



Pelikirja
avoimen liikunta- ja hyvinvointidatan hyödyntämiseksi

Sisältö


Johdanto	3
Mitä data on?	5
Liikunta- ja hyvinvointidatan keräämisen nykytilanne	7
Tulevaisuus ja mahdollisuudet.....	11
Haasteet, esteet ja pullonkaulat	16
Operointimallit	20
Datatalouden säännöt ja oppaat.....	23
Tietosuoja	25
Esimerkki toimintamallista.....	27
Lähteet	30

innokaupungit

JYVÄSKYLÄ 

KUOPIO **Lahti**

 **PIRKANMAA**
COUNCIL OF TAMPERE REGION

 Työ- ja elinkeinoministeriö
Arbets- och näringsministeriet

 Euroopan unionin
osarahoittama

- **Datan hyödyntäminen päätöksenteossa ja liiketoiminnan kehittämisessä on kasvava trendi.**
 - Useat kaupungit, kuten Kuopio, ovat laatineet dataohjelmia ja digitiekarttoja datan tehokkaaseen ja turvalliseen hyödyntämiseen.
 - Euroopan komission datastrategia pyrkii tekemään EU:sta maailman johtavan datatalouden.
- **Liikunta- ja hyvinvointidatan potentiaali on merkittävä julkisissa palveluissa, tutkimuksessa ja yritysinnovaatioissa.**
 - Teknologian kehitys mahdollistaa laajemman liikunta- ja hyvinvointidatan keräämisen.
 - Julkisen ja yritysten suljetun datan yhdistäminen voi kiihdyttää liiketoimintaa.
 - Datan avaaminen voi lisätä hallinnon läpinäkyvyyttä ja parantaa kuntalaisten palvelutasoa.
 - Liikuntadatan avulla voidaan kannustaa ihmisiä liikkumaan.

DATAN HYÖDYNTÄMINEN päätöksenteossa ja liiketoiminnan kehittämisessä on jatkuvasti kasvavan mielenkiinnon kohteena. Myös liikunta- ja hyvinvointidatan osalta potentiaalia hyödyntämiseen on niin julkisten palveluiden tuottamisessa, tutkimuksessa kuin yritystoiminnan innovaatioissakin. Teeman ympärillä on nähty viime vuosina uusia avauksia esimerkiksi eri kaupungeissa, oppilaitoksissa sekä huippu-urheilun tutkimusyksiköissä.

Teeman ajankohtaisuudesta kertoo se, että esimerkiksi useammat eri kunnat ja kaupungit ovat huomioineet datan hyödyntämisen strategiatyössään. Tämän myötä esimerkiksi Kuopion kaupunki on laatinut dataohjelman ja digitiekartan (dataohjelma, 2023) digikyvykkyyden edistämiseen ja datan tehokkaaseen sekä turvalliseen hyödyntämiseen. Käytännössä dataohjelman yhtenä tavoitteena on edistää datatalouden kehittymistä, eli luoda kykyä hyödyntää eri lähteistä kerättyä dataa eri toimijoiden kesken. Kaupunki pyrkii mahdollistamaan datataloutta mm. tietoliikenneinfrastruktuuria ja datan hyödyntämisen mahdollistavia alustoja rakentamalla.

Datan hyödyntäminen tulevaisuuden yhteiskunnan rakentamisessa on huomioitu myös kuntatasoa laajemmassa mittakaavassa. Euroopan komissio

(2020) on julkaissut eurooppalaisen datastrategian, jonka tavoitteena on rakentaa EU:sta kiinnostavin, turvallisin ja dynaamisin datatalous maailmassa. Tämän tavoitteen saavuttaminen vaatii koko alueen yhteisiä säännöksiä ja toimintamalleja.

Teknologian kehityksen myötä liikuntaan ja hyvinvointiin liittyvää dataa on mahdollista kerätä laajemmassa mittakaavassa, kuin aiemmin. Haasteet tiedon parhaaseen mahdolliseen hyödyntämisen liittyvät esimerkiksi tiedon siiloutumiseen – eri toimijoiden erilaisilla tavoilla keräämä ja hallinnoima data on heikosti ristiin hyödynnettävissä. Avoimemmalla toimintamallilla datan hyödynnettävyys paranee. Julkisen datan yhdistäminen yrityksiin suljettuun omaan dataan voi tarjota parempia mahdollisuuksia liiketoiminnan kiihdyttämiseen.

Datan avaamisen hyötyjä kaupungeissa voidaan tarkastella muutamastakin eri näkökulmasta. Siinä missä pääpaino voi olla yritysten ja innovaatiotoiminnan kehittämisessä, datan avaaminen lisää myös hallinnon läpinäkyvyyttä ja parhaassa tapauksessa nostaa kuntalaisten palvelutasoa.

Avointa dataa ei kuitenkaan pidä sekoittaa julkiseen tietoon. Erona on hyödynnettävyys – julkinen tieto on kaikkien katseltavissa, mutta avointa dataa voi jokainen toimija myös käyttää omiin tarpeisiinsa.

Tässä pelikirjan osiossa kuvataan datan keräämisen nykytilannetta Kuopion kaupungissa, lähinnä liikuntapaikkojen ja hyvinvointidatan osalta. Tämän lisäksi pelikirjassa pohditaan tulevaisuuden toimintamallin esteitä ja pullonkauloja, sekä myös erilaisia operointimalleja.

Mitä data on?

- **Data on laaja käsite, joka tarkoittaa kerättyjä arvoja, joista saadaan informaatiota ja tietoa.**
- **Yksittäinen datapiste ei kerro paljoa, mutta useamman arvon yhdistäminen tuottaa informaatiota.**
- **Avoin data on kaikkien käytettävissä olevaa digitaalista informaatiota, joka on koneluettavaa.**
- **Datatalous keskittyy datan hyödyntämiseen ja keruuseen, ja sen odotetaan kasvavan merkittävästi.**
- **EU:n datastrategia pyrkii luomaan datalle sisämarkkinat, joissa data voi liikkua vapaasti.**
- **Älykaupunki viittaa tulevaisuuden kaupunkiin, jossa hyödynnetään uusia teknologioita ja datalähteitä päätöksenteossa.**

DATA ON SUURIMMALLE osalle ihmisistä ainakin jollain tasolla terminä tuttu. Käsite on kuitenkin niin laaja, että on hyvä määritellä mitä datalla tässä kontekstissa tarkoitetaan ja minkälaista dataa voi olla mahdollista kerätä.

Laajan määritelmän mukaan data on kerättyjä arvoja, joista saadaan informaatiota ja pidemmälle jalostettuna voidaan kehittää uutta tietoa (Helsinki Institute of Physics, n.d.). Yksittäinen datapiste ei välttämättä kerro mitään, mutta sen yhdistyessä useampaan arvoon, voidaan jalostaa informaatiota. Opetushallitus (2024) määrittelee datan yksittäiseksi merkiksi tai symboliksi, informaation dataksi, jolla on merkitys ja tiedon informaatioksi, joka on asetettu oikeaan kontekstiin.

Kun taas puhutaan avoimesta datasta, tarkoitetaan digitaalisesti kaikkien käytössä olevaa informaatiota (digi- ja väestötietovirasto, 2023). Käytännössä siis datasta, jota jokainen voi hyödyntää omiin käyttö-tarkoituksiinsa avoimesti. Avoimen datan täytyy olla myös koneluettavaa – kaikki digitaalisessa muodossa oleva tieto ei täytä avoimen datan kriteerejä.

Datan lisäksi tämän kokonaisuuden teksteissä käytetään myös paljon termiä datatalous. Tämä viittaa talouden osa-alueeseen, jossa datan hyö-

dyntäminen ja keruu on keskiössä (Lähteenoja, 2023.) Datatalous tulee tulevaisuudessa läpileikkaamaan käytännössä kaikki liiketoiminnan osa-alueet ja muovaamaan yhteiskunnan tapaa toimia. Euroopan unioni arvioi EU:n datatalouden arvon kasvavan vuonna 2025 yli 820 miljardiin euroon. Euroopan yhteiseen datastrategiaan, datan saataavuuden ja hyödynnettävyyden parantamiseen ja datatalouden rakentumiseen myös panostetaan. EU on luomassa datalle sisämarkkinaa, jossa data voi liikkua vapaasti, yhteisiä säädöksiä noudattaen (Euroopan komissio, 2024).

Yksi tässäkin julkaisussa mainittava toistuva termi aiheen ympärillä on älykaupunki. Älykaupunki on laaja kattokäsite, jolle on vaikea löytää selkeää yksittäistä määritelmää (Mustonen ym., 2014). 2000-luvun alkupuolella yleistynyt ilmaisu viittaa tulevaisuuden kaupunkiyhteiskuntaan, jossa toimintaa voidaan kehittää uudenlaisia teknologioita hyödyntämällä. Tähän sisältyy suurena kokonaisuutena myös erilaisten datalähteiden hallinta sekä hyödyntäminen päätöksenteossa.



Liikunta- ja hyvinvointidatan keräämisen nykytilanne

- **Kaupunkiorganisaatiot keräävät dataa monipuolisesti, mutta yhteneväiset käytännöt puuttuvat.**
- **Liikuntapaikoilta kerätään tällä hetkellä lähinnä kävijämäärädataa.**
 - Kävijämääriä mitataan sekä sähköistä kulunvalvontaa hyödyntämällä, että ryhmäkokoja arvioimalla.
 - Uudet teknologiat, kuten konenäkö ja laserkeilaus, voivat parantaa kävijämäärien mittausta tulevaisuudessa.
- **Julkisilla liikuntapaikoilla on testausasemia huippu-urheilua ja tutkimusta varten.**
- **Laajempi ja avoimempi vuoropuhelu eri toimijoiden välillä parantaa datan hyödyntämistä.**
- **Avoimen datan portaalit ovat yleistyneet, mutta niiden käytännöt vaihtelevat.**
 - Datan avaaminen voi edistää uusien palvelujen innovointia ja tutkimusta.
 - Avoimen datan portaalille löytyy erilaisia operointimalleja, esimerkiksi datamarkkinapaikka.

NYKYPÄIVÄN kaupunkiorganisaatiossa kerätään dataa eri toimintayksiköissä monipuolisesti, mutta yhteneväiset käytännöt ovat vähissä. Osaa datasta voidaan kerätä ja hallinnoida kaupungin toimesta, mutta osassa tapauksista kolmas osapuoli teettää tutkimuksia ja tätä myöten omistaa kokonaisen data-aineiston.

Liikuntapaikkojen osalta tämän hetken käytännöt datan keräämisen suhteen rajoittuvat pitkälti erilaisiin kävijämäärämittauksiin. Sähköisten kulkuporttien paikoissa, esimerkiksi uimahalleissa, voidaan kävijädataa analysoida hyvinkin spesifillä tasolla. Jokainen asiakas leimaa itsensä sisään sähköisesti, joten muistiin jää tieto ajasta sekä lipputuotteesta, jota sisäänpääsyyn on käytetty. Myös esimerkiksi pukukaappien käyttöastetta voidaan seurata kattavasti – jokainen käyttäjä lukee kaappinsa avainranneketta hyödyntämällä.

Jos käytössä ei ole sähköisiä kulkuportteja, perustuu kävijämäärä jonkinasteiseen arvioon. Esimerkiksi liikuntahalleissa on tieto varausten täyttöasteesta, mutta yksittäisten liikkujien määrää ei tarkkaan tiedetä ryhmäkokojen vaihtelevuuden vuoksi. Sama pätee myös ulkoliikuntapaikkoihin.

Nykypäivän tekniikalla markkinoilta löytyy jo erilaisia, esimerkiksi konenäköä hyödyntäviä, kameratoteutuksia kävijämäärän mittaamiseen. Vielä toistaiseksi nämä eivät kuitenkaan isossa mittakaavassa ole tuoneet tarvittavaa tarkkuutta ja luotettavuutta kävijämäärien mittauksiin, mutta lähitulevaisuudessa nämä voivat olla ratkaisu kävijädatan keräämiseen. Erilaisten kameraratkaisuiden haasteena on luotettavuuden lisäksi myös yksityisyydensuoja, joka henkilöä kuvatessa nousee haasteeksi. Tämän vuoksi myös muiden ratkaisujen, kuten laserkeilauksella toteutettavan kävijämäärälaskennan, toivotaan yleistyvän ja monipuolistavan mahdollisuuksia laskennan toteuttamiseen.

Suomalaisissa kaupungeissa liikuntapaikkoja (stadioneita, halleja ja saleja) hyödynnetään paljon myös erilaisessa tapahtumakäytössä. Näitä ovat esimerkiksi urheilutapahtumat, konsertit ja erilaiset messut. Kävijälaskenta liikuntapaikoilla voi siis liikuntakäytön lisäksi palvella tapahtumateollisuutta ja tuoda mahdollisuuksia esimerkiksi erilaisten tapahtumien taloudellisten vaikutusten arviointiin, joka on oleellinen tieto esimerkiksi kaupunkien jakamien tapahtumatukien osalta.

Tulevaisuutta ajatellen on tärkeää pohtia, mitä muuta dataa kävijämäärien lisäksi liikuntapaikois-

ta olisi tarkoituksenmukaista ja mahdollista kerätä. Nykypäivän teknologiat mahdollista hyvinkin tarkan mittaamisen, mutta tarkoituksenmukaista on löytää ne arvoketjut, joihin datan tuottaminen tuo eniten hyötyä niin kuntaorganisaatiolle, kuin yritys-elämällekin. Julkisille liikuntapaikoille on myös rakennettu joissakin kaupungeissa erilaisia testausasemia palvelemaan huippu-urheilua sekä akateemista tutkimusta. Esimerkkeinä testausasemista toimivat esimerkiksi huippu-urheilun instituutti KIHU Jyväskylässä, Helsingin seudun urheilukatemian Urhean yhteydessä toimiva Helsingin urheilulääkäriasema HULA sekä Joensuu Areenaan Joensuun kaupungin ja Itä-Suomen yliopiston yhteisponnistuksella rakentuva TestLab Mehtimäki. Näiden testausasemien hyödyntäminen myös laajemmassa mittakaavassa voi tarjota uudenlaisia mahdollisuuksia sopivien data-alustojen mahdollistamana.

Liikuntapaikkojen ulkopuolella kaupunkiorganisaatioissa käsitellään hyvinvointiin liittyvää dataa myös muissa yksiköissä. Esimerkiksi koululaisille teetetään kouluterveyskyselyitä, joista saadaan näkemys kouluikäisten kokonaisterveydestä ja liikuntatottumuksista. Haasteina kouluterveyskyselyiden osalta on se, että kyselyt ovat valtakunnallisia ja kokonaisdata ei ole aina kaupunkiorga-

nisaation hallinnoitavissa. Lisäksi datan avaamista pohdittaessa terveystieteiden kysymykset ovat haastavia yksilönsuojan kannalta.

Julkisen puolen keräämän datan osalta tämän hetken tilannekuva pystytään maalaamaan suhteellisen selkeästi. Kaupunkiorganisaatioissa ei kuitenkaan välttämättä ole tarpeeksi tarkkaa tietoa siitä, mitä kolmansien osapuolien keräämä data voisi jo tällä hetkellä mahdollistaa. Tämän vuoksi toimijoiden välistä laajempaa ja avoimempaa vuoropuhelua tarvitaan. Selvitysten myötä voidaan löytää mahdollisuuksia, jotka jo pienelläkin työmäärällä parantavat datan hyödyntämisen potentiaalia.

Kuntaorganisaatioissa datan avaaminen on yleistynyt viimeisten vuosien aikana ja avoimen datan hyödyntämiseen esimerkiksi uusien palvelujen innovoinnissa, tutkimuskäytössä tai esimerkiksi tekniikan alan opinnoissa on kannustettu ahkerasti. Toisaalta esimerkiksi Ukrainan sodan seurauksena on nähtävissä myös vastaliikettä. Monissa paikoissa pohditaan esimerkiksi sitä, pitäisikö osa datasta, eritoten paikkatieto, sulkea turvallisuuskulmasta. Osasta isoimpia kaupunkeja löytyykin jo omia avoimen datan portaaleja. Kuopion Kaupungin datajalostamon lisäksi näitä ovat esimerkiksi Oulun dataportaali, Pääkaupunkiseudun

dataportaali (HRI), Lounais-Suomen dataportaali ja Tampereen dataportaali. Avoimen datan mahdollisuudet on siis tiedostettu useammassakin paikassa, vaikka matkaa täyden potentiaalihin hyödyntämiseen vielä onkin.

Avoimen datan portaalit ovat myös keskenään hieman erityyppisiä. Oulun, Helsingin, Lounais-Suomen ja Tampereen dataportaalit ovat keskittyneet vain avoimen datan jakamiseen, kun taas Kuopion Kaupungin datajalostamossa avoimenkin datan hyödyntäminen vaatii rekisteröitymisen. Osaan datalähteistä voidaan myös antaa pääsyoikeuksia vain osalle käyttäjistä tunnistautumisen perusteella. Kevytkin käyttöoikeushallinta voi vaikuttaa suuresti siihen, minkä tyyppistä dataa alustalla voidaan jakaa. Lisäksi rekisteröitymisen avulla voidaan seurata, kuinka paljon kutakin datalähdettä hyödynnetään.

Yksi vaihtoehto datan avaamiseen keskittyneen alustan hallinnointiin on luoda datamarkkinapaikka, jollaista on rakennettu esimerkiksi Tampereen IoT-alustan ympärille Tampereen datatori -nimellä. Datamarkkinapaikan ajatuksena on, että datan käyttäjä valitsee verkkokaupasta datalähteen, jota haluaa hyödyntää. Vaikka käytöstä ei perittäisi kukaan maksua, saadaan näin käyttäjä tunnistettua.

Kauppapaikkamalli mahdollistaa myös sen, että yksityiset toimijat voivat tuoda dataansa alustalle hyödynnettäväksi maksua vastaan.

Datan avaamisen yleistyttyä, on jatkuvaan kehittämiseen paneuduttu kaupunkiympäristöissä myös erilaisten hankkeiden kautta. Esimerkiksi Tampereen ammattikorkeakoulun toteuttamassa yhteisesti älykäs kaupunki (Yhtä) -hankkeessa selvitetään, kuinka yritykset voivat hyödyntää Tampereen kaupungin digitaalista ekosysteemiä. Hankkeessa tuotetaan pelikirja, joka sisältää teknisen dokumentoinnin siitä, kuinka kaupungin avointa dataa voi hyödyntää.



Tulevaisuus ja mahdollisuudet

- **Liikunta- ja hyvinvointidataa voidaan hyödyntää tiedolla johtamisessa, yritysten liiketoiminnassa, liikkumisen edistämässä ja tutkimuksessa.**
 - Tutkimusmaailma voi hyödyntää avointa dataa, mutta sensitiivisen datan käyttö vaatii lupa datan omistajalta.
 - Liikkumisen edistämässä dataa voidaan käyttää pelillistetyissä ratkaisuihin ja tekoälypohjaisissa sovelluksissa.
 - Datan hyödyntäminen yritysten liiketoiminnassa vaatii, että dataa on saatavilla jatkuvasti useasta eri lähteestä.
 - Julkiset organisaatiot voivat hyödyntää toisten avaamaa dataa tulevaisuuden hahmottamiseen, palveluiden suunnitteluun ja suoritusten seurantaan.
- **Huippu-urheilussa mittausratkaisut kehittyvät, ja anonymisoitu data voi hyödyttää myös muita tahoja.**
- **Liikkujat mittaavat itseään puettavilla laitteilla, ja dataa voitaisiin hyödyntää avoimien rajapintojen kautta.**
- **Laadukas ja kattava datavaranto on perusta kaikille tulevaisuuden kehityksille.**

JOTTA DATAN hyödyntämisen potentiaalia päästään valjastamaan parhaalla mahdollisella tavalla, täytyy tulevaisuutta pohtia myös pidemmällä tähtäimellä. Kun pidemmällä tulevaisuudessa siintävät tavoitteet saadaan määriteltyä, voidaan pienempiä kehitysaskelita suunnitella palvelemaan laajempaa kokonaisuutta.

Tulevaisuudessa tavoitteena on hyödyntää liikunta- ja hyvinvointidataa ainakin tiedolla johtamiseen kaupunkiorganisaatiossa, osana yritysten liiketoimintaa, liikkumiseen motivoivana keinona sekä tutkimustoiminnan mahdollistajana. Jokainen näistä kehityspoluista vaatii omanlaisia toimenpiteitä ja toimintamalleja.

Datan hyödyntäminen kaupunkiorganisaation tiedolla johtamisessa on tulevaisuuden visioinnissa siinä mielessä muita kokonaisuuksia yksinkertaisempi kuvitelma, että tarve ja hyöty on jo tässä vaiheessa helposti todennettavissa. Esimerkiksi kävijämäärien seurannassa ei myöskään ole tarvetta kaikista sensitiivisimmän datan keräämiseen, joten tietosuojan huomioiminen voi olla ratkaistavissa joitakin muita kokonaisuuksia kevyemmin. Huomioimatta tietosuojaa ei voi kuitenkaan tässä kokonaisuudessa jättää – jos esimerkiksi kävijätiedon voi jotain kautta yhdistää tiettyyn lip-

putuotteeseen, voidaan tämä käsittää henkilötiedoksi.

Datan hyödyntäminen liikkumisen edistämiseksi on tiedolla johtamisen lisäksi mielenkiintoinen tulevaisuuden näkökulma. Datan avulla voidaan rakentaa esimerkiksi erilaisia pelillistettyjä ratkaisuja, jotka parhaassa tapauksessa lisäävät liikkumista. Lisäksi esimerkiksi tekoälypohjaiset valmennus- tai avustussovellukset voivat tehostaa palvelutasoa liikuntapaikoilla. Datan hyödyntäminen onkin siis hyvin laaja kokonaisuus ja sisältää alleen hyvin laajan kokonaisuuden erilaisia mahdollisuuksia.

Jotta kerättyä dataa voidaan hyödyntää yritysten liiketoiminnassa, täytyy sen olla kattavasti hyödynnettävissä avoimien rajapintojen kautta. Avoimet rajapinnat mahdollistavat esimerkiksi sovellusten rakentamisen, jotka hyödyntävät datavirtaa käyttäjien hyväksi.

Viime aikoina varsinkin huippu-urheilun parissa ovat lisääntyneet erilaiset ratkaisut, joilla voidaan mitata urheilijoiden liikettä, nopeutta tai harjoittelutavaan lajiin liittyviä tekoja. Näiden mittausratkaisujen osalta on nähty kehitystä ja tulevaisuudessa ei välttämättä enää tarvita henkilökohtaisia senso-

reita, vaan mittaus voidaan toteuttaa liikuntatilaan rakennettavilla kiinteillä tai liikuteltavilla ratkaisuilla. Tämän tyyppinen data voisi tulevaisuudessa olla hyödynnettävää tietoa huippu-urheilun lisäksi myös monelle muulle taholle. Jos käytössä on anonymisoitua dataa liikuntatilaa käyttäneiden henkilöiden liikkeen määrästä ja toiminnasta, voidaan tästä saada tietoa siitä, kuinka paljon esimerkiksi nuoret liikkuvat liikuntatunnin tai joukkueharjoituksen aikana. Tai kuinka paljon esimerkiksi liikuntatilaan avattu ilmaisvuoro lisää liikettä. Tämäntyyppisen datan kerääminen voi olla tärkeässä roolissa liikkumattomuuden pommin purkamisessa ja terveellisemmän yhteiskunnan rakentamisessa.

Kaupunkiorganisaatioiden ja yritysmaailman lisäksi dataa voi hyödyntää myös muissa julkisissa organisaatioissa. Euroopan komissio (2023) kokosi julkisen sektorin mahdollisuuksia avoimen datan jälleen hyödyntäjänä osana julkaisuaan. Esiin nousi erityisesti kolme aluetta, joissa julkiset toimijat voivat hyötyä toisten julkisten toimijoiden avaamasta tai hallinnoimasta avoimesta datasta. Ensimmäinen näistä oli tulevaisuuden toimintojen hallinnointi, avointa dataa voidaan hyödyntää trendien ennustamiseen ja tulevaisuuden riskien ennakointiin. Toinen kokonaisuus oli palveluiden suunnittelu ja tuottaminen, datan avulla voidaan

tuottaa kansalaisille palveluita tehokkaasti ja tarpeeseen vastaten. Viimeisenä kokonaisuutena nostettiin esiin julkisen sektorin suorituksen seuranta sekä esimerkiksi tuottavuuden mittaaminen, jota avoimen datan avulla voidaan tehdä kokonaisvaltaisemmin.

Tutkimusmaailma voi mahdollisesti olla suurikin hyödyntäjä avatulle datalle. Tulevaisuudessa avoin data-alusta voi tarjota tutkimuksen tekijöille mahdollisuuden päästä käsiksi sellaiseen data-aineistoon, jota aiemmin on ollut haastava hyödyntää. Tutkimuskäyttöön hyödynnettävästä datasta suuri osa on kuitenkin siinä määrin sensitiivistä, että sen hyödyntämiseen tarvitaan lupa. Tämän ratkaisu on tulevaisuudessa suuressa roolissa. Millä tavoin henkilöltä kysytään lupa henkilökohtaisen datan hyödyntämiseen yksittäisessä tutkimuksessa? Suostumusta ei voi kysyä luovuttamishetkellä kaikkeen mahdolliseen hyödyntämiseen, eikä se myöskään palvelisi käyttäjiä parhaalla mahdollisella tavalla.

Helsingin Kaupunki on lähtenyt taklaamaan haastetta henkilökohtaisen datan hyödyntämisestä Helsinki-profiili hankkeella (Digitaalinen Helsinki, n.d.) Pitkän aikavälin visiona hankkeessa on luoda kaupungin asiakkaan tunnus, jonka avulla

voidaan asiointiin lisäksi hallinnoida henkilökohtaisen datan hyödyntämistä erilaisiin tarpeisiin. Tämän tyyppinen ratkaisu voi yleistyä jatkossa laajemminkin, jotta käyttäjä voi itse suostumusperustaisesti määritellä mihin tarkoituksiin hänestä kerättyä dataa voidaan hyödyntää.

Yhtenä selkeänä trendinä liikkumisessa on viime vuosina nähty liikkujien omatoiminen mittaaminen erilaisten puettavien mittauslaitteiden avulla. Urheilukellot, sovellukset ja esimerkiksi hyvinvointisormukset tuottavat dataa, joka parhaassa tapauksessa motivoi liikkumiseen ja oman suorituksen seuraamiseen. Moni mittauslaitteita tarjoava yritys mahdollistaa tällä hetkellä myös kerätyn datan hyödyntämisen avoimen rajapinnan kautta. Tulevaisuudessa kuntalaisten itse keräämän mittausdatan hyödyntäminen rajapinnan kautta anonymisoituna voisi tarjota uuden tyyppisiä mahdollisuuksia esimerkiksi kollektiiviseen liikkumiseen motivointiin tai tuottaa tietoa esimerkiksi eri asuinalueiden eroista. Toki jotta yksityishenkilöiden keräämää dataa saadaan hyödynnettyä, täytyy henkilöt jollain tavalla motivoida datan luovuttamiseen.

Markkinoille on noussut erilaisia palveluita (esimerkiksi urheilusovellus Strava), jotka keräävät

ja yhdistävät useasta eri lähteestä kerättyä dataa avoimen rajapinnan kautta. Kyseisessä sovelluksessa voi seurata omaa kehitystä ja harjoittelua, mutta myös esimerkiksi liittyä yhteisöllisiin haasteisiin. Tulevaisuudessa tämäntyyppisten palveluiden yhdistäminen kaupungin ylläpitämään data-alustaan voisi tarjota monentyyppisiä hyödyntämismahdollisuuksia. Tieto siitä missä, milloin ja millä tavoin kaupungin alueella urheillaan, on hyödynnettävissä niin kaupunkiorganisaatiossa, kuin esimerkiksi tutkimustoiminnassa. Yhdestä liikuntapaikoista puettavilla mittausvälineillä kerätystä datasta voidaan anonymisti saada tietoa esimerkiksi harjoituksen keskimääräisestä kestosta, intensiteetistä tai laadusta.

Liikuntadata, varsinkin liikkujaa mittaamisesta puhuttaessa, sekoittuu monissa tapauksissa myös terveysdatan kanssa. Terveysdatan avaaminen on yleisesti nähty viime aikoina mielenkiintoisena mahdollisuutena, mutta sen sensitiivisyys on nähty suurena haasteena. Irura (2019) nosti esiin oman julkaisussaan, että terveysdata sopii heikosti oletuksena kokonaan avoimeksi, mutta suljetun ja kokonaan avoimen datan välissä on myös välimalleja. Aineistoa voi avata osittain, anonymisoida tai muokata synteettiseksi, jotta dataa voidaan edes osittain hyödyntää tarpeen mukaisesti.

Huomioitavaa näissä kaikissa tulevaisuuspoluissa on, että jokaisen perustana on riittävän laadukas ja kattava datavaranto. Ensimmäinen, ja monessa mielessä haastavin, vaihe on luoda luotettavat dataputket ja näille toimittaja. Dataputkien osalta täytyy määritellä, mitä dataa sen kautta on saatavissa, millä käyttöoikeuksilla sitä voi hyödyntää ja millaisia sopimuksia datan hyödyntäminen vaatii. Lisäksi yksittäisten dataputkien tiedot on koottava kokonaiskuvaksi, jonka perusteella voidaan muodostaa esimerkiksi datakatalogi-ratkaisu. Katalogiratkaisun avulla tietoja datalähteistä ja hyödyntämismahdollisuuksista voidaan jakaa datan hyödyntäjille. Tämän jälkeen sovellusten ja muiden ratkaisujen rakentaminen on mahdollista. Alustalle syötettävälle datalle täytyy löytyä tarveperusta, mutta ratkaisuja voi rakentaa vasta toimivan dataputken päälle. Kattava datakatalogi tukee myös tiedonhallinnasta vastaavien työtä.

Monessa tapauksessa datan avaamista pohdittaessa jaotellaan data suljettuun ja täysin avoimeen dataan. Näiden kahden ääripään välissä on myös erilaisia vaihtoehtoja, jossa dataa avataan osittain luvittamisen kautta tietyille tahoille ja tiettyyn tarkoitukseen. Vaihtoehdot kahden ääripään välissä ovat tulevaisuuden kannalta hyvinkin mielenkiintoisia ja voivat monessa tapauksessa olla

mahdollistamassa uudenlaisia keinoja dataputkien hyödyntämiseen. Ratkaistava kysymys tässä datan osittaisessa avaamisessa on luvitusprosessin toimivuus. Periaatteiden, joilla lupa hyödyntämiseen annetaan, täytyy olla selkeitä ja datan omistajalla täytyy myös olla mahdollisuus hallinnoida sitä, mihin hänen dataansa hyödynnetään.

<p>Suljettu data</p> <p>Data on saatavilla tietyille organisaatiolle tai organisaation sisäiselle ryhmälle. Datalla on selkeä omistajuus ja organisaatio joka hallinnoi sen käyttöä ja jakelua.</p> <p>Suljettu data voi sisältää esimerkiksi henkilötietoja ja sen käsittelyyn voidaan tarvita perusteltu syy.</p>	<p>Sisäisesti tai sopimuksella jaettu data</p> <p>Dataa jaetaan organisaation osien välillä tai kumppaneiden kesken. Käyttäjät tiedetään tai tunnetaan pääsyä avattaessa ja käyttäjiä sitovat käyttöehdot.</p> <p>Sisäisesti tai sopimuksella jaettuun dataan voi sisältyä datan infrastruktuuriin ja esimerkiksi alustojen käyttökustannusten jakoa.</p> <p>Sopimuksella jaettua dataa voidaan myös myydä organisaation sisällä toimiville tytäryhtiöille tai kumppaneille.</p>	<p>Rajoitettu julkinen data</p> <p>Data on jaettu laajemmalle yleisölle. Kaikissa tapauksissa datan käyttäjiä ei välttämättä tiedetä tai yksilöidä suoraan, mutta pääsy datavarantoon on lähtökohtaisesti rajoitettu ja käyttäjät sitoutuvat tiettyihin käyttöehtoihin.</p> <p>Rajoitetusti jaettuun julkiseen dataan voi sisältyä ehtoja esimerkiksi datan kaupallisesta käytöstä. Rajoitetun julkisen datan käyttöä voidaan käyttöoikeuksien ja tunnusten hallinnan kautta seurata kattavammin kuin avoimen datan.</p> <p>Rajoitettua julkista dataa voidaan myös myydä ulkopuolisille organisaatioille ja sen edelleen käyttöä voidaan kontrolloida erilaisin lisenssein.</p>	<p>Avoim data</p> <p>Dataa, joka on julkaistu esimerkiksi avoimella lisenssillä käytettäväksi, kopioitavaksi ja jaettavaksi ilman rajoituksia.</p> <p>Avoimen datan käyttäjiä ei tyypillisesti pystytä yksilöimään ja sen mahdollisten väärinkäytösten seuraaminen on haastavaa.</p> <p>Avoim data on tyypillisesti käyttäjilleen maksutonta ja jakelun kustannuksista vastaa datan avaaja.</p>
--	---	---	--

Datan avautuessa sen hallinta monimutkaistuu, mutta mahdollisuudet uusien käyttökohteiden tunnistamiseen kasvavat!

Haasteet, esteet ja pullonkaulat

- **Datatalouden haasteet jakautuvat kahteen osaan: edellytysten rakentaminen (sääntely, infrastruktuuri) ja yksittäisten mahdollisuuksien hyödyntäminen (yritysten liiketoiminta).**
- **Suomessa on korkea osaamistaso ja laaja tietoyhteiskunta, mutta datatalouden toimenpiteet ovat hajanaisia ilman yhteistä kansallista strategiaa.**
 - TKI-rahoituksen lisääminen voi kehittää infrastruktuuria ja lisätä ymmärrystä datataloudesta.
- **Tietosuoja ja GDPR ovat yksi suurimmista haasteista liikuntapaikkadatan hyödyntämisessä.**
- **Ulkoliikuntapaikkojen kävijädatan kerääminen on haastavaa alueiden laajuuden ja mittausratkaisujen luotettavuuden vuoksi.**
- **Datan kerääminen ja hyödyntäminen vaatii resursseja ja selkeitä hyödyntämismahdollisuuksia.**
- **Avoimeen dataan perustuvien liiketoimintojen jatkuvuus ja vastuunjako on turvattava sopimuksilla.**
- **API-rajapintojen hyödyntämisessä haasteina ovat puutteellinen ymmärrys, liiketoimintatarpeen tunnistaminen ja löydettävyys.**
 - Datatalouden perusinfrastruktuuriin kuuluvat keskitetyt data-alustat voivat helpottaa rajapintojen löydettävyyttä ja hyödynnettävyyttä.

DATATALOUDEN HAASTEISTA puhuttaessa monesti yhdistetään kaksi hieman toisistaan poikkeavaa kokonaisuutta: datatalouden edellytysten rakentaminen (esimerkiksi sääntely ja infrastruktuuri) sekä yksittäisen mahdollisuuden hyödyntäminen (esimerkiksi yksittäisen yrityksen liiketoiminnan kehittäminen). Molempiin näihin liittyviä ongelmia täytyy ratkoa hyödyntämisen sujuvoittamiseksi, mutta toimenpiteet ovat erityyppisiä. Laajempaan datatalouden edellytysten rakentamiseen liittyen Suomessa on korkea osaamistaso sekä laaja tietoyhteiskunta, mutta datataloutta edistävät toimenpiteet ovat toistaiseksi jääneet hajanaisiksi eikä yhteistä kansallista strategiaa datakehitykseen ole täysin saatu luotua ja jalkautettua (Ahvonen ym, 2022). Kaupunkien hallinnoimat avoimen datan portaalit lisäävät mahdollisuuksia datatalouden edistämiseen, mutta on tärkeää, että ne linkittyvät toisiinsa myös kansallisen strategian kautta. Ahvonen ym. (2022) nostivat esiin Sitran julkaisemassa muistiossa datatalouden kehityksen mahdollistavana toimenpiteenä myös TKI-rahoituksen lisäämisen, jolla voidaan kehittää tarvittavaa infrastruktuuria sekä lisätä ymmärrystä ja osaamista datatalouden ympärillä.

Yksi suurimpia ratkaisun vaativia haasteita luotettavan liikuntapaikkadatan hyödyntämisessä

on tietosuoja ja GDPR. Kuten aiemmin mainittu, sähköisten kulkuporttien kohteissa kävijämäärien ja -profiilien seuraaminen on verrattain helppoa myös yksilönsuojan osalta. Esimerkiksi ulkoliikuntapaikoissa tilanne on kuitenkin haastavampi. Suurin osa kävijämittausratkaisusta toteutetaan kameroiden avulla, joka luo omat erityispiirteensä esimerkiksi kasvojentunnistukseen.

Toinen haaste kävijädatan keräämiseen ulkoliikuntapaikoista tuo alueiden luonne. Suurin osa ulkoliikuntapaikoista kattaa laajan alueen, eikä sisään-tulo välttämättä tapahdu vain yhtä kulkuväylää pitkin - kuinka luotettavasti mitata kulkijoiden määrä, kun koko alueen kattaminen kameroilla on lähes mahdotonta? Lisäksi useammalla mittauspisteellä alueen kattaminen luo mahdollisuuden sille, että sama kohde kirjautuu useampaan kertaan mittaustuloksiin. Kun käyttäjät liikkuvat vapaasti ympäri mittausaluetta, ei useampaan havainnointipisteeseen päätymistä voi estää. Jos data ei ole tarpeeksi luotettavaa, ei se ole myöskään arvokasta hyödynnettävää. Uudenlaisia ratkaisuja siis tarvitaan.

Datan keräämisen osalta on myös tärkeää huomioida, että se harvoin, jos koskaan, on ilmaista. Datan kerääminen, oli se sitten kävijädata tai esi-

merkiksi käyttäjien erilainen mittaaminen, vaatii resursseja. Jotta resursseja voidaan perustellusti ohjata datan keräämiseen, täytyy olemassa olla jo ajatus erilaisista hyödyntämismahdollisuuksista. Datan kerääminen itsessään ei vielä lisää valmiuksia tiedolla johtamiseen tai avaa uusia innovaatiomahdollisuuksia, jos kertynyttä aineistoa ei ole kyvykkyyttä hyödyntää. Datatalouden liiketoimintamallit ovat suurilta osin vielä muodostumassa ja hyvien toimintamallien löytäminen niin kustannusten jakoon kuin varsinaiseen liiketoiminnan mahdollistamiseen vaatii kokeiluja ja uutta ajattelua.

Avoimeen dataan perustavien liiketoimintojen osalta myös jatkuvuus on yksi ratkaistavista ja pohdittavista asiakokonaisuuksista. Jos liiketoiminta rakentuu esimerkiksi avoimen datalähteen päälle rakennettuun sovellustoteutukseen, täytyy pystyä luottamaan siihen, että data on saatavilla myös tulevaisuudessa. Datan toimittamisen vastuiden täytyy myös olla selkeät. Joissakin tapauksissa datan toimittaja, alustan ylläpitäjä ja loppukäyttäjä ovat kaikki eri tahoja, joten vastuunjako on hyvä turvata esimerkiksi sopimuksilla.

Datan keräämisen lisäksi myös sen hyödyntäminen vaatii resursseja. Avoimen datan hyödyntämi-

nen innovaatiotoiminnassa on yksi suuri tavoite kokonaisuudessa, innovaatioiden avulla voidaan luoda uutta liiketoimintaa jopa kansainvälisessä mittakaavassa. Koski ym. (2017) nostivat esiin valtioneuvoston kanslian selvityksessä, että vuonna 2017 yli 250 henkilöä työllistävistä yrityksistä yli viidennes oli hyödyntänyt avointa dataa innovaatiotoiminnassaan, kun taas 10–49 henkilöä työllistävässä yrityksissä määrä oli vain runsas kymmenes. Tulevaisuudessa olisi oleellista saada avoimen datan hyödyntäminen enemmän saavutettavaksi myös pienemmille yrityksille, jotta täysi potentiaali liiketoiminnan edistämiseksi saadaan hyödynnettyä. Tilastokeskuksen (2023) tilaston mukaan yrityksistä julkisen sektorin avointa dataa hyödynsi vain 9% yrityksistä. Potentiaali laajempaan hyödyntämiseen on siis olemassa vielä reilusti.

Euroopan Unioni on pyrkinyt viime aikoina rakentamaan yhteistä sääntelyä datan hyödyntämiselle kiinnostuksen jatkuvasti kasvaessa aiheen ympärillä. Mattila ym. (2022) nostavat elinkeinoelämän tutkimuslaitoksen ETLA:n muistiossa esiin, että EU:n sääntely, esimerkiksi datasäädös (engl. Data Act), tekee dataan liittyvästä sääntelystä liian raskasta pienten yritysten kannalta. Tämä voidaan nähdä haasteena tulevaisuuden kehityksen kannalta.

Datalouden ja datan hyödyntämisen kannalta tarvitaan onnistuneita case-esimerkkejä datan hyödyntämisestä, jotta toimijoita saadaan mukaan tuottamaan dataa hyödynnettäväksi avoimen datan alustalle. Poikola ym. (2024) tunnistivat Sitran data-avaruuksiin liittyvässä selvityksessä yhdeksi haasteeksi datan hyödyntämiseen toimijoiden välisen luottamuksen saavuttamisen. Joissain tapauksissa varsinkin yritykset pitävät datan jakamista riskinä, koska joku toinen taho voi hyödyntää sitä heitä laadukkaammin. Tämä voi johtaa kilpailuedun heikkenemiseen markkinoilla. Datan avaamisen hyödyt täytyy siis tuoda selkeästi esille, jotta mahdollisuudet ovat tätä huolenaihetta painavammat päätöksenteossa.

Jos pohditaan mahdollisia hyödyntämismahdollisuuksia kaupunkiorganisaation näkökulmasta, nämä voivat linkittyä laajemmin strategiaan tavoitteisiin. Esimerkiksi Kuopion kohdalla Pohjois-Savon heikot terveys- ja mielenterveystilastot ovat osaltaan vaikuttaneet hyvän elämän pääkaupunki-tavoitteen asettamiseen ja Mieliteko-hankkeen aloittamiseen, joissa onnistumiseen voidaan dataa hyödyntää. Resurssien kohdentamisen lisäksi esimerkiksi kävelijälaskurien asettaminen kaduille voi tuoda lisää tietoa liikenteen päästöjen hallintaan ja liikuntapaikkojen kävijädata voi valottaa

kuntalaisten terveyden edistämiseen suunnattujen toimien onnistumista ja toisaalta ongelmakohtia. Yksittäisistä dataputkista saatava tieto voi monesti olla hyödynnettävyydeltään parhaimmillaan muun tiedon kanssa yhdistettäessä.

Tietoyhteiskunnan kehittämiskeskus ry (2021) tunnisti kansallisten API-periaatteiden kehittämishankkeessa erilaisia haasteita API-rajapintojen hyödyntämisessä. Ensimmäinen haasteista oli puutteellinen ymmärrys rajapintojen tarjoamista mahdollisuuksista. Tämä haaste on hyvä huomioida myös avointa dataa tarjoavan alustan näkökulmasta. Tarvitaan onnistunutta viestintää, jotta datan potentiaaliset hyödyntäjät saadaan

kiinnostumaan mahdollisuudesta. Konkreettiset menestystarinat ja esimerkit onnistuneesta avoimen datan hyödyntämisestä ovat hyviä keinoja mahdollisuuksista viestimiseen. Kuntatoimijan ja data-alustan ylläpitäjän näkökulmasta mielenkiintoisia todettuja haasteita olivat myös haasteet liiketoimintatarpeen tunnistamisessa sekä API-rajapintojen löydettävyyks. Näitä molempia pystytään ehkäisemään laadukkaalla viestinnällä ja yhteistyökumppaneiden kouluttamiselle. Älykaupunki-infrastruktuuriin liittyvällä avoimella data-alustalla voidaan myös helpottaa rajapintojen löytämistä, kun dataa on saatavilla monipuolisesti samasta paikasta.

Operointimallit

- **Datan avaamisessa on hyvä tehdä valikointia hyödyllisyysnäkökulmasta ja tunnistaa selkeät käyttötarpeet.**
 - On tärkeää kartoittaa, missä määrin data avataan täysin julkiseksi ja missä määrin hallitaan käyttöoikeuksilla.
 - Yksi mahdollinen rooli kaupungille datan avaamisessa on mahdollistaa kolmansien osapuolien hyödyntäminen.
 - Datan avaaminen vaatii vastuiden ja veloitteiden ymmärtämistä.
- **Dataa voidaan avata tiedostomuodossa tai API-rajapinnan avulla, joista API mahdollistaa jatkuvan päivittymisen.**
- **Hyödynnettävän datan tulee täyttää FAIR-periaatteet (findable, accessible, interoperable, re-usable).**
- **Yhteentoimivuutta voidaan parantaa tukeutumalla kansainvälisiin laatukriteereihin ja ohjekokonaisuuksiin datainfrastruktuurin rakentamisessa.**

JULKISEN DATAN AVAAMISEN suhteen on ensimmäisenä järkevää kartoittaa missä määrin on hyödyllistä avata data täysin julkiseksi ja missä määrin julkisuutta voidaan hallita erilaisilla tokeneilla ja käyttöoikeuksilla. Käyttöoikeuksien myötä dataa voidaan avata myös osittain julkiseksi, jolloin voidaan harkintaperustaisesti pohtia, kuka dataan pääsee käsiksi. Toki varsinkin julkisorganisaation hallinnoiman datan kohdalla perusteiden käyttöoikeuksien myöntämiseen täytyy olla läpinäkyvät ja selkeät, jotta mahdollisuus on saatavilla kaikille kriteerit täyttävälle toimijoille.

Isommassa kuvassa tärkeä pohdinta on se, mikä on julkisen organisaation, tässä tapauksessa kaupungin, rooli on datan avaamisessa ja sen hallinnoinnissa. Yksi näkökulma on ajatella rooli samoin, kuin kaupungin muussakin toiminnassa. Yleisellä tasolla kaupunki rakentaa infrastruktuuria, joka mahdollistaa kaupunkilaisille ja yrityksille laadukkaat mahdollisuudet toimia. Samoin myös tulevaisuuden älykaupunkikonaisuudessa kaupunki rakentaa digitaalista perusinfrastruktuuria, kuten esimerkiksi Datajalostamo, joka mahdollistaa laadukkaan liiketoiminnan kehittämisen digitaalisen alustan avulla. Vaikka dataa kerätään ja hyödynnetään myös kaupungin omiin toimiin, on pääajatuksena kolmansien osapuolien mahdollisuus hyödyntää alustaa.

Dataa voidaan avata hyödynnettäväksi käytännössä kahdella eri tavalla, joko tiedostomuodossa tai API-rajapinnan avulla. Tiedostomuodossa avattu data on käyttäjien ladattavissa, mutta ei mahdollista aktiivista päivittymistä. API-rajapintaa hyödyntämällä mahdollistetaan jatkuvat päivittyminen ja näin ollen dataa voi hyödyntää esimerkiksi toisen ohjelmiston tai sovelluksen pohjalla. Rajapintojen hyödyntäminen voi kuitenkin olla haastavampaa, siinä missä tiedostomuotoista dataa voi hyödyntää useimmilla ohjelmilla suoraan, rajapinnan hyödyntäminen vaatii usein ohjelmointia. Esimerkiksi Kuopion Kaupungin Datajalostamo-alusta hyödyntää API-rajapintoja ja tarjoaa tätä kautta mahdollisuuden esimerkiksi sovellusten rakentamiseen datan hyödyntämiseksi.

Julkisen datan avaamisen suhteen valikointia on hyvä tehdä hyödyllisyysnäkökulmasta – minkä dataputken avaamiselle on suurin tarve ja selkeitä hyödynnettävyydenmahdollisuuksia. Dataputkea voikin pohtia lopusta alkuun, millaisen datan avaamiselle löytyisi selkeitä käyttötarkoituksia? Käyttötarkoitusten tunnistamisen jälkeen voidaan rakentaa malleja keräämiselle ja datan säilyttämiselle.

Avattua julkista dataa voivat hyödyntää toimijat sekä yksityisellä, että julkisella puolella. Julkisella puo-

lella avointa dataa voidaan hyödyntää esimerkiksi tiedolla johtamiseen, mutta myös avoimen datan päälle rakennetut sovellustoteutukset tuovat parhaimmillaan lisäarvoa kuntalaisten palvelutasoon. Useammassakin kaupungissa on yhtenä esimerkkinä avattu avoimen rajapinnan kautta lumiaurojen sijaintitiedot kuntalaisten hyödynnettäväksi. Vastaavan tyyppisiä hyödyntämismalleja voi hyvin löytää myös urheiluun ja liikuntaan linkittyvästä datasta.

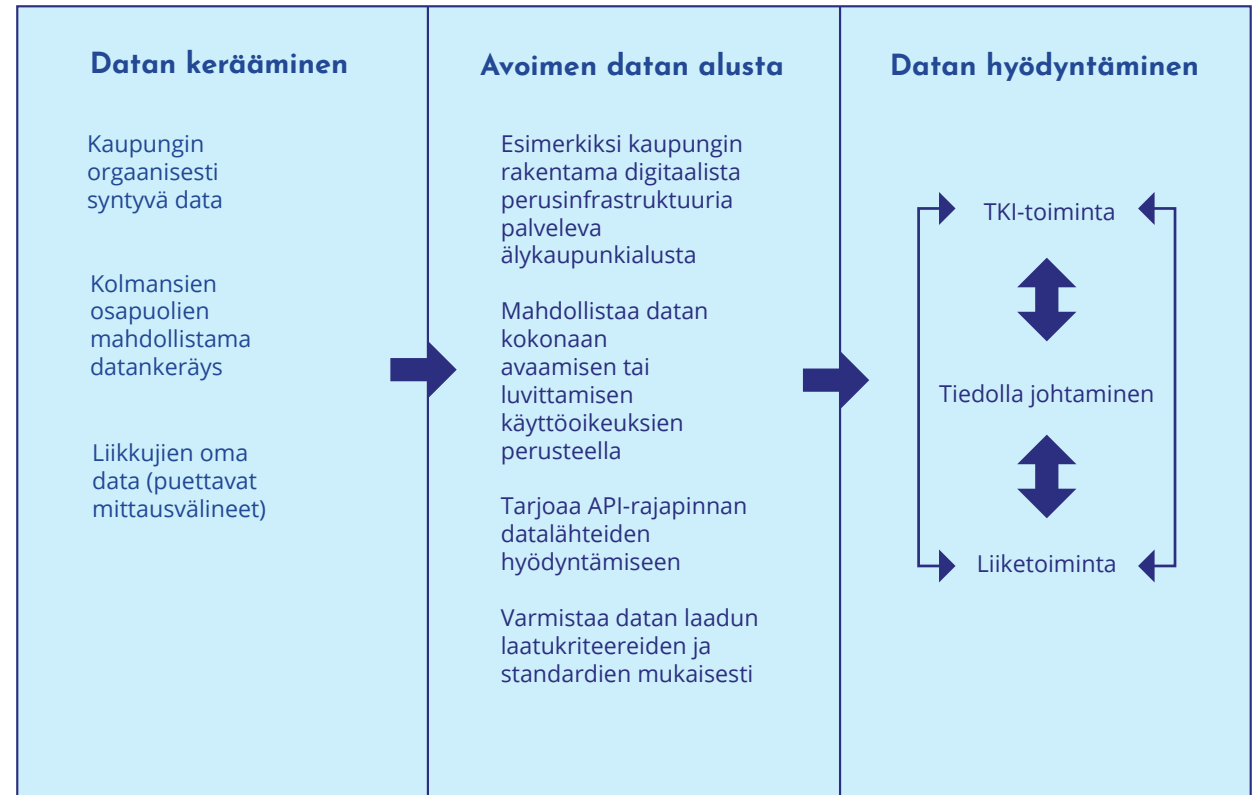
Datan avaaminen julkiseen käyttöön vaatii laajasti erilaisten vastuiden ja veloitteiden ymmärtämistä. Digi- ja väestötietovirasto (2023) on luonut tiedon jakamisen toimintamallin, joka koostuu kymmenestä huomioitavasta vaiheesta datan avaamiseen liittyen. Toimintamalli toimii hyvänä ohjenuorana esimerkiksi rooleille ja vastuille, joita tiedon avaamiseen liittyy.

Datan hyödyntämisen osalta on ymmärrettävä, että vain laatukriteerit täyttävä aineisto tarjoaa täyden mahdollisuuden hyödyntämiseen. Euroopan komissio (2020) on datastrategiassaan maininnut, että avattavat datat olisi hyvä noudattaa FAIR (findable, accessible, interoperable, re-usable) periaatteita. Aineiston on siis oltava löydettävissä yksilöllisellä tunnisteella ja kaikkien käyttävien

saavutettavissa. Yhteistoimivuus tarkoittaa, että tietoaaineistossa käytetään yleisesti ymmärrettävää kieltä ja metatiedot ovat saatavilla. Lisäksi aineiston on myös tärkeää olla uudelleen käytettävissä.

Paikallisella tasolla datainfrastruktuurin rakentamisessa ja datasäännösten rakentamisessa kannattaa tukeutua kansainvälisesti luotuihin laatukriteereihin ja ohjekokonaisuuksiin. Ohjeiden rakentaminen kansainvälisen viitekehyksen mukaisesti avaa myöhemmässä vaiheessa mahdollisuuden myös datan kansainväliseen ristiin hyödyntämiseen. Kansainvälisiin verrokkeihin vertaaminen tai paikallisen datan yhdistäminen laajempaan kontekstiin voi tuoda laajasti lisäarvoa kaikille osapuolille ja avata myös liiketoimintamahdollisuuksia.

Yhteentoimivuutta voidaan myös parantaa hyödyntämällä etenkin uusien palveluiden kehittämisessä valmiita tietojärjestelmien rakenteita ja sisältöjä sekä tiedon kansallisia ja kansainvälisiä standardeja.



Datatalouden säännöt ja oppaat

KUTEN AIEMMIN MAINITTU, data ja sen hyödyntäminen on ollut kasvavan mielenkiinnon kohteena myös laajemmassa mittakaavassa niin Suomessa, kuin myös koko EU:n alueella. Tämä on johtanut siihen, että on nähty tarpeelliseksi luoda erilaisia sääntöjä datatalouteen. Yksi esimerkki luodusta säännöstöstä on Sitran reilun datatalouden sääntökirja (2022). Vaikka jokaisella dataverkostolla onkin omat erityispiirteensä, kannattaa tietyissä periaatteissa ja säännöissä hyödyntää olemassa olevia perusmalleja. Yhteisen toimintamallien ja sääntöjen hyödyntäminen auttaa muun muassa yhteistyön tekemistä eri dataverkostojen kanssa. Lisäksi se parantaa yhdessä verkostossa kehitetyn tuotteen tai palvelun laajentamismahdollisuuksia.

Sitran reilun datatalouden sääntökirja (2022) tarjoaa oppaan dataverkoston perustamiseen ja sopimusten yleisten ehtojen määrittelyyn. Tämä on tärkeä vaihe dataverkoston perustamisessa, sillä yhteisillä säännöillä ja sopimuksilla voidaan varmistaa, että dataverkoston toiminta perustuu voimassa olevaan lainsäädäntöön ja esimerkiksi tietoturvaan sekä väärinkäyttöön liittyvät riskit on otettu tarpeellisella laajuudella huomioon.

Sitran lisäksi myös digi- ja väestötietovirasto (2023) on luonut yleiseen käyttöön avoimen datan tekni-

sen viitekehysten. Kyseinen viitekehys päivittyy säännöllisesti uuden tiedon mukaisesti, jotta dokumentti pysyy ajan tasalla uusimman kehityksen kanssa jatkuvasti muuttuvassa ympäristössä. Dokumentti sisältää tällä hetkellä kuvauksen FAIR-periaatteista, kitkattoman datapaketin mallista sekä viiden ja seitsemän tähden mallit datatalouteen liittyen. Tavoitteena näiden periaatteiden avulla on parantaa datan laatua sekä standardisoida sen rakenne, jotta käsittely ja hallinnointi onnistuu helpommin ja aina samoilla periaatteilla.

Yksi uusi ratkaisu datahallinnan alalla ovat ns. Data-avaruudet (data spaces). Sitran tulevaisuussanastossa data-avaruus määritellään yhteisesti sovittujen periaatteiden ja pelisääntöjen muodostamaksi kokonaisuudeksi, joka on tarkoitettu datan jakamiseen ja vaihtoon tietyllä toimialalla tai toimialojen välillä. Data-avaruudet liittyvät siis enemmän datan hallintaan yhteistyösopimusten eikä niinkään yhteisten data-alustojen avulla.

Vuonna 2020 hyväksytyssä Euroopan Datastrategiassa asetettiin tavoitteeksi yhteisten eurooppalaisten data-avaruuksien kehittäminen keskeisille toimialoille. Näiden data-avaruuksien tavoitteena on edistää uusien datavetoisten tuotteiden ja palvelujen kehittämistä EU:ssa ja muodostaa eurooppalaisen datatalouden ytimen. Käytännössä ne mahdollistavat eri toimijoille (yritykset, julkishallinnot ja yksityishenkilöt) heidän oman datansa jakamisen innovointitarkoituksiin turvallisesti, hallitusti ja itse datansa käyttöä valvoen. Yhteisiä eurooppalaisia data-avaruuksia on tällä hetkellä kehitteillä 14 eri osa-alueella: maatalous, kulttuuriperintö, energia, rahoitus, green deal, terveys, kieli, valmistusteollisuus, media, liikkuvuus, julkishallinto, tutkimus ja innovointi, osaaminen sekä matkailu (Euroopan komissio, 2024). Data-avaruuden kehittämiseen liittyvää tietoa on kerätty mm. The International Data Spaces Association:in (2024) laatimaan sääntökirjaan IDSA Rulebook.

- **Tietosuoja on keskeinen tekijä datan avaamisessa.**
- **Tietosuojalaki (1050/2018) ja GDPR ohjaavat henkilötietojen käsittelyä.**
- **Henkilötiedoksi luokitellaan kaikki tieto, joka voidaan liittää tunnistettavissa olevaan henkilöön.**
- **Synteettinen data, joka jäljittelee alkuperäisen datan ominaisuuksia ilman henkilötietoja, voi tulevaisuudessa auttaa tietosuojahaasteiden minimoimisessa.**

DATAN AVAAMISEN YHTEYDESSÄ on hyvin keskeistä huomioida tietosuoja ja sen vaikutus siihen, millaista dataa on mahdollista avata julkiseksi ja missä muodossa. Erityisesti tietosuojalaki (1050/2018) ohjaa mahdollisuuksia henkilötietojen käsittelyyn yhdessä EU:n yleisen tietosuoja-asetuksen (GDPR) kanssa (avoindata.fi, 2020).

Käytännössä avoimessa datassa henkilötietojen käsittelyssä täytyy noudattaa varovaisuusperustetta, suurilta osin sensitiivisen datan käsittely ei ole mahdollista. On myös tärkeää hahmottaa, missä laajuudessa tieto voidaan käsittää henkilötiedoksi, jotta riskejä voidaan arvioida. Lähtökohtaisesti kaikki tieto, joka voidaan liittää tunnistettavissa olevaan henkilöön, luokitellaan henkilötiedoksi (Tietosuojatyöryhmä, 2007). Vaikka data ei siis itsessään sisältäisi henkilötietoja, jos henkilö voidaan tunnistaa esimerkiksi yhdistämällä kokonaisuus johonkin toiseen tietoon, on kyseessä henkilötieto (Tietosuojavaltuutetun toimisto, n.d.) Esimerkiksi aiemmin mainitussa liikuntapaikkojen kävijädatassa yksittäinen lipputuotteen perusteella rekisteröity käynti liikuntapaikassa ei ole henkilötietoa, mutta jos käynti voidaan yhdistää tiettyyn lipputuotteeseen ja sen ostajaan, voidaan datapiste käsittää tietosuojan näkökulmasta henkilötiedoksi. Henkilötietoja käsitellessä täytyy myös huomioida

rekisteröidyn oikeudet. joita on esimerkiksi oikeus poistaa tiedot sekä rajoittaa tietojen käsittelyä.

Helpottamaan datan avaamisen tietosuojaa ja tietoturvallista avaamista, Espoon Kaupunki on luonut tähän liittyvän oppaan osana Suomen kuuden suurimman kaupungin 6aika -hanketta. Datan tietosuoja ja tietoturvallinen avaaminen -opas esittelee kuusiportaisen toimintamallin tarpeen tunnistamisesta datan julkaisemiseen ja ylläpitoon. Toimintamallin mukaan tarpeen tunnistamisen jälkeen datalle tehdään alustava hyöty-, riski- ja kustannusanalyysi, jonka perusteella voidaan suunnitella toimenpiteet, joita datan avaaminen vaatii. Espoon Kaupungin mallissa analyysi toteutetaan standardoidulla lomakkeella, jotta jokainen datalähde arvioidaan samoilla menetelmillä. Analyysilomake auttaa myös datan avaajaa ja tuottajaa arvioimaan dataansa ja tunnistamaan sen mahdollisia tietosuojariskejä. Analyysin arvioinnin jälkeen voidaan tehdä päätös datan julkaisemisesta tai sen julkaisematta jättämisestä. Jos päätös on julkaista data, täytyy vielä varmistua sen julkaisuvalmiudesta ja suunnitella ylläpito, jotta datalähde pysyy toiminnassa sekä ajantasaisena.

Yhtenä vaihtoehtona tietosuojahaasteiden minimoimiseksi esiin voi tulevaisuudessa nousta synteettisen datan hyödyntäminen. Ajatuksena synteettisessä datassa on luoda esimerkiksi tekoälyn avulla alkuperäistä data-aineistoa ja sen ominaisuuksia jäljittelevä data-aineisto. Tämä synteettisesti luotu data-aineisto jäljittelee alkuperäisen datan ominaisuuksia, mutta aineisto ei enää koostu oikeista ihmisistä, eikä siitä voi tunnistaa yksittäisen ihmisen henkilötietoja. (Honkela, 2023.) Synteettisen datan haasteena on se, kuinka aineistossa pystytään säilyttämään alkuperäisen datan ominaispiirteet tarpeeksi tarkasti niin, että tilastollisten päätelmien kautta ei voida tunnistaa yksittäistä henkilöä. Tähän dilemmaan on haettu vastauksia erilaisissa yhteyksissä. Esimerkiksi Jyväskylässä kaupungin ja yliopiston yhteisessä Wellbeing Data Lab -hankkeessa kehitetään liikunta- ja hyvinvointidatasta synteettistä dataa hyödynnettäväksi.

Esimerkki toimintamallista

- **Jatkuva prosessi:** Datan hyödyntämisen toimintamalli on kehä, jossa viimeisen vaiheen jälkeen palataan alkuun arvioimaan tuloksia.
- **Käyttötapauksen tunnistaminen:** Aloitetaan käyttötapauksen tunnistamisella ja laaja-alaisella pohtimisella, mikä auttaa tunnistamaan kumppaneita ja resursseja.
- **Strategiset tavoitteet:** Käyttötapauksen tulisi linkittyä strategiaan tavoitteisiin, kuten ulkoliikuntapaikkojen kävijämittaukseen resurssien suuntaamiseksi.
- **Tilannekuvan rakentaminen:** Tunnistetaan kaikki toimijat ja heidän tehtävänsä dataputken toiminnassa.
- **Pilotointi:** Suunnitellaan toiminta rajatun pilotin ympärille, jotta voidaan varmistua datan laadusta ja hyödyntämismahdollisuuksista.
- **Riskien ja kustannusten arviointi:** Pohditaan datan keräämisen riskejä ja kustannuksia sekä tarvittavia sopimuksia.
- **Monipuoliset käyttötarkoitukset:** Alkuperäisen käyttötarkoituksen ympärille voidaan pohtia useampia käyttötarkoituksia, mikä lisää datan arvoa.
- **Palautesykli:** Pilotoinnin jälkeen palataan aiempiin vaiheisiin saadun palautteen perusteella.

KÄYTÄNNÖN TASOLLA DATAN hyödyntämisen toimintamallia voidaan havainnollistaa jatkumona (kuva seuraavalla sivulla). Vaikka kuva selkeyttää toiminnassa vaadittavat askeleet, on askeleet hyvä ymmärtää lineaarisen jatkumon sijaan kehänä, jossa viimeisen vaiheen jälkeen palataan takaisin alkuun arvioimaan testauksen perusteella saatua informaatiota käyttötapauksen todellisesta potentiaalista.

Kuten jo aiemminkin tässä pelikirjassa mainittu, dataputken suunnittelu alkaa parhaassa tapauksessa käyttötapauksen tunnistamisella ja laaja-alaisella pohtimisella. Kun datan hyödyntämistä pohditaan tarveperustaisesti, päästään tunnistamaan mahdollisia kumppaneita, tarvittavia resursseja sekä hahmottamaan mahdollisia hyödyntäjätahoja kattavasti jo suunnitteluvaiheessa. Käyttötapauksen tunnistaminen linkittyy parhaassa tapauksessa myös johonkin asetettuun strategiseen tavoitteeseen, jotka ohjaavat toimintaa.

Käytännön esimerkki strategisesta tavoitteesta, joka ohjaa datan avaamista voisi tulevaisuudessa olla esimerkiksi ulkoliikuntapaikkojen kävijämittaus, jonka avulla voidaan suunnata resurssointia ja tehdä suunnitteluun liittyviä ratkaisuja käytön perusteella. Kun käyttötarve on tunnistettu, voidaan

pohtia kumppaneita toteuttamiseen, esimerkiksi datan keräämiseen liittyen. Jos kävijämittausta toteutetaan esimerkiksi laserkeilaustekniikalla, voidaan kävijämäärän lisäksi datasta saada tietoa käyttäjien liikkumisesta sekä alueen, esimerkiksi ulkokuntosalin, vilkkaimmista alueista. Tämän myötä alkuperäisen käyttötarkoituksen, tässä tapauksessa kävijämäärien mittaamisen, ympärille voidaan pohtia myös useampia käyttötarkoituksia, joka heti lisää arvoa datan avaamiselle. Ulkokuntosalin tapauksessa esimerkiksi kuntosalilaitteiden toimittajat voisivat avattua dataa hyödyntämällä tutkia, kuinka yksittäisen laitteen ulkonäön muokkaaminen vaikuttaa käyttöaktiivisuuteen.



Käyttötapausten tunnistamisen jälkeen on tärkeää rakentaa tilannekuva siitä, mitä kaikkia toimijoita kyseisen dataputken toimintaan liittyy ja millaisia tehtäviä linkittyvillä toimijoilla on. Kappaleen aloittavassa kaaviossa on mainittu tärkeimpiä kysymyksiä, joilla kokonaisuutta voi hahmottaa. Huomioon otettavia asioita on kokonaisuudessa lukuisia, mutta tärkeää on tunnistaa oleellimmat, jotka huomioonottamalla voidaan edetä pilottivaiheeseen. Suuri osa kysymyksistä voidaan käydä läpi data-alustan, esimerkiksi datajalostamon, käyttöohjeissa. Myös datan kerääjän ja loppukäyttäjän vastuulle jää myös pohdittavia asioita esimerkiksi datan riskeihin ja kustannuksiin liittyen.

Aiemmin mainitussa ulkokontosaliin liittyvässä esimerkissä loppukäyttäjien tunnistaminen on yksi asioista, jonka pohtiminen jo ennen pilotointia avaa lisämahdollisuuksia. Vaikka ensisijainen datan hyödyntäjä on kaupunkiorganisaatio, dataa voidaan hyödyntää myös tutkimuksessa tai laitetoimittajan taholta. Tämän ymmärtäminen avaa myös ymmärrystä siitä, millaisia sopi-

muksia tarvitaan ja minkä tahojen kanssa näitä on hyvä solmia. Lisäksi riskien pohdinta vaikuttaa siihen, millä tavoin datankeräys toteutetaan. Jos kävijädataa kerätään kameroilla, täytyy mahdollinen henkilötiedon syntyminen ottaa tarkasti huomioon. Laserkeilausta hyödynnettäessä dataa voidaan hyödyntää laajemmin ilman, että henkilötietoa syntyy. Hyödyntämismahdollisuuksia laajemmin pohtimalla voidaan siis löytää mahdollinen perustelu kustannuksiltaan korkeamman vaihtoehdon käyttämiseen datan keräämisessä.

Viimeinen porras toimintamallissa on pilotointi. On tärkeää huomioida, että hyvinkin suunniteltu dataputki ei välttämättä todellisuudessa anna teoriasa toivottuja tuloksia. Tämä voi johtua mittauksen epätarkkuudesta tai muusta haasteesta datan laadussa. Suunnittelemalla toiminta ensin rajatun pilotin ympärille voidaan varmistua datan laadusta sekä hyödyntämismahdollisuuksien laajuudesta. Prosessissa voidaan myös pilotissa saatujen havaintojen myötä palata aiempiin vaiheisiin.

Ahvonon, K. Bremer, O. Djakonoff, V. Koponen, J. Mikkonen, J. Toivanen, M. 2022. Suomen vahvuudet, haasteet ja mahdollisuudet datatalouden rakentamisessa. Sitra. <https://www.sitra.fi/julkaisut/suomen-vahvuudet-haasteet-ja-mahdollisuudet-datatalouden-rakentamisessa/>

Avoindata.fi. 2020. Selvitä rajoitteet ja riskit. <https://www.avoindata.fi/fi/opas/selvita-rajoitteet-ja-riskit>.

Digi- ja väestötietovirasto. 2023. Avoimen datan tekninen viitekehys. <https://www.avoindata.fi/fi/tietoa-avoimesta-datasta/avoimen-datan-tekninen-viitekehys#tekninen-viitekehys>

Digi- ja väestötietovirasto. (2023). Tiedon jakamisen toimintamalli. Versio 3. Saatavilla: <https://www.avoindata.fi/fi/toimintamalli>. Viitattu [päiväys].

Digitaalinen Helsinki. N.d. Helsinki-profiili (MyData-periaatteet). <https://digi.hel.fi/esittely/helsinki-datastrategia/helsinki-datastrategia-luku-3/luku-35/>

Espoon Kaupunki. Datan tietosuoja ja tietoturallinen avaaminen. N.d. https://static.espoo.fi/cdn/ff/zHVvWDY4KE_GVDUkXaEtQo6-xXkvh-jKBcZZewhqK840/1629461421/public/2021-08/sovellusesimerkki_Datan_tietosuoja_ja_tietoturallinen_avaaminen_ei%20saavutettava.pdf

Euroopan komissio. 2020. A European strategy for data. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:-52020DC0066&from=EN>

Euroopan komissio. 2023. Creating public sector value through the use of public open data. https://data.europa.eu/sites/default/files/report/Creating%20public%20sector%20value%20through%20the%20use%20of%20open%20data_ENG_0.pdf

Euroopan komissio. 2024. Yhteiset eurooppalaiset data-avaruudet. <https://digital-strategy.ec.europa.eu/fi/policies/data-spaces>.

Helsinki Institute of Physics. N.d. Mitä data on? <https://avoin-data-opas.github.io/data/mita.html>

Honkela, A. 2023. Synteettinen data voi suojata yksityisyyttä. Helsingin yliopisto. <https://www.helsinki.fi/fi/matemaattis-luonnontieteellinen-tiedekunta/ajankohtaista/synteettinen-data-voi-suojata-yksityisyytta>.

International Data Spaces. 2024. IDSA Rulebook. <https://docs.internationaldataspaces.org/ids-knowledgebase/idsa-rulebook>.

Irura, M. (2019). Open data and health. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), The state of open data: Histories and horizons (pp. 166–180). Cape Town and Ottawa: African Minds and International Development Research Centre. <http://stateofopendata.od4d.net>

Koski, H. Honkanen, M. Luukkonen, J. Pajarinen, M. & Ropponen, T. 2017. Avoimen datan hyödyntäminen ja vaikuttavuus. Valtioneuvoston selvitys- ja tutkimustoiminta. https://etene.fi/documents/10616/3866814/40_avoimen+datan+16032017.pdf

Kuopion Kaupunki. 2023. Dataohjelma. https://www.kuopio.fi/uploads/2024/01/kuopio_dataohjelma-1.pdf.

Lähteenoja, V. 2023. Mitä on datatalous. Sitra. <https://datataloudentiekartta.fi/mita-on-datatalous/>.

Mattila, Juri, Seppälä, Timo, Bützow, Alexander, Hynönen, Kalle & Puittinen, Mika. 2022. Teollisuuden sopimukset, uusi datasäädös ja digivihreä siirtymä. ETLA Muistio No 110., <https://pub.etla.fi/ETLA-Muistio-Brief-110.pdf>.

Mustonen, V., Koponen, J. & Spilling, K. 2014. Älykäs kaupunki – Smart City: katsaus fiksuihin palveluihin ja mahdollisuuksiin. https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/77892/Julkaisu_12-2014.pdf?sequence=1&isAllowed=y.

Opetushallitus. 2024. Miten käsitteet, kuten 'data', 'informaatio' ja 'tieto' liittyvät toisiinsa. <https://www.oph.fi/fi/digiosaaminen/datatalousosaamisen-perusteita-perusopetukseen-ja-toiselle-asteelle/mihin-0>.

Poikola, A. Wong, D. Lähteenoja, V. Turpeinen, M. 2024. Suomalaisen data-avaruuksien tilannekuva. Sitran selvityksiä. <https://www.sitra.fi/julkaisut/suomalaisten-data-avaruuksien-tilannekuva/>.

Sitra. 2022. Reilun datatalouden sääntökirja <https://www.sitra.fi/julkaisut/reilun-datatalouden-saantokirja/>.

Tietosuojatyöryhmä. 2007. Lausunto 4/2007 henkilötietojen käsitteestä. http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_fi.pdf.

Tietosuoja-valtuutetun toimisto. Henkilötietojen käsittely. N.d. <https://tietosuoja.fi/henkilotietojen-kasittely>.

Tietoyhteiskunnan kehittämiskeskus ry. 2021. Ehdotus toimintamallista API:n yhteiskehittämiseen. <https://tieke.fi/wp-content/uploads/2023/08/Toimintamalli-API.docx.pdf>.

Tilastokeskus. 2023. Data-analytiikkaa teki 41 % yrityksistä vuonna 2023. Data-analytiikkaa teki 41 % yrityksistä vuonna 2023 | Tilastokeskus.

Pelikirja on toteutettu osana kansallisen Sport Suomi -verkoston toimintaa. Verkosto perustuu Jyväskylän, Lahden ja Kuopion kaupunkien yhteiseen ja Euroopan aluekehitysrahastosta tuettavaan vetovastuuhankkeeseen, jota toteutetaan 1.4.2023 – 31.12.2024 välisenä aikana.