tilastointi ei tuota jatkuvia, esim. päivittäin päivittyviä analyysijä. Tilastollinen seuranta tuottaa nykyisellään kuukausittain, neljännesvuosittain ja vuosittain päivittyvää tietoa, mutta sekään ei tuota monissa tilanteissa tarvittavaa reaaliaikaista tilannetietoa. Ennustavaa analyysiä ja skenaariotyöskentelyä tehdään mm. Liikenteen ja väestön kehitykseen liittyen. Tekoälyn menetelmiä on pienissä määrin hyödynnetty toimialojen pilottihankkeissa, mutta skaalaaminen kaupunkitasoksi vaatii merkittäviä panostuksia.

1.4. Datan jakamisen vaikeudet

Toimialojen välinen tiedon jakaminen, yhdistäminen ja hyödyntäminen on vähäistä ja tiedon jakaminen toimialarajojen yli koetaan vaikeaksi tai mahdottomaksi. Suurimmat syyt datan yhdistämisen vähäisyyteen liittyvät jakamista tukevien tietojärjestelmien puutteisiin, datan rakenteellisuuden ja semanttisen yhteentoimivuuteen liittyviin ongelmiin, yksilöiden tunnisteiden puutteisiin sekä lainsäädännön asettamiin rajoituksiin ja laillisten käsittelyperusteiden vaikeuden arviointiin.

Jakamista tukevien tietojärjestelmien puutteeseen liittyy itse järjestelmien lisäksi puuttuva kaupunkiyhteinen ohjeistus, ydintiedon hallinta ja tietoarkkitehtuuri. Näiden puuttuessa datan yhdistämiseen liittyvät kehityshankkeet toteute-

taan toimialojen välisinä tiedonsiirtoina, joissa tieto kopioidaan yhdestä järjestelmästä toisen toimialan järjestelmään vaihtelevin menetelmin. Syntyvä tiedonsiirtoyhteys on toimiva ratkaisu, mutta kopioitu tietovaranto hyödyttää yhteyden rakentamisen jälkeen vain yhden toimialan ja käyttötapauksen tarpeita. Tietoarkkitehtuuri tulisi suunnitella niin, että se sopii tiedon tallentamiseen ja jakamiseen kaupungin ja sidosryhmien kesken. Hyvin suunniteltu tietoarkkitehtuuri mahdollistaa myös jakamisen skaalautumisen.

Tiedon yhdistämisen ja laajemman hyödyntämisen arviointiin ei löydy selkeää prosessia, jossa juridiset näkökulmat voitaisiin arvioida. Tällainen prosessi ja ohjeistus on tehty vain EU:n yleisestä tietosuojasetuksesta (GDPR). Ohjeistuksesta vastaa kaupungin tietosuojatiimi. Kaupungin tuottamia palveluja ohjaa kuitenkin laajempi lainsäädäntö, jonka noudattaminen on toimialojen lakimiesten osaamisalueella. Tietojen yhdistämisen ja hyödyntämisen arviointi vaatii useamman toimialan lakimiesten työpanoksen, mikä hidastaa oikeutusten arviointia.

2. Kaupunki data-alustana

2.1. Kaupunkia ympäröivä dataekosysteemi

Ekosysteemi tarkoittaa joustavaa kokonaisuutta, jossa yhteisöt, ihmiset, palvelut ja teknologiat kytkeytyvät luontevan asiakastarpeen kautta yhteen. Tällaisessa ekosysteemissä kaikki toimijat täydentävät toistensa toimintaa ja hyötyvät yhteistyöstä. Kaupunkistrategian tavoitteena on tehdä Helsingistä maailman edistyksellisin julkisen sektorin ekosysteemi.

Kuva 2 hahmottelee Helsingin roolin data-alustana kaupungin ympärillä toimivassa ekosysteemissä. Kaupungin ekosysteemiin voidaan lukea seuraavat osapuolet:

- **Kaupunkilaiset, kaupungissa työskentelevät ja matkailijat**
- **Yritykset:** Palveluntarjoajat, startupit, suuryritykset
- **Yhteisöt:** MyData, Open Street Map, jne.
- **Toimialat:** KASKO, KYMP, KUVA, SOTE (+ KANSLIA)

- **Liikelaitokset ja virastot:** Palke, Talpa, Stara, Tyhe, HKL, Tarkastusvirasto
- **Konserniyhtiöt**, joissa kaupungilla määräysvalta, kuten Helen ja Forum Virium Helsinki
- **Muut viranomaiset ja kaupungit:** valtion virastot ja ministeriöt sekä muut kaupungit
- **Yliopistot ja tutkimuslaitokset:** Korkea-koulut, yliopistot

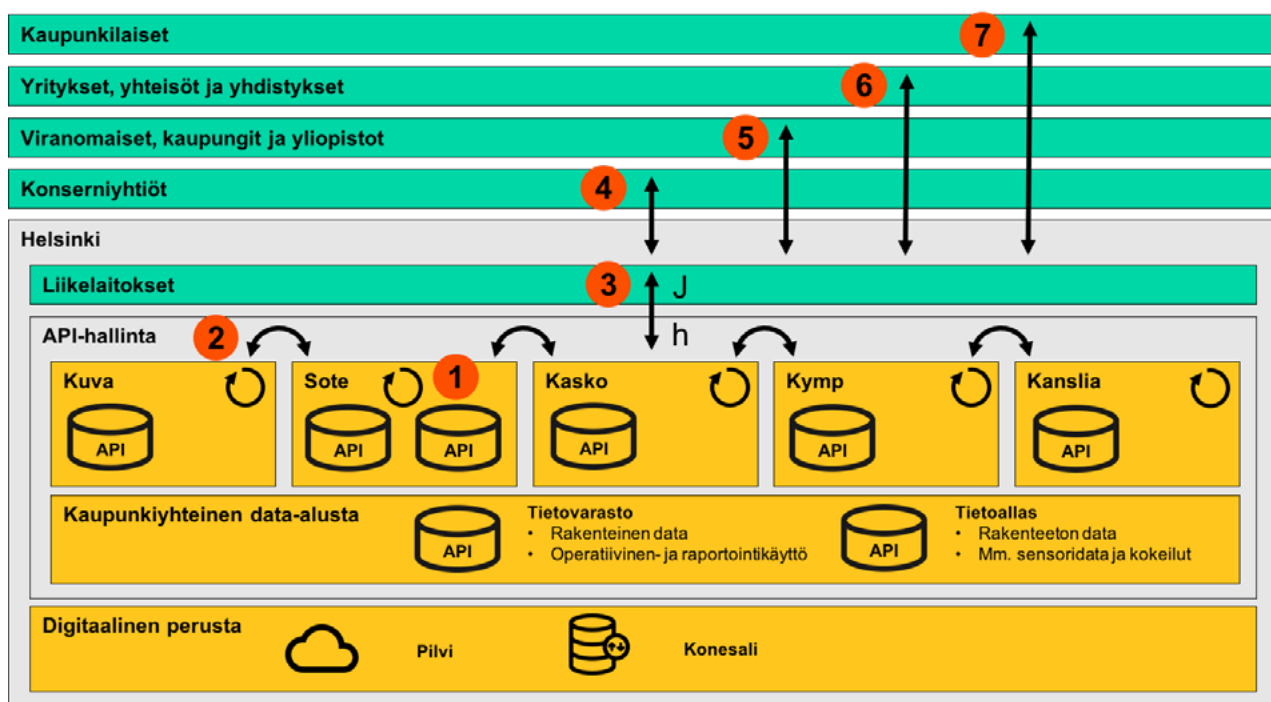
Kaupungin tavoitetilä on, että kaikki toimialat ovat tasavertaisesti mahdollisia datan tuottajia ja hyödyntäjiä yli toimialarajojen. KYMP:llä ja KANSLIA:lla on erityinen rooli kaupunkiyhteisen datan tuottamisessa ja jakamisessa kaikille toimialoille. Kuvan alalaidassa on hahmoteltu kaupunkiyhteistä data-alustaa ja sen pohjana on tietysti digitaalinen perusta ml. verkko, konesali ja pilviratkaisut.

Datastrategian laajempi tavoite on se, että datan avulla myös ulkopuolisille toimijoille syntyy mahdollisuuksia kehittää uusia palveluita ja innovaatioita. Kaupunki haluaa tarjota menestyvän alustan yritysten liiketoiminnalle ja muulle taloudelliselle aktiviteetille. Kaupungin tehtäviin kuuluu myös yhteisöllisyyden edellytysten lisääminen. Datan ja algoritmien lähdekoodin avaaminen sekä kehittämismenetelmien avoin tarjoaminen ekosysteemin käyttöön ovat oleellinen osa kaupungin datastrategiaa. Yritykset ja organisaatiot pääsevät käyttämään kaupungin

avointa dataa ja sovitussa raameissa kaupungin muutakin dataa sekä kehittämään palveluita kaupunkilaisille. Kaupunki haluaa tarjota dataa myös ei-kaupallisten tahojen kuten järjestöjen hyödynnettäväksi. Näin he voivat kehittää omaa toimintaansa ja palvella näin helsinkiläisiä sekä lisätä Helsingin houkuttelevuutta myös vierailijoiden keskuudessa. Kaupunki yleishyödyllisenä toimijana on hyvässä asemassa dataekosysteemin rakentamisessa erilaisille toimijoille siten, että kaikki tahot hyötyvät. Kaupunki pyrkii myös aktiivisesti itse hyödyntämään muualla tuotettua dataa.

Helsingin kaupungin dataekosysteemi tukee myös kaupungin rajat ylittäviä ekosysteemejä. Dataekosysteemit voivat olla Suomen tasoisia sekä myöhemmin Euroopan laajuisia. Euroopan komission datastrategiadokumentti² helmikuulta 2020 asettaa tavoitteeksi eurooppalaisten "datatilojen" (data spaces) rakentamisen määrittelemällä regulatorisen kehikon datanhallintaan, pääsyoikeuksiin ja datan hyödyntämiseen yritysten välillä, yritysten ja julkisen vallan välillä sekä eri julkisen vallan toimijoiden välillä. Helsingin kaupunki toimii Euroopan etujoukoissa regulatorisen datanhallintakehikon rakentamisessa ja elävän dataekosysteemin luomisessa.

Kuva 2. Helsinki data-alustana ympäröivässä ekosysteemissä



² European Commission: A European Strategy for Data. White Paper. 19.2.2020.

Kuvassa 2 on hahmoteltu numeroiduilla sinisillä nuolilla seitsemän eri tapaa jakaa ja hyödyntää dataa kaupungin sisällä ja ulkopuolisten tahojen kanssa. Jokaisella nuolella on kaksi suuntaa: kaupunki sekä jakaa omaa dataansa (j) että hyödyntää muiden tuottamaa dataa (h). Näihin eri tapoihin on pyritty tunnistamaan kaikki tavat, jotka joko teknisistä tai lainsäädännöllisistä syistä on syytä pitää erillään:

1. Toimialan sisällä

- Rekisterirajat rajoittavat tiedon jakamista myös toimialan sisällä
- Tiedonhallinta- ja tietosuojalainsäädäntö

2. Toimialojen välillä

- Tarvitaan kaupunkiyhteisiä ratkaisuja

3. Liikelayritysten kanssa

4. Konserniyhtiöiden kanssa

- Omistuksen kautta kaupungille strategisia kumppaneita

5. Viranomaisten, kaupunkien ja yliopistojen kanssa

- PSI-direktiivi, tiedonhallinta-, yms. lainsäädäntö
- Tietojen haku viranomaisten rekistereistä (Vtj, Ytj, Ktj)
- Tutkimuslupa

6. Yritysten, yhdistysten ja yhteisöjen kanssa

- PSI-direktiivi, kilpailu-, yms. lainsäädäntö
- Datan monetisointi

7. Kaupunkilaisten kanssa

- Tietosuojalainsäädäntö
- MyData-periaatteet

Helsingin kaupungin dataekosysteemi tukee myös kaupungin rajat ylittäviä ekosysteemejä. Dataekosysteemit voivat olla Suomen tasoisia sekä myöhemmin Euroopan laajuisia. Euroopan komission datastrategiadokumentti helmikuulta 2020 asettaa tavoitteeksi eurooppalaisten ”datatilojen” (data spaces) rakentamisen määrittelemällä regulatorisen kehikon datanhallintaan, pääsyoikeuksiin ja datan hyödyntämiseen yritysten välillä, yritysten ja julkisen vallan välillä sekä eri julkisen vallan toimijoiden välillä. Helsingin kaupunki toimii Euroopan etujoukoissa regulatorisen datanhallintakehikon rakentamisessa ja elävän dataekosysteemin luomisessa.

2.2. Kaupungin yhteinen data-alusta

Datastrategian yhtenä tavoitteena on se, että kaupunki toimii data-alustana, joka jakaa ja hyödyntää dataa mahdollisimman yhteneväisillä tavoilla niin sisäisesti kuin ulkopuolisten kanssa. Myös ulkopuoliset voivat toimia palveluntuot-

tajina ja innovaattoreina toimialojen tapaan. Kaupunki on siis alusta myös liiketoimintamielessä eikä vain teknisessä mielessä. Yhteisen data-alustan päätehtävänä on tukea ja innostaa toimialoja ja alustan ulkopuolisia osapuolia hyödyntämään ja jakamaan dataa keskenään. Yhdessä toimialojen kanssa haetaan tasapaino, jossa toisaalta toimialat voivat toimia itsenäisinä tuottajina, mutta samaan aikaan noudattavat yhteisiä alustan pelisääntöjä. Yhteisten pelisääntöjen määrä pyritään pitämään mahdollisimman pienenä keskittymällä olennaisiin ja lakien velvoittamiin linjauksiin.

Teknisen infrastruktuurin osalta toimialat ovat jossin tehneet omia ratkaisujaan. Kaupunki pyrkii tarjoamaan niin hyvän yhteisen tietovarasto ja tietoaallas ratkaisun, että pidemmällä aikavälillä myös infrastruktuurit olisivat mahdollisimman yhtenäisiä ja kustannustehokkaita. Näin ollen kaupungin data-alustan arkkitehtuuria voidaan kokonaisuutena tarkastella mikropalvelutyypin arkkitehtuurina, jossa yksittäiset toimialojen tietolähteet ovat erillisissä järjestelmissä ja kommunikoivat keskenään rajapintojen kautta. Data-alustan teknisiä ratkaisuja käydään tarkemmin läpi [luvussa 6](#).

2.3. API-linjaukset ja muut tukevat rajaresurssit

Jotta alustan ympärille syntyisi aktiivisesti toimiva ja arvoa tuottava ekosysteemi, alustan täytyy tukea tuottajia tarjoamalla niille niin kutsuttuja rajaresursseja. Rajaresursseilla tarkoitetaan alustan tuottajilleen tarjoamia rajapintoja, sääntöjä ja työkaluja. Alustassa täytyy olla avoimet rajapinnat, jotta tiedon jakaminen olisi ylipäänsä mahdollista. Säännöt määrittävät yhteiset toimintatavat, jotta alusta pysyy yhtenäisenä ja kolmansien osapuolien tuotokset laadukaina. Työkalut tukevat alustan tuottajia heidän työssään. On tärkeä huomata, että myös sisäiset toimialat ja liikelaitokset ovat tuottajia kaupungin data-alustassa.

Helsingin kaupungin data-alustan keskeisimmät rajaresurssit ovat:

- API:t (application programming interfaces, ohjelmointirajapinnat)
- API-linjaukset
- Datan jakamisen ja hyödyntämisen ennakkotapauspankki
- Tietovarantojen ja ydintietojen kuvaukset
- Helsinki-profiili

Datan jakaminen laajamittaisesti ei ole mahdollista, jos kaupunki ei avaa dataansa ja rajapintoja ulkopuolisille. Tämän vuoksi kaupunki on laatinut API-linjaukset, jotka velvoittavat tietojärjestelmähankinnoissa hankkimaan aina myös API:n ja avaamaan sen lainsäädännön mahdollistamissa puitteissa. Vaikka sisäisesti kaupunki periaatteessa pystyy hyödyntämään ja jakamaan dataa suoraan järjestelmistä ohi rajapintojen, jatkossa myös sisäisessä käytössä on ensisijaisesti käytettävä API-rajapintoja. Poikkeava toimintatapa on perusteltava.

API-rajapintoihin tulee suhtautua kuin tuotteisiin. Niiden tuoteomistaja on tyypillisesti toimiala tai liikelaitos, joka vastaa palvelusta tai tietojärjestelmästä. API:n hallintatyökalu on kaupunkitasoinen, jotta kaupungin sisäiset ja ulkopuoliset toimijat pystyvät löytämään API:t keskitetysti ja niiden käyttöä voidaan seurata kaupunkitasoisesti. Hallintatyökaluun kerätään API-rekisteri ja API:n käyttötiedot toimialojen tietojärjestelmistä. Avoimet API:t listataan ulkopuolisille selkeästi yhtenäisen palvelun, pääkaupunkiseudun kaupunkien yhteisen avoimen datan palvelun Helsingin Region Infosharen (HRI) kautta. Kaupungin API-linjaukset ovat datastrategian liitteenä.

Data-alustaa tukevana työkaluna ja sääntöko-

koelmana tarjotaan kaupungin sisäinen datan jakamisen ja hyödyntämisen ennakkotapauspankki, jota ylläpitää tiedon hyödyntämisen työryhmä. Työryhmä varmistaa, että datan jakaminen tapahtuu tiedonhallinta-, tietosuojaja muun datan jakamiseen liittyvän lainsäädännön mukaisesti. Ennakkotapausprosessia kuvataan tarkemmin [osiossa 5](#).

Jotta datan hyödyntäminen ekosysteemissä ja toimialojen välillä olisi mahdollista, tarvitaan myös yhteinen tietomalli eli kuvaukset eri tietojärjestelmien datamalleista. Tärkeimmässä roolissa on yli toimialojen käytetyn ydintietojen kuvaus. Tätä käydään tarkemmin läpi [osiossa 3.4](#). Ydintietojen hallinta.

Kaupunkilainen on Helsingin datastrategian keskiössä, ja siksi Helsinki noudattaa MyData-periaatteita. Vaikka kaupunkilaista koskeva data on kaupungin hallussa ja hallinnoimaa, kaupunkilainen voi itse päättää, miten hänen dataansa jaetaan ulkopuolisille. Periaatteiden mukaisesti kaupunki kehittää omaa Helsinki-profiilia, joka toimii myös yhtenäisenä asiakastunnuksena, ja jota voidaan siten myös hyödyntää (lain niin salliessa) tiedon yhdistämisessä eri rekisterien välillä. Helsinki-profiilia käsitellään tarkemmin [osassa 3.5](#).

3. Tietovarannot ja ydintiedon hallinta

Kaupungilla on käytössä yhteensä noin 900 lähdejärjestelmää, jossa kussakin operoidaan yhtä tai useampaa kaupungin tarjoamaa palvelua. Palveluissa ja niihin liittyvien järjestelmien kautta käsitellään kaupungin käytössä olevia dataa. Samaa prosessiin liittyvät tietojoukot muodostavat tietovarantoja, jotka ovat loogisia tietojen kokonaisuuksia. Esimerkiksi terveydenhuollon asiakkaiden tiedot muodostavat tietovarannon.

3.1. Tietovarantojen hallinnan nykytila

Järjestelmiin saatetaan tehdä tiedon päivityksiä hieman eri tavalla, käsittelyprosessit saattavat erota toisistaan ja työntekijät kirjaavat tietoja mahdollisesti eri tavoilla. Mm. näistä syistä datan oikeellisuus ja eheys voivat kärsiä, mikä vaikuttaa datan laatuun.

Datan laatu muodostuu ongelmalliseksi, kun tietovarantoja pyritään hyödyntämään yhdistämällä niitä toisiinsa tarkoituksena tehdä raportteja, analyyseja tai mallinnusta. Yhdistäminen on usein vaikeaa tai jopa mahdotonta, mikäli tietoa yhdistävät tunnisteet, yhteiset käsitteet tai tietomallit puuttuvat. Yhdistämisen vaikeus voi johtua myös datan huonosta yhteentoimivuudesta, kuten erilaisesta tavasta käsitellä ja käsitellä tietoa.

Kun datan laatuun panostetaan nykyistä enemmän, datan toissijainen käyttö saadaan lisääntymään tulevaisuudessa merkittävästi. Datan hyödyntämisen tulisi olla sujuvaa ja sen avulla tehostetaan tiedon avulla johtamista. Tiedontuotannon turhat toisteisuudet ja ylimääräiset vaiheet poistetaan mahdollistamalla yhteisen tietopohjan käyttö. Samalla edesautetaan mm. tilastointia, tutkimuksen, kehittämis- ja innovaatiotoiminnan, päätöksenteon, viranomaisohjauksen ja -valvonnan sekä lukuisten selvitysten, suunnittelutehtävien ja ennusteiden tarvitseman tietopohjan laatua.

3.2. Tiedonhallintalain velvoittamat toimenpiteet ja tietovarantojen hallinta

Tietovarantojen hallintaan liittyy keskeisesti tiedonhallintalaki, joka astui voimaan 1.1.2020. Lain tarkoituksena on varmistaa viranomaisten

tietoaaineistojen yhdenmukainen ja laadukas hallinta sekä tietoturvallinen käsittely julkisuusperiaatteen toteuttamiseksi, mahdollistaa viranomaisten tietoaaineistojen turvallinen ja tehokas hyödyntäminen sekä edistää tietojärjestelmien ja tietovarantojen yhteentoimivuutta. Valtiovarainministeriön yhteyteen perustettu julkisen hallinnon tiedonhallintalautakunta arvioi ja ohjaa valtion ja kuntien viranomaisten tiedonhallinnan toteuttamista. Tiedonhallintalakiin ja muussa laissa säädettyyn tiedonhallintaan liittyvien tehtävien kokonaisohjauksesta vastaa Helsingin kaupungilla tiedonhallintaryhmä.

Tiedonhallintalaki edellyttää kaupungilta toimenpiteitä ja kartoituksia mm. seuraavista asioista:

- Tietovarantojen käyttäjät ja omistajat, tietosisältö erityisesti tiedon luottamuksellisuutta koskien, tiedot tietojen säilytysajoista, arkistoinnista ja tuhoamisesta, sekä selostus, miten tietovarantoja käsitellään tietoturvallisesti
- Tietovirtakartoitus, josta ilmenee viranomaisten välillä tapahtuva tiedonsiirto, tiedonsiirtotapa ja siirrettävä tietosisältö
- Tiedon yhteentoimivuuden parantaminen ministeriöiden antamia ohjeistuksia noudattaen
- Tiedot tietojen säilytysajoista, arkistoon siirtämisestä, arkistointitavasta ja arkistopaikasta tai tuhoamisesta

Kaupungin tiedonohjaussuunnitelma (TOS) on osa tiedonhallintalain edellyttämää tiedonhallintamallia. TOS ohjaa mm. siihen merkittyjen digitaalisten asiakirjatietojen elinkaarten hallintaa. Tiedonohjaussuunnitelmalla ei hallita tietojärjestelmien tietoaaineistoja tietoyksikö-tasoisesti. Sen sijaan TOS:n tulee sisältää eri hakuperusteilla tietoyksiköistä muodostettavat digitaaliset asiakirjat kuten tilastot, tutkimukset, raportit ym.

Lain tarkoitukset ovat osittain samoja, mitä datastrategiassa tavoitellaan tietovarantojen hallinnassa, jakamisessa ja hyödyntämisessä. Tietovarantojen hallinnan kannalta onkin tärkeää, että tiedonhallintalain edellyttämissä kartoituksissa ja toimenpiteissä otetaan huomioon myös datastrategia. Lain vaatimusten lisäksi toteutetaan tarvittaessa toimenpiteitä, jotka edistävät yleisesti kaupungin datan hyödyntämistä. Yhtenäiseen tietomalliin nojaavaan, eri toimijoiden yhdessä muodostamaan dataan, on mahdollista rakentaa modulaarisesti hallitut palvelut ja silti antaa asiakkaalle yhtenäinen palvelukokemus.

3.3. FAIR - tietovarantojen kaupunki-yhteinen hallintamalli

Kaupungin tietovarantoja käsitellään EU-komission Horizon 2020 -ohjelman FAIR-mallin mukaisesti. Malli määrittelee alla esitellyt periaatteet reiluun ja tehokkaaseen datan hyödyntämiseen organisaatioissa. Näiden teknisempien periaatteiden lisäksi [osiossa 2.3](#) listataan muita rajaresursseja, jotka tukevat ja kiihdyttävät datan hyödyntämistä.

- **Löydettävissä (Findable)** Datan käyttäjä löytää helposti tarvitsemansa tiedon. Datan löytämistä varten käytössä on datakatalogi, erityinen hakupalvelu, joka sisältää tietoa tiedosta eli metatietoa. Metatieto sisältää datan löydettävyyden ja hyödyntämisen kannalta oleelliset tiedot, kuten mistä data löytyy, pysyvän tunnuksen dataan, sekä tietoa datan sisällöstä ja mahdolliset rajoitteet datan hyödyntämisessä.
- **Saavutettava (Accessible)** Data tai sen metatieto on noudettavissa tunnisteiden avulla standardia yhteyskäytäntöä noudattaen. Yh-

teyskäytäntö on avoin, maksuton ja yleisesti käyttöön otettavissa ja se tukee tunnistautumista ja valtuuttamista, kun se on tarpeen.

- **Yhteentoimivaa (Interoperable)** Kun datassa käytetään yhteisesti sovittuja tunnisteita, käsitteitä ja ontologioita, datan yhdistettävyyden muiden datojen kanssa helpottuu. Tiedon yhteentoimivuuden parantamiseksi siirrytään käyttämään kansallisia käsitteitä sitä mukaa, kun määritteet valmistuvat. Datan yhteentoimivuutta tavoitellaan kaupungin sisäisen yhteentoimivuuden lisäksi myös kaupungin ulkopuolisten toimijoiden kanssa, kansallisella tasolla ja mahdollisesti myös EU-tasolla.
- **Uudelleenkäytettävissä (Reusable)** Investointi datan hyödyntämiseen tehdään ajatellen datan uudelleenkäytettävyyttä, ei vain ensimmäistä käyttötarkoitusta. Data on uudelleenkäytettävää, kun datasisältö sisältää runsaasti tarkkoja ja hyödyllisiä attribuutteja, datan käyttöoikeuksissa on otettu huomioon monikäyttöisyys, selkeä käyttölisenssi on saatavilla ja data on aina yhdistettävissä alkuperäänsä.

Kuva 3. Tietovarantojen hallintamalli FAIR.



*Muokattu EU-komission Horizon2020-ohjelman FAIR-datamallista (2016)

3.4. Ydintiedon hallinta

Ydintieto on pysyvästi tarvittavaa perustietoa, jonka olemassaolo on välttämätöntä useiden prosessien ja toiminnan kannalta. Ydintiedon idea on, että se tieto, jota käsitellään ydintietona, ei ole organisaatiossa kiistanalaista, vaan koko organisaatiolla on yhteinen käsitys siitä, mistä tiedosta on kyse, mikä tiedon sisällön pitää olla ja mistä tieto löytyy. Yksiköt, jotka tarvitsevat ydintietoa, eivät muodosta datasta ylimääräisiä kopioita, vaan käyttävät ydintietoa suoraan sen lähteestä käsin.

Ydintiedon hallinta nostaa organisaation tehokkuutta sekä laatua tiedonkäsittelyssä. Parantunut tehokkuus ja laatu syntyvät ydintiedon oikeellisuudesta: oikea tieto löytyy helposti ja on käytössä läpi organisaation.

Kaupungin organisaatiossa ydintietoa on kahdella tasolla: toimialan sisäistä ja kaupungin yhteistä. Niitä voidaan kuvata seuraavasti:

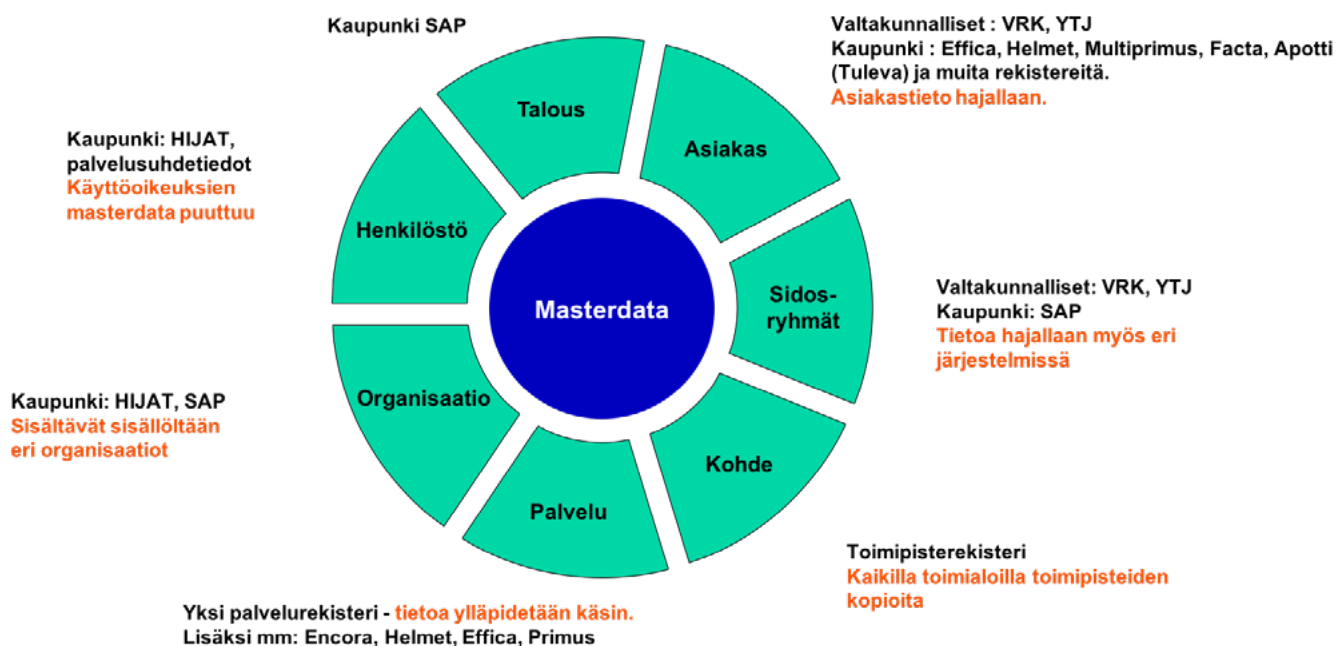
- Toimialan sisäiset ydintiedot palvelevat ensisijaisesti toimialan omia prosesseja ja tarpeita. Sisäistä ydintietoa voidaan jakaa tarvittaessa, mikäli datan jakaminen on mahdollista, toimialan ulkopuolelle, mutta sille ei ole jatkuvaa tarvetta kaupungin yhteisissä prosesseissa
- Kaupungin yhteiset ydintiedot ovat käytössä useissa kaupungin toiminnoissa

Erityisesti kaupungin yhteiset ydintiedot ovat kaupungin datastrategian kannalta merkityksellisiä, sillä ydintiedon avulla tietovarantoja pystytään yhdistämään luotettavasti ja tehokkaasti eri käyttötapauksia varten. Kaupungin ydintiedon on oltava kaupunkitason jaettua tai avuttua tietoa, jotta sitä voi vapaasti käyttää kaupungin sisäisesti.

Datastrategiassa on tunnistettu oheiset kaupungin ydintiedon osa-alueet (ks. kuva 5), joiden systemaattinen käyttäminen helpottaa merkittävästi toimialojen välistä tiedonvaihtoa, tiedon yhdistelyä sekä kaupunkitason toimintoja. Kaikille näille alueille kehitetään kaupunkitasoinen ydintieto.

- **Asiakas.** Kaupungin palveluita käyttävät tai tarvitsevat toimijat, kuten asukkaat, säännölliset vierailijat, turistit, yritykset ja yhteisöt sekä kaupungin sisäiset asiakkaat
- **Sidosryhmät.** Valtio, muut kunnat, toimittajat ja kumppanit, oppilaitokset, kolmas sektori
- **Organisaatio.** Organisaatorakenne, delegointirakenne, päätöksenteon rakenne
- **Palvelu.** Kaupungin tarjoamat fyysiset ja digitaaliset palvelut
- **Kohde.** Sisältäen korkeamman tason kohde-
luokat, kuten rakennukset, rakennetut kohteet, luontokohteet ja kaupungin palvelutilat
- **Henkilöstö.** Kaupungin henkilökunnan palvelusuhdetieto
- **Talous.** Kustannukset, tilit, budjetin seuranta-kohteet

Kuva 4. Tunnistetut kaupungin yhteiset ydintiedot ja ydintietoa sisältävät tietojärjestelmät. Merkittävät puutteet on merkitty punaisella.



Jotta ydintieto saadaan kaupungin yhteiseen käyttöön, on ensin määriteltävä datojen ja järjestelmien nykytila sekä osa-alueittain kartoittaa se tavoittila, johon halutaan päästä. Kartoituksen pohjalta määritetään tiekartta, jonka perusteella ydintiedon hallintaa kehitetään.

3.5. Helsinki-profiili (MyData-periaatteet)

Helsinki-profiili on käynnissä oleva hanke, jossa kehitetään kaupunkilaisen henkilökohtaista käyttäjätunnusta kaupungin digitaalisiin palveluihin. Helsinki-profiilin visiona on kaupungin asiakkaan tunnus, jonka avulla kaupungin asiakas voi yhdellä kirjautumisella asioida kaupungin palveluissa, hallita omia henkilökohtaisia tietojaan ja antaa suostumuksia tietojen jakamiseen myös muuhun käyttöön kuin ne on alun perin kerätty.

Helsinki-profiilin toiminta-ajatuksena on hakea asiakkaan perustiedot reaaliaikaisesti Digi- ja väestötietoviraston järjestelmästä, jossa on tallennettuna kaikkien kansalaisten nimet, henkilötunnukset ja pysyvät osoitteet. Helsinki-profiilissa olisi tällöin aina ajantasaiset perustiedot kaupungin toimijoiden käytettävissä. Helsinki-profiilissa asiakas voi täydentää tätä tietoa omilla sähköisillä yhteystiedoillaan ja toissijaisilla osoitteillaan. Helsinki-profiili mahdollistaa sen, että kaupunkilainen voi antaa tietyt avaintiedot (esim. puhelinnumeron, sähköpostiosoitteen) vain kerran, eikä hänen tarvitse syöttää niitä jokaiseen palveluun erikseen. Itse palvelut täydentävät tarvittaessa Helsinki-profiilin tietoja palvelukohtaisilla tiedoilla omissa tietojärjestelmissään.

Yhteisen käyttäjätunnuksen lisäksi Helsinki-profiili tarjoaa MyData-periaatteiden mukaisen suostumustenhallintamekanismin sekä pitää kirjaa muista datankäsittelyperusteista tietosuojasetuksen mukaisesti. Suostumustenhallinnan avulla asiakas voi halutessaan määrittää, mitä ja miten hänestä kerättyä tietoa kaupunki saa hyödyntää kaupungin ekosysteemissä. Pohjana kaikessa datan jakamisessa ja hyödyntämisessä on Suomen ja EU:n tietosuojalainsäädäntö.

3.6. Tietovastaavien työryhmä

Kaupungin ydintietojen harmonisointi ei tapahdu ilman järjestelmällistä yhteistyötä kaupungin eri toimijoiden kesken. Toimialojen datan laadun ja yhteisen käytettävyyden parantamiseksi perustetaan eri toimialojen tietovastaavista (esim. tietoarkkitehdit) koostuva työryhmä. Työryhmä vastaa ydintiedon määrittelyistä, yhteisten tunnisteen valinnasta keskeisille tietovarannoille sekä laatii yhteisiä datamääritelmiä, konseptuaalisia datamalleja, standardeja sekä ontologioita. Tietovastaavien työryhmä määrittelee ja valvoo FAIR-mallin toteutuksen edistymistä sekä pitää yllä tietovarantoon liittyvää dokumentaatiota ja ohjeita. Ryhmä vastaa myös kansallisten tiedon yhteentoimivuushankkeiden yhteistyöstä sekä tiedon semanttisen ja rakenteellisen yhteentoimivuuden toteutumisesta toimialoilla.

Käynnissä olevia kansallisia tiedon tehokkaampaan hyödyntämiseen tähtääviä hankkeita ovat mm. Yhteisen tietopohjan hyödyntämisen tietoarkkitehtuuri, #Tietokiri, Yhteentoimivuusalusta ja -menetelmät sekä AuroraAI.

4. Datan luokittelu

Datan luokittelu tarkoittaa tietoaineiston luokittamista ennalta määriteltuihin kategorioihin. Datan luokittelussa pyritään tunnistamaan tietojoukkoja, joihin pätevät yhteiset säännöt ja toimintatavat. Luokittelua voidaan hyödyntää esimerkiksi:

- Datan käytön linjauksissa kuten ”muuta kuin rajoitetuksi luokiteltua tietoa voidaan pääsääntöisesti käyttää kaupunkitasoisen tilannekuvan muodostamiseen”
- Datan käytettävyyden varmistamisessa. Sopimalla yhteiset metatietokäytännöt: käyttäjä tietää metatiedon perusteella, mistä lähteestä data on alun perin peräisin, milloin päivitetty ja miten käytettävissä
- Data-arkkitehtuurissa kuten ”sensoridata tallennetaan pääsääntöisesti kaupunkiyhteiseen pilviympäristöön”

Datan käsittely luokkien avulla tuo suoria hyötyjä datan hallintaan seuraavasti:

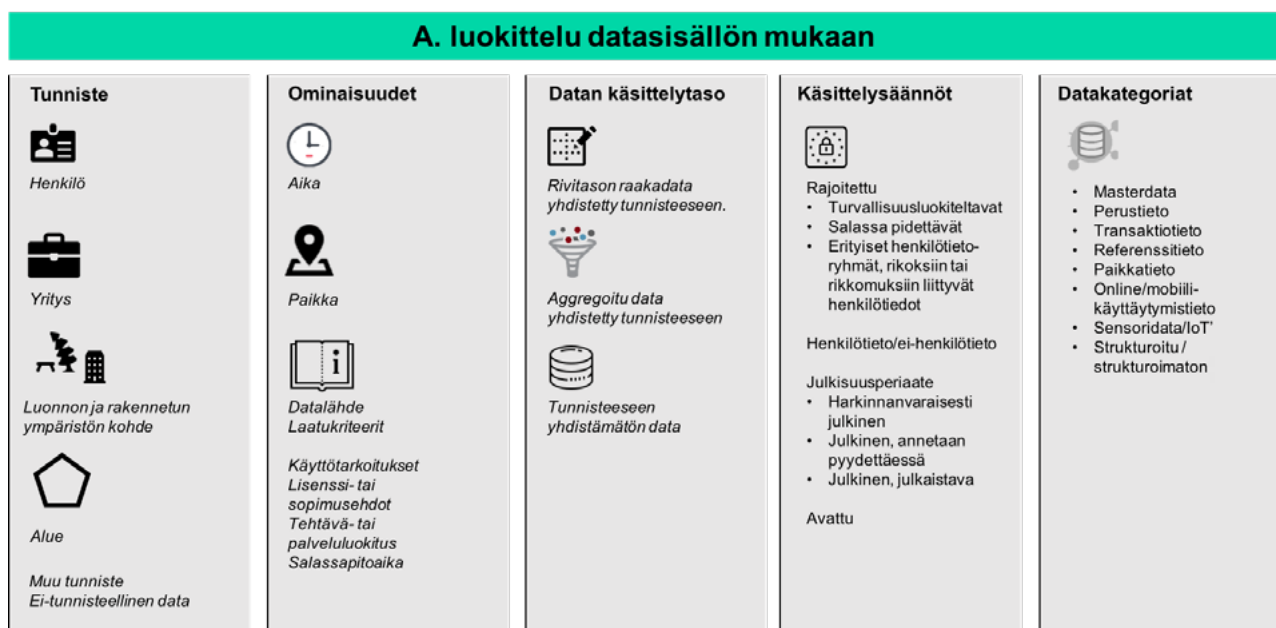
- Datan käsittely tehostuu, kun käsittelyä ohjataan prosesseilla ilman tarvetta tutustua tietosisältöön useaan kertaan

- Tiedonhallinnan virheet pienenevät automaattisen ja systemaattisen käsittelyn avulla
- Tietoturva paranee, koska oikeudet tietoon määritellään datan luokkien perusteella, eikä esimerkiksi erillisen prosessin tai harkinnan kautta
- Tiedon hyödyntämisen työryhmän työ helpottuu, kun käytettävien tietojen luokat on yksiselitteisesti määriteltä
- Tiedon hyödyntämistä on helpompi viestiä ja ennakkotapaukset löytää standardoidun datan luokittelun avulla, jolloin läpinäkyvyys tiedon hyödyntämiseen kasvaa

Datavision mukainen datan hyödyntäminen edellyttää datan luokittelua ja siihen liittyvien pelisääntöjen määrittelyä. Myös tietorakenteiden ja -järjestelmien pitää tukea näitä metatietoja yhdenmukaisella tavalla. Koska datan luokittelun ymmärtäminen on olennaista myös datastrategian kokonaisuuden ymmärtämiseksi, seuraavassa kappaleessa esitetään tämänhetkinen luokittelumalli taustatiedoksi. Luokittelumalli on jaettu kahteen osaan, tietosisältöön ja datan käyttöön perustuviin luokitteluihin. Nämä molemmat näkökulmat käydään läpi seuraavissa kahdessa alakappaleessa. Kappale 4.3 pureutuu tarkemmin käsittelysääntöihin.

4.1. Luokittelu tietosisällön mukaan

Kuva 5. Dataluokat tietosisällön mukaan



Tunnistedata	Määrittää, mihin tunnistetason data on linkitetty. Henkilötiedon osalta tunniste voi olla esimerkiksi Helsinki-profilin tunniste, henkilötunnus tai sähköpostiosoite, mutta myös laite tai eväste tunnistaa yksilön. Tunnistetaso voi olla myös yritys (esimerkiksi y-tunnus), luonnon tai rakennetun ympäristön kohde (kohdetunnus) tai alue (esimerkiksi postinumeroalue). Tunniste voi olla myös esimerkiksi verkkosivun url. Data voi olla myös ei-tunnisteellista (esim. tilasto).
Ominaisuudet	Tietoa, joka liittyy tietoon. Esimerkiksi tiedon lähde, keräysaika tai -paikka, käyttötarkoitus ja siihen liittyvät käsittelysäännöt. Kaupunkimallin kohdalla myös esimerkiksi mittakaava ja sijaintitarkkuus. Keskeinen osa metadataa, hyödyllistä datan käytössä sekä tiedon elinkaaren hallinnassa.
Datan käsittelytaso	Määrittää, millä tasolla dataa käsitellään. Käsittelytaso voi olla rivitason raakadata (esimerkiksi henkilön asiointihistoria tai sensorin lähettämä mittaustieto), tunnisteeseen yhdistymätön data (esimerkiksi tilastotieto) tai aggregoitu data yhdistetty tunnisteeseen (esimerkiksi asiakkaaseen liitetty tieto siitä, montako kertaa käyttänyt kulttuuripalveluja vuoden aikana, tai aluetason tilastotietoa).
Datan käsittelysäännöt	Kuvaa tietoa käsittelyn rajoitteiden näkökulmasta: onko kyseessä henkilötieto, erityiset henkilötietoryhmät, salassa pidettävä tai turvaluokiteltu tieto, vai onko tieto julkista tai avointa. Ks. kohta 4.3
Datakategoriat	Kuvaavat erityyppisiä tietovarantoja. Eri kategorioilla on erilaisia vaatimuksia datan käsittelyn suhteen. Esimerkiksi ydintiedon osalta tarvitaan prosessit varmistamaan yhteisen datan ylläpito ja käyttö keskitetystä lähteestä, kun taas online-käyttäytyminen ja anturidata voidaan halutessa tallentaa pilviympäristöön suurten datamassojen vuoksi. Samantyyppisten käytösääntöjen voidaan ajatella pääsääntöisesti pätevän saman kategorian datoille, joskin toimialan erityispiirteet ja muut luokittelukriteerit voivat tuoda tähän variaatioita. Dataa voi luokitella myös teemoittain, ks. esim. hri.fi, jossa avoimet datat on kategorisoitu JHS-suosituksen (JHS183) pohjalta.

4.2. Luokittelu datan käytön mukaan

Kuva 6. Dataluokat käytön mukaan



Käsittely ja säilytys

Kuua datan sijaintia: onko datan käsittelyalusta itse tuotettu vai palveluna hankittu, mihin data on tallennettu. Sijaintitiedolla on merkitystä datan käytön kannalta, sillä lainsäädäntö ja kaupungin omat linjaukset saattavat määrätä, missä dataa saa säilyttää.

Tiedon elinkaari

Tiedonohjaussuunnitelmaan merkittyjen tietoineistojen elinkaaren hallinta perustuu tiedonohjaussuunnitelmassa dokumentoituihin metatietoihin. Myös muulle tiedolle tulee määritellä säilytysaika: poistetaanko tieto heti palvelun tuottamisen jälkeen, liittykö siihen määräaikainen säilytysaika tai -velvollisuus, vai säilytetäänkö tietoa pysyvästi.

Käyttötapaukset

Käyttötapaukset liittyvät kiinteästi siihen, mihin tarkoitukseen dataa voidaan kerätä ja käyttää. Kaikissa datan käsittelyn linjauksissa käyttötapa on keskeisessä osassa. Käyttötapauksia on kuvattu tarkemmin tämän dokumentin alussa.

Juridinen toimintakehikko

Kaupunkitasolla datan hyödyntämisessä on huomioitava erilaiset lainsäädännön asettamat velvoitteet ja rajoitteet. Julkisuuslain lisäksi on huomioitava esimerkiksi tietosuojalaki, tiedonhallintalaki, digipalvelulaki, arkistolaki, PSI-direktiivi, laki sosiaali- ja terveystietojen toissijaisesta käytöstä (toisilaki) ja muu toimialakohtainen lainsäädäntö.

Käyttäjät ja käyttöoikeudet

Datan käyttöä hallinnoidaan viime kädessä käyttöoikeuksien kautta. Käyttäjäröhmien tunnistaminen auttaa linjaamaan, millä käyttöjäröhmillä on millainen pääsy minkä tyyppiseen dataan.

4.3. Luokittelu datan käsittelysääntöjen näkökulmasta

Kuva 7: Tiedon luokittelu käsittelysääntöjen näkökulmasta

Rajoitettu Salassa pidettävien tietojen käsittely on aina rajoitettua.
Turvaluokiteltavat tiedot ja asiakirjat ovat julkisuuslain nojalla salassa pidettäviä asiakirjoja, joissa salassa pidon perusteena on erityinen turvallisuuteen ja yleiseen etuun liittyvä syy. Tällaiset asiakirjat liittyvät muun muassa henkilöiden, rakennusten- ja tietojärjestelmien turvallisuusjärjestelyihin, poikkeusoloihin varautumiseen sekä maanpuolustukseen.
Salassa pidettävät tiedot ja asiakirjat: Tiedot ja asiakirjat jotka on määritetty julkisuuslaissa ja erityislaeissa salassa pidettäviksi.
Tietosuoja-asetuksessa on määritelty erityiset henkilötietoryhmät , joiden käsittelylle on asetetut tiukemmat perusteet. Erityisiä henkilötietoja ovat rotu, etninen alkuperä, poliittinen mielipide, uskonnollinen tai filosofinen vakaumus, ammattiliiton jäsenyys, geneettiset tai biometriset tiedot, terveyttä tai seksuaalista suuntautumista koskevat tiedot.
Henkilötieto. Henkilötiedolla tarkoitetaan kaikkia tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyviä tietoja. Henkilötietoja voivat olla muun muassa osoite, sähköpostiosoite, henkilötunnus, sijaintitieto, valokuva, ääni- tai videotallenne, auton rekisterinumero, sormenjälki tai muu biologinen näyte.
Julkinen (A. harkinnanvaraisesti. B. annetaan pyydettyäessä C. julkaistava) Kaupungin data joka lähtökohtaisesti saatavissa niin sisäiseen kuin ulkoiseen käyttöön, mutta dataa ei avattu yleiseen käyttöön rajapintojen kautta.
Avattu Kaupungin kaikille avoin data joka avattu rajapintojen kautta tai tiedostomuodossa yleiseen käyttöön
Huom. • Ulkoisten datalähteiden osalta sopimus voi rajoittaa datan käsittelyä silloinkin, kun data ei kuulu rajoitetun tiedon joukkoon. • Harkinnanvaraisesti julkisia tai julkisia datalähteitä yhdistelemällä on mahdollista tuottaa dataa, jotka ovat salassa pidettävää tai muuten rajoitettua tietoa (esim. yhdistelemällä tilastoja tavalla, josta yksilö on arvattavissa).

Julkisuuslain nojalla kaupungin keräämä tieto on lähtökohtaisesti julkista, ja kaupungin strategian mukaista on julkaista tietoa avoimena datana muiden toimijoiden käyttöön silloin, kun se on mahdollista. Kaupungin datavarantojen käsittelyä säätelevät kuitenkin monet lait, jotka rajoittavat tietojen käsittelyä. Tiedon luokittelua tästä näkökulmasta on havainnollistettu kuvassa 7.

Seuraavien tietojoukkojen käsittely on aina rajoitettua:

- **Turvaluokiteltavat tiedot ja asiakirjat** ovat julkisuuslain nojalla salassa pidettäviä asiakirjoja, joissa salassapidon perusteena on erityinen turvallisuuteen ja yleiseen etuun liittyvä syy. Tällaiset asiakirjat liittyvät muun muassa henkilöiden, rakennusten ja tietojärjestelmien turvallisuusjärjestelyihin, poikkeusoloihin varautumiseen sekä maanpuolustukseen.³

- **Salassa pidettävät tiedot ja asiakirjat:** Asiakirjat, jotka on määritetty julkisuuslaissa ja erityislaeissa salassa pidettäviksi. Julkisuuslain 24§ sisältää määrittelyn viranomaisten salassa pidettävistä asiakirjoista. Salassa pidettäviä tietoja voivat olla myös salattu puhelinnumero tai turvaosoite.
- Tietosuoja-asetuksessa (GDPR) on määritelty **erityiset henkilötietoryhmät**, joiden käsittelylle on asetettu tiukemmat perusteet. Erityiset henkilötietoryhmät vastaavat pitkälti aikaisemman henkilötietolain arkaluonteisia tietoja. Erityisiä henkilötietoja ovat rotu, etninen alkuperä, poliittinen mielipide, uskonnollinen tai filosofinen vakaumus, ammattiliiton jäsenyys, geneettiset tai biometriset tiedot, joita käytetään henkilön tunnistamiseksi, terveyttä koskevat tiedot sekä seksuaalista käyttäytymistä ja suuntautumista koskevat tiedot.

³ Turvallisuusluokittelua koskevat tiedonhallintalain säännökset (18 §) koskevat valtionhallintoa, mutta eivät kuntaa. Kaupunki selvittää parhaillaan mahdollisuuksia soveltaa turvallisuusluokittelua myös kaupungilla. Asiaa on tiedusteltu julkisen hallinnon tiedonhallintalautakunnan virkamiehiltä, mutta kanta on ollut kaupungille kielteinen ja kaupungille on todettu, että kunnilla ei ole toimivaltaa päättää turvallisuusluokittelusta. Salassapitomerkitöiden tekemisestä on säädetty erikseen julkisuuslaissa ja kaupunki voi tehdä julkisuuslain puitteissa salassapitomerkitöitä. Turvallisuusluokittelun käyttöönoton osalta on syytä odottaa selvitysprosessin tuloksia.

Henkilötiedon käsittely on myös rajoitettua ja sitä säätelee EU:n tietosuoja-asetus sekä tietosuojalaki

- **Henkilötiedolla** tarkoitetaan kaikkia tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyviä tietoja. Tunnistettavissa olevana pidetään henkilöä, joka voidaan tunnistaa suoraan tai epäsuorasti erityisesti tunnistetietojen kuten nimen, henkilötunnuksen, sijaintitiedon, verkkotunnistetietojen tai henkilölle tunnusomaisten fyysisen, fysiologisen, geneettisen, psyykkisen, taloudellisen, kulttuurillisen tai sosiaalisen tekijän perusteella.
- Henkilötietoja voivat olla muun muassa osoite, sähköpostiosoite, valokuva, ääni- tai videotallenne, auton rekisterinumero, sormenjälki tai muu biologinen näyte.

Lähtökohtaisesti kaupungin tuottama tieto on julkista ja saatavissa ulkoiseen tai sisäiseen käyttöön. Se voi tällöin olla

- **Harkinnanvaraisesti julkista.** Esimerkiksi valmistelussa olevat asiakirjat eivät välttämättä ole julkisia ennen valmistumistaan.
- **Julkista dataa, jonka saa pyydettyä** tai tekemällä sopimuksen datan käytöstä.
- **Julkista dataa, joka on julkaistavissa** avoimena datana.

Kaupungilla on myös dataa, joka on avattu rajapintojen kautta tai tiedostomuodossa vapaaseen käyttöön koneluettavassa muodossa. Avoin data on julkaistu JHS189-suosituksen mukaisella avoimella Creative Commons Nimeä 4.0 -lisenssillä. Muuhun kaupungin julkaisemaan dataan voi liittyä lisenssiin perustuvia käyttörajoituksia, joista myös pyritään luomaan kaupunkiyhteiset linjaukset.

Ulkoisten tietolähteiden osalta sopimus voi rajoittaa datan käsittelyä silloinkin, kun data ei kuulu rajoitetun tiedon joukkoon. Julkisia tietolähteitä yhdistelemällä on mahdollista tuottaa dataa, jotka ovat salassa pidettävää tai muuten rajoitettua tietoa (esim. yhdistelemällä tilastoja tavalla, josta yksilö on arvattavissa).

Tiedon luokittelun käyttöönoton edistämiseksi tiedon luokittelu tulee validoida testaamalla luokittelua käytännössä valituilla esimerkkiaineistoilla ja -käyttötapauksilla sekä vertaamalla muihin käytössä oleviin luokitteluihin (esim. tiedonohjaussuunnitelma, HRI:n ja kaupunkitutkimuksen datakatalogit, tietojärjestelmät ja

niihin liittyvät dokumentit ja ohjeistukset, ulkoiset luokitukset). Tämän pohjalta varmistetaan luokittelun yhdenmukaisuus sekä tarkennetaan käsitteiden määrittelyjä. Lisäksi tulee huolehtia koulutusten ja ohjeistusten avulla luokittelun käyttöönotosta kaupungin toiminnoissa ja prosesseissa (tiedonohjaussuunnitelma, tietojärjestelmiin liittyvät luokitukset ja järjestelmäsuunnittelu, kaupungin datatoiminnot).

5. Datan jakamisen ja hyödyntämisen prosessi

Datan hyödyntämistä kaupungilla ohjaa laaja lainsäädäntö, joka tekee tiedon hyödyntämisen laillisten perusteiden arvioinnista vaikeaa. Datan hyödyntämisen tehostamiseksi kehitetään datan jakamisen ja hyödyntämisen laillisten perusteiden arvioinnin menetelmiä siten, että arviointiprosessista tulee selkeä ja tehokas.

5.1. Tiedon hyödyntämiseen liittyvä lainsäädäntö

Tiedon hyödyntämisessä on huomioitava viranomaisten tiedonhallintaan liittyvä lainsäädäntö, jolla määritellään, miten kaupungin hallinnassa olevaa tietoa tulee käsitellä, julkaista ja hyödyntää. Tässä kappaleessa on kuvattu keskeistä lainsäädäntöä, joka liittyy kaupungin hallinnassa olevan tiedon hyödyntämiseen.

Laki viranomaisten toiminnan julkisuudesta (621/1999, **julkisuuslaki**) sääntelee viranomaisten asiakirjojen ja muiden tietoaineistojen julkisuutta ja salassapitoa sekä tiedon antamista asiakirjasta ja siihen liittyvää menettelyä. Laki sääntelee myös viranomaisten velvollisuutta toteuttaa hyvää tiedonhallintatapaa, johon sisältyy tietoturva vaatimuksia.

Tiedon julkisuustavoitteita toteuttaa EU:ssa direktiivi avoimesta datasta ja julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (EU 2019/1024, **PSI-direktiivi**). Direktiivi velvoittaa varmistamaan, että tietosisältöä, joihin direktiiviä sovelletaan, voidaan käyttää uudelleen kaupallisiin tai ei-kaupallisiin tarkoituksiin silloin, kun julkissektorin tietosisältöjen saatavuutta ei ole kansallisesti rajoitettu.

Laki digitaalisten palvelujen tarjoamisesta (306/2019, **digipalvelulaki**) velvoittaa viranomaiset tarjoamaan asiakkaille mahdollisuuden asioida viranomaisen kanssa sähköisesti tietoturvallisia menetelmiä hyödyntäen, huolehtimaan että sähköiset palvelut ovat saatavilla myös asiointipisteiden aukioloaikojen ulkopuolella, sekä koska palvelussa tulee käyttää vahvaa tunnistautumista.

Julkisen hallinnon tiedonhallinnasta annetun lain (906/2019, **tiedonhallintalaki**) tarkoituksena on varmistaa viranomaisten tietoaineistojen yhdenmukainen ja laadukas hallinta sekä tietoturvallinen käsittely julkisuusperiaatteen toteuttamiseksi. Laki edellyttää kaupungin tietovarantojen ja tietovirtojen kartoituksen sekä tietojärjestelmien, tietosisältöjen ja tietovarantojen parempaa yhteentoimivuutta organisaation sisällä mutta myös suhteessa muihin viranomaisiin. Lain keskeisiä kohtia on avattu tarkemmin [kappaleessa 3.2](#).

Laki hallinnon yhteisistä sähköisen asioinnin tukipalveluista (571/2016, ns. **Kapa-laki**) parantaa julkisten palvelujen saatavuutta, laatua, tietoturvallisuutta, yhteentoimivuutta ja ohjausta sekä edistää julkisen hallinnon toiminnan tehokkuutta ja tuottavuutta.

Suuri osa kaupungin hallinnassa olevasta datasta sisältää **tietosuoja-asetuksen ja tietosuoja-lain** (1050/2018) mukaista henkilötietoa, johon liittyy asetuksen määrittämiä tiedon käytön rajoituksia ja ehtoja. Erityisesti tulee huomioida se, että kaupunki on viranomainen, jolloin henkilöiden antamaan suostumukseen liittyy lisäksi erityisiä rajoituksia.

Arkistolain mukaan viranomaisen on määrättävä tehtävien hoidon tuloksena kertyvien asiakirjojen säilytystavat ja -ajat ja ylläpidettävä niistä arkistonmuodostussuunnitelmaa. Kaupungilla on tätä varten yhteinen tiedonohjaussuunnitelma tiedonohjausjärjestelmässä. Sen sisältämät metatiedot mahdollistavat myös tietoaineiston pysyvän ja pitkäaikaissäilytyksen.

Sosiaali- ja terveystietojen käyttöä säätelevät lukuisat lait, jotka asettavat erityisiä ehtoja tietojen käsittelyyn sekä rajoittavat tiedon hyödyntämistä. Keväällä 2019 eduskunta hyväksyi lain sosiaali- ja terveystietojen toissijaisesta käytöstä (552/2019, **toisioalaki**). Laki mahdollistaa datan toissijaisen käytön SOTE-toimialan sisällä tietyin edellytyksin, mutta toimialan ulkopuolella vain rajatusti.

Myös muilla toimialoilla on runsaasti tiedon käsittelyyn vaikuttavaa lainsäädäntöä. Lainsäädännön runsas määrä ja monimuotoisuus on osaltaan hidastanut datan hyödyntämistä ja jakamista eri palvelualueiden välillä.

5.2. Tiedon hyödyntämisen työryhmä

Tiedon hyödyntämisen tehostamiseksi kaupungille perustetaan **tiedon hyödyntämisen** työryhmä, jonka tehtävänä on datan hyödyntämisen ohjaaminen juridiselta ja strategiselta kannalta. Ryhmän tarkoituksena kiihdyttää kaupungin datan hyödyntämistä. Ryhmä selkeyttää ja nopeuttaa juridisten näkökulmien tarkastelua, antaa päättävälle viranhaltijoille suosituksia datan hyödyntämisestä käytötapauskohtaisesti sekä kerää ja ylläpitää datan käytön ennakkotapauspankkia. Tiedon hyödyntämisen työryhmä on resursoitava hyvin, jotta se pystyy ottamaan kantaa datan hyödyntämisen käytötapauksiin ripeällä aikataululla.

Työryhmässä ovat edustettuina toimialojen liiketoiminta, data- ja analytiikkaosaaminen, kaupungin tietosuojatiimi ja toimialojen juridiikka. Ryhmä kokoontuu säännöllisesti, vähintään kerran kuukaudessa kaupunkiin perustettavan data- ja analytiikkayksikön johdolla.

Työryhmän päätehtävänä on käsitellä datan ja analytiikan käytötapauksia ja antaa niistä suosituksia toimialoille. Työryhmä pitää kirjaa annetuista suosituksista ja kehittää niiden pohjalta yhteisiä käytänteitä ennakkotapauspankkiin. Ennakkotapauspankin avulla datan hyödyntämisen reunaehdot selkeytyvät ja käytötapauksen käsittely nopeutuu jatkossa. Ennakkotapauspankki on myös kaupungin toimijoiden käytössä ja auttaa arvioimaan uusien tiedon hyödyntämisen tapauksien toteutettavuutta.

Suosittelujen antamisen lisäksi työryhmä auttaa linjausten toteutumista ja valvoo niiden noudattamista kaupunkitasolla. Käyttöoikeudet data-alustalla olevaan dataan myönnetään työryhmän linjausten perusteella.

Tutkimushankkeisiin ja tutkimuslupiin liittyvien käytäntöjen käsittely rajataan pääsääntöisesti tiedon hyödyntämisen työryhmän tehtävien ulkopuolelle. Tutkimuslupiin on olemassa vaikiintuneet käytännöt, joita kuvataan tarkemmin [kohdassa 5.4](#).

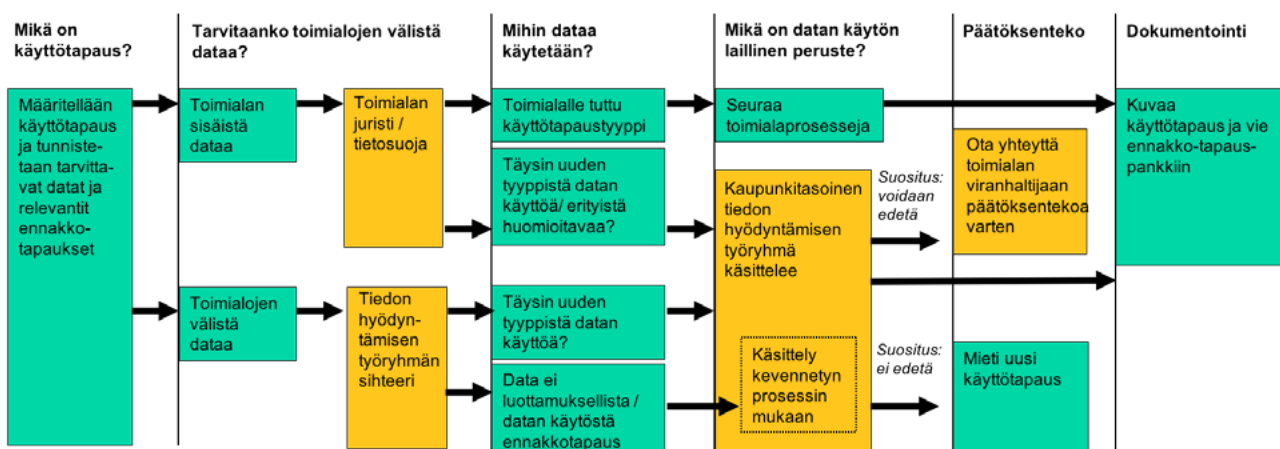
5.3. Käyttötapausten käsittelyn prosessi

Käyttötapausten käsittelyn sujuvoittamiseksi tiedon hyödyntämisen työryhmä käsittelee käyttötapausta kuvan 8 mukaisella prosessilla. Käyttötapausten omistajat (esim. toimialojen liiketoimintaedustajat) kuvaavat käyttötapaustensa työryhmän määrittelemälle lomakkeelle mahdollisia ennakkotapauksia hyväksi käyttäen. Käyttötapauslomakkeella kysytään tietoa datasta luokiteltuna tietosisällön ja datan käytön mukaan (4.1 ja 4.2). Prosessissa on eriytetty toimialan sisäisen datan käsittely, jonka hyödyntämisestä päätetään toimialan sisäisesti silloin kun käyttötapaustyyppi on tunnettu. Toimialojen välisen datan hyödyntämisen sekä täysin uudentyyppiset tiedon hyödyntämisen tapaukset käsitellään tiedon hyödyntämisen työryhmässä, joka antaa suositukset etenemisestä toimialan viranhaltijalle ja kirjaa tapauksen ennakkotapaushankkiin.

Jos käyttötapaukselle löytyy samankaltainen ennakkotapaus aiemmin käsiteltyjen tapausten joukosta tai hyödynnettävä data ei ole luottamuksellista (kyse esimerkiksi ei-henkilötietoa sisältävästä datasta, kuten Helsingin puut tai terassit), käyttötapaus voidaan käsitellä ns. kevennetyn prosessin mukaisesti. Tällöin toimialan juristi tai tiedon hyödyntämisen työryhmän sihteeri tekee suosituksen perusteluineen ja vie sen tiedoksi työryhmälle. Kaikki käyttötapaukset kuvataan perusteluineen (miksi/miksi ei dataa avattu) ja tallennetaan ennakkotapaushankkiin. Tavoitteena on se, että toimintamallin vakiintuessa käyttötapausta voidaan ratkaista enenevässä määrin suoraan ennakkotapaushankkiin kerättyjen esimerkkien avulla ilman tiedon hyödyntämisen työryhmän erillistä käsittelyä.

EU:n tietosuoja-asetus antaa mahdollisuuden hyödyntää lainsäädännön alaisia aineistoja tilasto- ja tutkimustoimintaan. Tutkimushankkeisiin on luotu omia lupakäytänteitä eikä tilasto- ja tutkimushankkeisiin käytettäviä aineistoja tarvitse viedä em. käyttötapauskäsittelyprosessin läpi.

Kuva 8. Tiedon hyödyntämisen toimintamalli.



5.4. Datan hyödyntäminen ja jakaminen tutkimushankkeissa

Myös kaupungin sisäisissä sekä kaupungin ulkopuolisten tutkijoiden toteuttamissa tutkimushankkeissa voidaan hyödyntää kaupungin tuottamia ja ylläpitämiä dataa. Tutkimushankkeita ovat tietopalveluna tuotettavat aineistot, tilastotarkoituksiin tuotettavat aineistot sekä konsulttisopimuksella kaupungin puolesta tehtäviin selvityksiin ja suunnitelmiin tehtävät aineistot ja luovutukset. Tutkimushankkeissa tarvittavien data-aineistojen luovuttamisesta päätetään erillisen tutkimuslupaprosessin kautta. Tutkimuslupien käsittelyyn on kaupungilla olemassa toimialakohtaiset käytännöt, jotka noudattavat voimassa olevaa lainsäädäntöä.

Tutkimuslupien käsittelyprosessia kehitetään edelleen. Tavoitteena on yhdenmukaistaa ohjeistusta lupien hakemisesta ja käsittelystä sekä luoda kaupunkiyhteinen toimintamalli tilanteisiin, joissa tarvitaan päätös useamman toimialan vastuulle kuuluvista aineistoista.

Tutkimuslupien hakua ja käsittelyä helpotetaan mahdollistamalla tutkimuslupan hakeminen sähköisesti. Päätös tutkimuslupan myöntämisestä tehdään viranhaltijapäätöksenä sillä toimialalla ja/tai palvelukokonaisuudessa, jonka aineistoista luvassa päätetään. Lupaprosessin kehittämisen yhteydessä arvioidaan mahdollisuudet lupakäsittelyn keventämiseksi esimerkiksi kaupungin sisäisissä tutkimushankkeissa.

Tutkimuslupien käsittelyssä esiin nousevissa uudentyypisissä aineistojen yhdistämiseen liittyvissä erityiskysymyksissä voidaan hakea tukea tiedon hyödyntämisen työryhmältä.

Tutkimuslupaprosessin kehittämisestä vastaa kanslian ja toimialojen tutkimuslupa-asiantuntijoista koostuva työryhmä, jota koordinoi kanslian kaupunkitutkimus ja -tilastot -yksikkö.

5.5. Eettiset periaatteet ja algoritmi-rekisteri

Kaupunki ottaa eettiset periaatteet huomioon kaikessa tekoälyn ja edistyneen analytiikan hyödyntämisessä. Kaupunki kehittää yhdessä suomalaisten ja kansainvälisten kumppanien (mm. Sitra, Oikeusministeriö, Saidot, Espoo ja Amsterdam) kanssa vastuullisen tekoälyn hyödyntämisen viitekehyksen, eli metadatatmallin⁴.

⁴ <https://www.saidot.ai/resources>

⁵ European Commission (2020): Artificial Intelligence - A European approach to excellence and trust. White Paper.

Metadatatmalli kokoaa yhteen eri elementit, joiden avulla algoritmien toimintaperiaatteet tehdään läpinäkyviksi ja ymmärrettäviksi. Kaupungin digitaalisissa palveluissa käytettävät tekoäly-algoritmit kootaan yhteiseen algoritmirekisteriin, jonka tarkoituksena on dokumentoida ja viestiä avoimesti, miten kaupungin käyttämät mallit on rakennettu ja mihin dataan ne perustuvat. Tämä helpottaa kaupungin asiakkaita ymmärtämään palvelujen takana olevan päätöslogiikan.

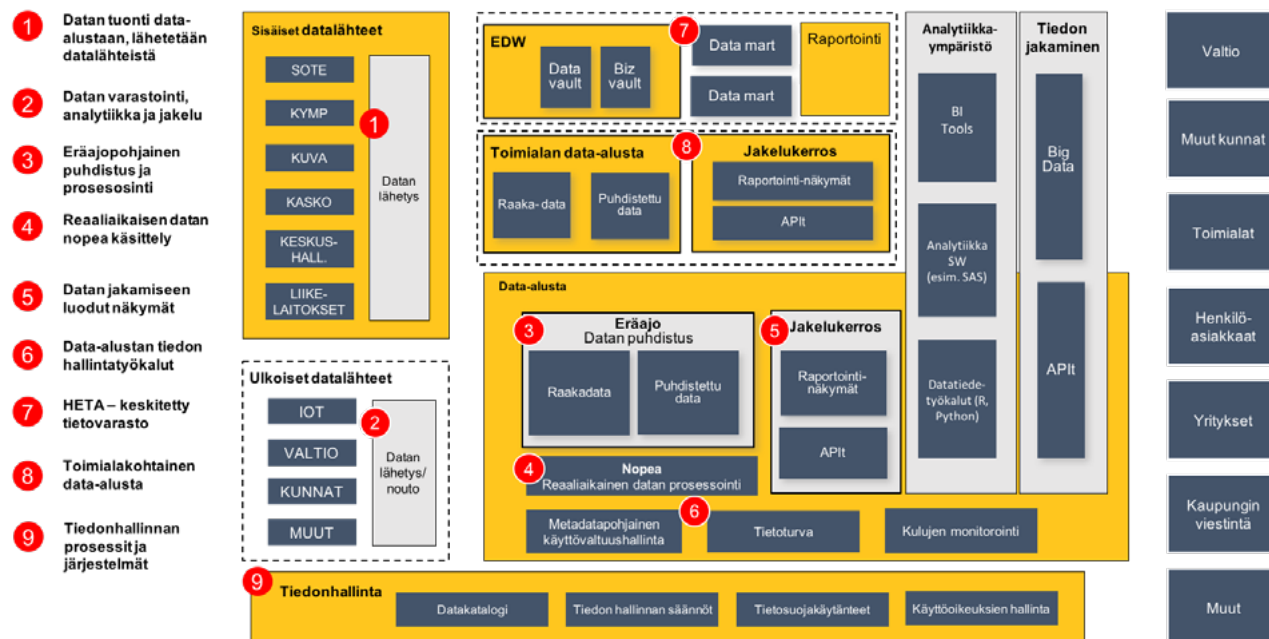
Myös Euroopan komission tekoälystrategia⁵ painottaa vahvasti tekoälyn läpinäkyvyyttä ja selitettävyyttä. Eurooppalaisten pitää pystyä luottamaan tekoälyn suosituksiin ja ratkaisuihin. Komissio ehdottaa uusia, merkittäviä investointeja dataan ja tekoölyyn, mutta myös uusia sääntöjä ja regulaatiota liittyen erityisesti ns. korkean riskin tekoälysovelluksiin. Tällaisina alueina mainitaan esimerkiksi terveydenhuolto, poliisivoimat ja liikenne. Komission valkoinen paperi on nyt julkisella kommentointikierroksella ja seuraavien vuosien aikana siitä kuoriutuu oletettavasti EU:n tekoälydirektiivi.

6. Tavoitearkkitehtuuri ja teknologiat

Kaupungin datastrategian keskeisenä periaatteena on hyödyntää kaupungin datavarantoja huomattavasti nykyistä laajemmin ja tehokkaammin. Datavarantojen laaja jakaminen, yhdistäminen ja hyödyntäminen tuo uusia vaatimuksia etenkin kaupungin keskitettyyn tietoarkkitehtuuriin, johon syntyy uusia tarpeita datan jakamiselle, ei-rakenteellisen tiedon käsittelylle ja edistyneen analytiikan kyvykkyyksille. Uusia vaatimuksia datavarantojen hallinnassa luo esimerkiksi kaupungin erilaisissa monitoringipalveluissa syntyvä sensoridata (ts. IoT-data), jota kerätään jo tällä hetkellä automaattisesti ja jatkuvasti. Sensoridatan erityisominaisuuksina ovat isot datavolyymit sekä tiedon reaaliaikaisuus. Kaupungin tietoarkkitehtuurissa ei ole tällä hetkellä osoittanut selkeää paikkaa talentaa sensoridataa, minkä takia data kerätään pääsääntöisesti toimittajan palvelimille. Datan jatkohyödyntäminen ja historiallisen datan keräytyminen ovatkin riippuvaisia toimittajasta ja hankintaehdoista, mikä ei ole ideaalinen tilanne.

6.1. Tavoitearkkitehtuuri

Kuva 9: Tavoitearkkitehtuuri tietovarantojen hallintaan



Kaupungin datavarannot on pääosin tallennettu lähdejärjestelmien yhteydessä oleviin tietokantoihin. Tavoitearkkitehtuurissa lähdejärjestelminä toimivat toimialojen ja liikelaitosten operatiiviset järjestelmät sekä ulkoiset toimijat kuten DVV, Tilastokeskus ja HSL. Kaupungin API-linjausten mukaisesti lähdejärjestelmien ja data-alustan väliset yhteydet toteutetaan ensisijaisesti API-pohjaisesti, ja jokaiseen hankittavaan järjestelmään tulee sisältyä API, jotta jo kerättyä ja kerättävää dataa voidaan hyödyntää laajasti. Ulkopuoliset datat noudetaan data-alustalle ulkopuolisilta toimijoilta heidän tarjoamin menetelmin. Mikäli toimijan tiedonsiirtomenetelmät eivät täytä nykyaikaisen tietojenkäsittelyn vaatimuksia, kaupungin tulee vaatia niiden kehittämistä käytettävälle tasolle.

Tarvittaessa on mahdollista myös siirtää raakadataa suoraan lähdejärjestelmistä data-alustalle dataeksploraatiota ja analyttistä mallinnusta varten. Datatieteen (esim. neuroverkot) luonteen kuuluu se, että datasta pyritään löytämään piileviä yhteyksiä ja riippuvuussuhteita ilman, että etukäteen tiedetään tarkkaan, mitkä data-attribuutit vaikuttavat tutkittavaan ilmiöön.

Kun raakadata on siirretty data-alustalle, tulee sille suorittaa esikäsittely, joka sisältää mm.

mahdollisten virheiden korjauksen, tietojen yhdistelyn ja mahdollisen aggregoinnin. Tähän kuuluu myös konseptuaalinen, looginen ja fyysinen datamallinnus yhtenäistä jatkokäsittelyä varten. Sensoridatan ja muun reaaliaikaisuutta vaativan datan käsittelyä varten on käytössä erillinen data-allas, jossa data käsitellään ajantasaisuusvaatimukset huomioiden. Rakenteellisemman datan käsittely voi tapahtua tietovarastossa.

Datavarantojen yhdistäminen ja analyttinen käsittely tapahtuu analytiikkaympäristössä, josta löytyvät tehokkaat työkalut datavarantojen käsittelyyn, yhdistämiseen sekä edistyneen analytiikan käyttötapauksien, kuten ennustemallien, toteutukseen. Valmiiksi käsitellyt datat, analyysit ja mallit julkaistaan rajapintojen kautta kaupungin sisäiseen ja ulkoiseen käyttöön sillä laajuudella kuin käyttöoikeudet julkaisun sallivat.

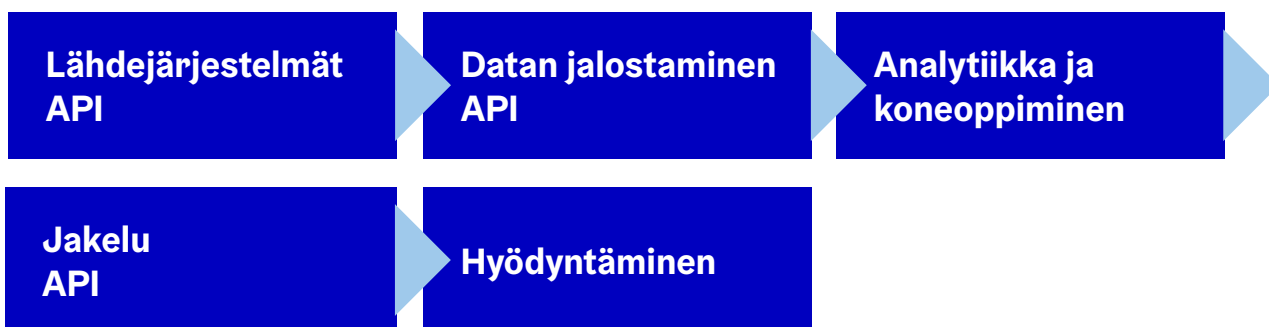
6.2. Data-alustan palvelumalli

Kaupungin datakyykykyyksien nostamiseksi uudelle tasolle luodaan kaupunkiyhteinen data-alusta, joka tarjoaa kehikon erilaisten datojen yhdistämiseen ja edistyneen analytiikan monipuoliseen hyödyntämiseen. Selvitetään yhteisen data-altaan (data lake) rakentamista sekä HETA-tietovaraston roolia yhteisessä data-alustassa. Yhteisesti hyödynnettävästä data-alustasta tehdään niin sujuvasti toimiva, että datan hyödyntäminen helpottuu merkittävästi. Data-alustalla tuotettu koodi ja data ovat kaupungin omistuksessa ja jaettavissa toimialojen kesken.

Kaupungin keskitetty data- ja analytiikkayksikkö kehittää toimialoille data-alustaa yhteisenä palveluna. Yksikön palvelujen avulla toimialat saavat nopeasti käyttöönsä data-alustaympäristön sekä toimialan omaa käyttöä varten että datan jakamiseen kaupungin ekosysteemissä. Data-alustan palvelumalli sisältää itse alustan lisäksi asiantuntijoiden palveluita ja tiedon jakamista, joiden avulla osaaminen kertyy toimialoille. Keskitetystä analytiikkaympäristöstä pystytään sujuvasti lukemaan tietoa toimialan data-alustalta sekä päinvastoin toimialan ympäristöstä pystytään lukemaan tietoa keskitetystä data-alustasta.

6.3. Dataputki ja käytetyt teknologiat

Data-alusta ja muut viitearkkitehtuurin teknologiaratkaisut muodostavat yhtenäisen kokonaisuuden, jossa raakadata jalostetaan vaiheittain tietovarannoista informaatioksi ja ennustemalleiksi. Data jalostuu seuraavasti data-alustan eri komponenteissa, ns. dataputkessa:



Kuva 10. Datan hyödyntämisen prosessi

Lähdejärjestelmät. Data luetaan data-alustalle lähdejärjestelmistä API:n kautta API-linjausten mukaisesti. Tietoaineisto tallennetaan data-alustalle esimerkiksi data-altaaseen (data lake),

yhden loogisen kokonaisuuden sisältäminä tiedostoina. Esimerkiksi yhtenä päivänä kerätty tieto talletetaan yhteen tiedostoon.

Datan jalostaminen. Sisältö luetaan tiedosto kerrallaan ja jatkojalostetaan käsittelyä varten. Jalostus pitää sisällään tiedon oikeellisuuden tarkistuksen, mahdollisten virheiden korjauksen, yhdistämisen aiempiin tiedostoihin laajemmaksi kokonaisuudeksi sekä tarvittaessa datan aggregoinnin, anonymisoinnin tai pseudonymisoinnin. Jalostusvaiheessa suoritetaan kaikki tarvittavat tehtävät tiedon mallintamiseksi sellaiseen muotoon, joka tekee datan käyttövalmiiksi seuraavaa vaihetta eli analytiikkavaihetta varten. Käsiteltyä dataa kutsutaan usein nimellä datatuote. Datatuotteet tallennetaan data-alustan puhdistetun datan alueelle. Työkaluina käytetään data-alustan tarjoamia datan jalostustyökaluja ja skriptikieliä kuten Python.

Analytiikka ja koneoppiminen. Tiedon arvo syntyy tiedon hyödyntämisen kautta. Tämä voi tapahtua joko luomalla datatuotteista raportteja ja analyyskejä tai käyttämällä koneoppimis- ja tekoälymalleja esimerkiksi ennustemallien tai oppivien algoritmien luomiseen. Lopputulokset voidaan tallettaa uusiksi datatuotteiksi. Analyysiin ja mallinnukseen käytettävät menetelmät ja työkalut valitaan käyttötapauksen mukaan. Esimerkiksi avoimen lähdekoodin Python ja R tarjoavat monipuoliset datankäsittely- ja mallinnuskirjastot. Kaupungin tavoitteena on myös julkaista kehitetyt algoritmit avoimena lähdekoodina, jolloin sidosryhmät pystyvät verifioimaan mallin toiminnan ja oikeellisuuden.

Jakelu. Analytiikkavaiheen tuloksena syntyvät datat tallennetaan tiedon jakamisen kerrokseen relaatiotietokantaan tai muuhun tarkoituksenmukaiseen formaattiin, josta tulokset voidaan hakea rajapintojen avulla seuraavaan vaiheeseen eli datan hyödyntämiseen.

Datan hyödyntäminen. Analytiikan tulosten hyödyntäminen voi tapahtua monilla eri tavoilla ja samaa tulosjoukkoa voidaan myös hyödyntää monella eri tavalla samanaikaisesti. Yksinkertaisimmillaan tulokset julkaistaan raporttina, mutta yhä enemmän pyritään luomaan interaktiivisia graafisia visualisointeja, jotka muokkaavat datan korkeammalle abstraktiotasolle ja kertovat visualisointien avulla tarinoita jotka auttavat lukijaa ymmärtämään ilmiöitä ja trendejä. Raportoinnin ja visualisoinnin lisäksi dataa voidaan hyödyntää jakamalla dataa API:n kautta ulkopuolisille laitoksille tutkimuskäyttöön tai dataa hyödyntäville sovelluksille. Esimerkiksi rakennushankesuunnitelmissa voidaan mahdollistaa avatun datan ja kannustaa lupien hinnoittelulla tekemään yhtenäisiä suunnitelmia katuosuuksittain tai alueittain.

6.4. Käyttötapausten toteutus data-alustalla

Alla kuvataan esimerkkeinä, miten kolme pääkäyttötapausta toteutetaan data-alustalla. Ennen analyttisen mallin rakentamista käyttötapaukseen liittyvät datat on tuotava data-alustalle, ellei käyttötapausten datatuotteita ole vielä olemassa. Kun datat on tuotu alustalle, puhdistettu ja esikäsitelty vaadittavalla tavalla, voidaan tietoa hyödyntää analytiikkaympäristössä.

Asiakaskokemuksen parantaminen

Kaupungin digitalisaatio-ohjelman mukaisesti kaupunki panostaa proaktiivisiin ja henkilökohtaisiin palveluihin. Proaktiivisuus vaatii data-alustalta kyvykkyyksiä ennakoida asiakkaan henkilökohtaiset tarpeet ennen kuin asiakas itse ottaa kaupunkiin yhteyttä.

Data-alustan kannalta kyseessä on palvelutarpeiden asiakaskohtainen ennakointi, jossa käytetään syötteenä yksittäisen asiakkaan tai asiakasryhmän tietoa. Kun algoritmi löytää asiakkaan kriteereihin sopivan palvelun, voidaan palvelua suositella erilaisissa asiakaskohtaamisissa esimerkiksi asiakaspalvelutapahtuman yhteydessä tai lähettämällä asiakkaalle palvelusta viesti käyttäen hyväksi esimerkiksi Helsinki-profiilin kautta löytyviä yhteystietoja ja asiakkaan anta-

mia viestintäasetuksia. Data-alustan kannalta viestien lähetys voi tapahtua joko eräajona tai API:n kautta, mikä mahdollistaa myös reaaliaikaisten viestien lähetyksen suoraan operatiivisista järjestelmistä. Esimerkkinä ennakkoinnista on koulupaikan tai kielikurssin tarjoaminen asiakkaalle pohjautuen taustatietoihin.

Kaupungin henkilökunnan avustaminen palveluprosesseja optimoimalla

Kaupungin tuottamissa palveluissa syntyy iso määrä tietoa, jota keräämällä ja analysoimalla voidaan luoda dataan pohjautuvia tilannekuvia sekä optimoida palveluprosesseja.

Data-alustan kannalta dataa kerätään ja yhdistetään eri datavarannoista sekä julkaistaan visualisointeja varten.

Palveluprosessien optimoinnit suoritetaan optimointi- ja simulointialgoritmeilla, joilla haetaan prosessin vaihtoehtoja ne, jotka parhaiten optimoivat algoritmille syötetyt tavoitteet. Optimoinnin tuloksena voi olla esimerkiksi arvioitu henkilökunnan tarve palvelupisteellä tietyllä kellonajalla, jolloin ryhmänvetäjä voi laatia optimaaliset työllistat ennakkotiedon perusteella.

Tiedolla johtaminen

Tiedolla johtamisen käyttötapauksessa yhdistetään useita datavarantoja, jotta pystytään analysoimaan eri ilmiöiden ristikkäisvaikutuksia. Analytiikkavälineitä ja -menetelmiä kehittämällä voidaan monipuolistaa päättäjille suunnattua tietotarjontaa. Tämä tapahtuu hyödyntämällä perinteisten tilastojen ja historiallisten datojen lisäksi uusin tiedonkeruumenetelmin hankittuja dataa ja soveltamalla niihin ennustemalleja. Mallien luomista varten käytetään historiallista tietoa opettamaan mallia koneoppimisen menetelmin. Kun malli on opetettu, mallilla pystytään tuottamaan ennusteita ja skenaarioita päättäjille päätöksenteon tueksi. Skenaarioita voidaan visualisoida käyttämällä esimerkiksi dashboard-mittaristoja. Kaupunkimallin data-aineistojen ja tietomallien yhdistelmä luo alustan ja digitaalisen kaksosen, jonka integrointi kaupungin palveluprosesseihin tuo hyötyjä tehokkuudessa, tuottavuudessa, toiminnan suunnittelussa ja tarkkailussa sekä läpinäkyvyydessä.

7. Kompetenssit ja organisoituminen

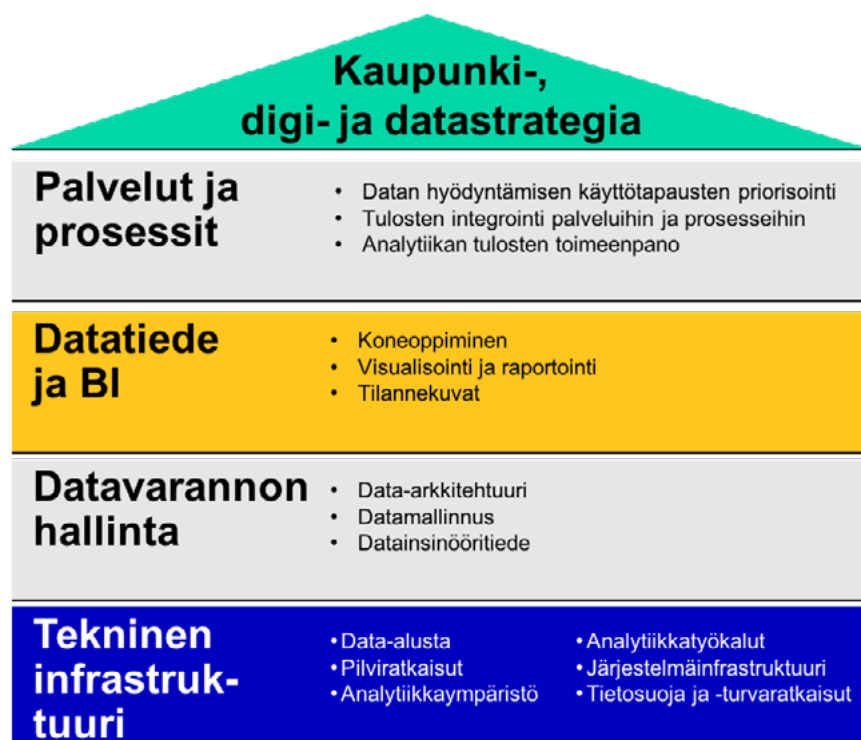
Datakyvykkyyksien rakentaminen perustuu osaaviin ihmisiin. Kaupungin paras analytiikka-osaajien keskittymä löytyy tällä hetkellä kaupunkitutkimus ja -tilastot -yksiköstä, jossa on tutkijoita, joilla on osaamista mm. tilastotieteen menetelmistä ja työkaluista. Lisäksi HETA-ympäristön toimialojen pääkäyttäjistä löytyy raportointiosaamista sekä teknologia-arkkitehti, joka vastaa teknisestä alustasta. 3D-tiimissä on osaamista simulaatio- ja visualisointimallien rakentamisesta. Muissa yksiköissä, kuten TPR-alustassa, löytyy raportointia tekeviä asiantuntijoita.

Kaupungilla on kuitenkin vähemmän kovan datatieteen osaajia, kuten datatieteilijät (data scientist), joiden osaamista ovat edistyneen analytiikan, tiedonlouhinnan ja koneoppimisen menetelmät. Lisäksi datavarantojen ylläpito on suurelta osin ulkoisten palvelukumppanien varassa ja kaupungilla olisi siten tarve omille datainsinööreille (data engineer), joiden osaamista ovat pilvi-, analytiikka- ja tietoaallasympäristöjen hallinnointi ja kehittäminen.

Toimialojen liiketoimintapuolella data- ja analytiikkaosaaminen ei ole nykypäivän vaatimuksiin ja mahdollisuuksiin nähden riittävää. Omia tavoitteita ei aina osata kääntää data- ja datatiedevaatuiksi tai ymmärtää, miten dataa voisi hyödyntää tehokkaasti tavoitteisiin pääsemisessä. Tätä voidaan edistää datan hyödyntämisen käyttötapauksiin erikoistuneiden ”tulkkien” eli datastrategien avulla. Lisäksi liiketoiminta- ja muille ihmisille tullaan järjestämään koulutuksia, joissa avataan datan, analytiikan ja tekoälyn mahdollisuuksia.

Kuva 11 esittää neljä keskeistä kyvykkyyksialuetta, jotka kaikki tarvitaan datan tehokkaaseen hyödyntämiseen. Toimialojen palvelut ja prosessit ohjaavat datan hyödyntämisen tavoitteita. Datatiede ja BI tuottavat datasta arvoa raportoinnin ja analytiikan (esim. koneoppiminen) menetelmin. Tietovarannon hallintaan kuuluu kaikki data operatiivisissa järjestelmissä ja data-alustoilla, esim. datan laatu, integroinnit, tietomallinnus, standardit, metatiedot ja muut asiat. Tekninen infrastruktuuri liittyy teknisten tietovarastojen ja tietoaltaiden kehitykseen ja ylläpitoon.

Kuva 11. Datastrategian kannalta olennaiset kyvykkyyksialueet



Dataosaajia voidaan kategorisoida monella tavoin, mutta kaupungin kannalta kriittisimmät roolit on listattu seuraavassa taulukossa ryhmiteltyinä edellä mainittuihin kyvykkyyksalueisiin (tarkemmat roolikuvaukset liitteessä):

Rooli	Kuvaus	Osaamisalue
Data- ja tekoälystrategi	Kääntää liiketoiminnan (toimialojen) vaatimukset data- ja tekoälykielelle, toimii projektipäällikkönä toteutusprojekteissa ja varmistaa, että tulokset otetaan käyttöön liiketoiminnoissa.	Palvelut ja prosessit
Datatieteilijä	Hallitsee koneoppimisen, tilastotieteen ja ohjelmoinnin; rakentaa analyyttisiä malleja ja algoritmeja; kommunikoi tuloksia tehokkaasti.	Datatiede ja BI
Data-arkkitehti	Määrittää kaupungin/toimialojen data-arkkitehtuurin; tietokantaratkaisut ja komponentit; tietokantakuvaukset; ydintiedon hallinta ja datastandardit	Tietovarannon hallinta
Datainsinööri	Suunnittelee ja rakentaa datan siirtämiseen tarvittavan infrastruktuurin (pilviympäristöt, dataputket), joka sisältää esim. taulut, tietokannat, automatisoinnin ja skaalautuvuuden.	Tietovarannon hallinta

Kaupunkiyhteinen data ja analytiikkatiimi

Data- ja analytiikkaosaamista täytyy nostaa ja rekrytointia tehdä sekä kaupunkiyhteisissä yksiköissä että toimialoilla. Jotta Helsinki pystyy ottamaan seuraavan askeleen kehittyneen analytiikan hyödyntämisessä, digitalisaatioyksikköön tarvitaan erillinen data ja analytiikkatiimi tai muu hallinnollinen ratkaisu, joka varmistaa strategiassa esiin nostettujen toimenpiteiden toteutumisen. Tiimiin kootaan kaupungin data-alustan rakentamisen kannalta kriittiset resurssit ja rekrytoidaan riittävän vahva tiimi datatieteilijöiltä ja datainsinöörejä. Tiimin osaamisen tulee olla riittävän kattava, jotta tiimi kykenee itsenäisesti kehittämään kaupungin yhteistä data-alustaa, tukea toimialoja dataintensiivisten palvelujen kehittämisessä ja lisäksi auttaa viemään kaupungin päätöksentekoa datavetoiseksi.

Käsitteitä

Aggregointi	Aggregointi on tiedon muokkaamista yleiselle tasolle tai sellaiseen muotoon, etteivät yksittäisiä henkilöitä tai henkilöryhmiä koskevat tiedot ole enää tunnistettavassa muodossa.
Anonymisointi	Anonymisointi on tiedon saattamista peruuttamattomasti muotoon, josta se ei ole suoraan eikä epäsuoraan yhdistettävissä tunnistettuun tai tunnistettavissa olevaan henkilöön. Tämä voi tapahtua joko aggregoimalla tai pilkkomalla yksilötason tieto siten, että yksilö ei ole tunnistettavissa (esimerkiksi yhteen sessioon liittyvä verkkokäyttäytyminen).
Algoritmi	Sarja täsmällisesti määriteltäviä käskyjä tai toimenpiteitä jonkin tehtävän tai prosessin suorittamiseksi.
Analytiikka	Menetelmä, jonka tarkoituksena on raakadataa tutkimalla ja analysoimalla tehdä johtopäätöksiä datan sisältämästä tiedosta.
API	API (Application programming interface) eli ohjelmointirajapinta on määritelmä, jonka mukaan eri ohjelmat voivat tehdä pyyntöjä ja vaihtaa tietoja eli keskustella keskenään.
Avoim data	Avoimella datalla tarkoitetaan julkishallinnon, organisaatioiden tai yritysten tuottamaa tai niille kertynyttä julkista tietoa, joka on avattu rakenteisessa muodossa vapaasti ja maksutta kaikkien hyödynnettäväksi.
Data	Data on tietoa, jolla itsessään ei ole välttämättä semanttista merkitystä tai informatiivista järjestystä.
Edistynyt analytiikka	Edistyneellä analytiikalla tarkoitetaan koneoppimismenetelmiä laajasti. Niihin kuuluvat menetelmät, joissa käytettävä algoritmi oppii itse yhteyksiä datasta. Oppimisen lisäksi koneoppiminen/tekoäly mahdollistaa ennakoivien mallien rakentamisen.
IoT-data	IoT (Internet of Things) tarkoittaa teknisten laitteiden suorittamia automaattista tiedonsiirtoa internet-verkon kautta. IoT-data on kyseisten teknisten laitteiden avulla kerättyä dataa, esimerkiksi lämpötila-anturin keräämä mittaustieto kohteesta.
Johdon työpöytä	Johdon työpöytä (engl. Dashboard), on auton mittaristoa muistuttava näkymä, jolla pyritään seuraamaan keskeisiä liiketoimintamuuttujia reaaliaikaisesti.
Koneoppiminen	Koneoppiminen (engl. Machine Learning, ML) on toimintaa, jossa ohjelma oppii toistuvista tapahtumista tai datasta ja muuttaa itse toimintaansa annetun tavoitteen suuntaan saamansa palautteen perusteella. Koneoppimisella pyritään automatisoimaan tiedon tulkintaa ja laajentamaan ohjelman havainnointikykyä.
Metatieto	Metatieto (engl. Metadata) on tietoa tiedosta eli kuvailevaa ja määrittävää tietoa datajoukosta.
Mikropalvelu	Mikropalvelu (engl. Microservices) on sovelluskehityksen lähestymistapa, jossa iso sovellus rakennetaan modulaarisista komponenteista tai palveluista.
Pseudonymisointi	Henkilötietojen käsittely siten, ettei niitä voida suoraan yhdistää rekisteröityyn käyttämättä henkilön yksilöinnin mahdollistavia lisätietoja. Lisätiedot tulee säilyttää erillään henkilötiedoista ja niihin tulee soveltaa teknisiä ja organisatorisia toimenpiteitä, joilla varmistetaan, ettei yhdistämistä tunnistettuun tai tunnistettavissa olevaan henkilöön voi tapahtua.
Rajaresurssi	Rajaresursseilla (engl. Boundary resources) tarkoitetaan sopimuksellisia ja muita yhteistoiminnallisia säännöksiä sekä teknisiä ohjelmistotyökaluja ja rajapintoja. Ne toimivat avoimina vuorovaikutuskanavina digitaalisen alustan ja minkä tahansa kolmannen osapuolen välillä. Rajaresurssit ovat tyypillisesti Internetissä vapaasti ja ilmaiseksi (tai pientä korvausta vastaan) saatavilla.

Tekoäly	Tekoäly on yleistermi algoritmeille, jotka pystyvät oppimaan ilman, että ihminen ohjelmoi toimenpiteitä tai vaiheita erikseen.
Tiedonohjaus	Tiedon käsittelyn ohjaaminen tietojärjestelmässä. Tiedonohjaussuunnitelma ohjaa asiakirjatiedon syntymistä ja käsittelyä tuottamalla metatietoarvoja ja käsitte-lysääntöjä (JHS 191). Helsingin kaupungilla tiedonohjaussuunnitelma ohjaa digitaalisen tiedon lisäksi myös manuaaliaineiston käsittelyä
Tieto	Tietoa synnytetään datasta antamalla datalle semanttinen merkitys.
Tietoaineisto	Tietoaineistolla tarkoitetaan asiakirjoista ja muista vastaavista tiedoista muodostuvaa tiettyyn viranomaiseen tehtävään tai palveluun liittyvää tietokokonaisuutta. Lähde tiedonhallintalaki (906/2019) 2 §.
Tietoallas	Tietoallas (engl. Data lake) on tietotekninen arkkitehtuuriratkaisu, joka perustuu isojen tietomassojen käsittelyyn soveltuvien teknologioiden hyödyntämiseen. Tietoaltaaseen pystytään nopeasti tallettamaan isoja tietomassoja sekä strukturoimatonta dataa.
Tietomalli	Tietomalli kuvaa, mitä tietokenttiä tietoaineistossa on ja miten ne liittyvät toisiinsa.
Tietovaranto	Kaupungin tehtävien hoidossa tai muussa toiminnassa käytettävä dataa sisältävä kokonaisuus yhdessä tai useammassa tietojärjestelmässä.
Tietovarasto	Tietovarasto (engl. Data Warehouse) on erillinen relaatiotietokantoihin pohjautuva arkkitehtuuri, johon eri järjestelmissä hajallaan oleva strukturoitu data poimitaan ja ladataan raportointia, analytiikkaa ja muuta käyttöä varten.
Ydintieto	Ydintieto (engl. Master data) on pysyvästi tarvittavaa perustietoa, joka on tunnistettu ja jaettu läpi organisaation ja jonka olemassaolo on välttämätöntä organisaation useiden prosessien ja toimintojen kannalta.
Yhteentoimivuus	Tietojärjestelmien kyky viestiä keskenään sellaisella tavalla tai siinä laajuudessa, että ne voivat rutiininomaisesti käyttää toistensa tuloksia. Tietojen hyödyntämistä ja vaihtoa eri tietojärjestelmien välillä siten, että tietojen merkitys ja käytettävyys säilyvät.

The logo for Helsinki, featuring the word "Helsinki" in white, bold, sans-serif font, enclosed within a white, stylized speech bubble or shield-like border. The background is a solid dark blue.

Helsinki

Helsingin kaupunki
Kaupunginkanslia

Pohjoisesplanadi 11–13
00170 Helsinki
PL 1
00099 Helsingin kaupunki
Puhelinvaihde 09 310 1641

www.hel.fi