DATAHUONE

Datahuone tuo ajantasaisista viranomaisrekistereistä tehdyt analyysit nopeasti päätöksenteon tueksi.

Jari Järvenpää, Oskari Nokso-Koivisto ja Otto Toivanen

Helsinki Graduate School of Economics 2021

1. Johdanto, selvityksen tarkoitus ja toteutus sekä tärkeimmät johtopäätökset

Julkinen valta tekee jatkuvasti kansalaisten ja yritysten toimintaan vaikuttavia päätöksiä. Päätösten taloudellinen ja muu yhteiskunnallinen merkittävyys vaihtelee, kuten vaihtelee myös päätöksentekoon käytettävissä oleva aika. Aika ajoin suuriakin päätöksiä tehdään nopealla aikataululla.

Suomessa on poikkeuksellisen laadukas ajantasaisten, julkisten, yksilö- ja yritystasoisten rekisteriaineistojen perusinfrastruktuuri. Suuressa osassa päätöksentekoprosesseista ei kuitenkaan ole tällä hetkellä mahdollista hyödyntää näitä aineistoja. Tämä koskee erityisesti nopeita päätöksentekotilanteita ja tilanteita, joissa tarvittaisiin erityisen ajantasaisia aineistoja. Merkittäviä ongelmia ovat aiheuttaneet niin viiveet käytössä olevien aineistojen päivityksissä, pitkät lupaprosessit kuin osaavan henkilöstön puute päätöksiä valmistelevissa organisaatioissa.

Koronaan liittyvän päätöksenteon tueksi keväällä 2020 perustettu Helsinki GSE:n Tilannehuone on raapaissut ajantasaisten rekisteriaineistojen hyödyntämisen pintaa ja useilla esimerkeillä osoittanut, että nykyistä merkittävästi nopeampi, rikkaampi ja laadukkaampi datan hyödyntäminen niin hitaissa kuin nopeissakin päätöksentekotilanteissa on mahdollista. Toiminnan kehittämisen ja ylläpitämisen vaatimat taloudelliset resurssit ovat todennäköisesti hyvin vähäiset verrattuna datan tehokkaasta hyödyntämisestä kumpuaviin taloudellisiin säästöihin ja muihin hyötyihin.

Miltei välittömästi Helsinki GSE:n Tilannehuoneen ensimmäisten analyysien julkaisemisen jälkeen ministeriöiden edustajat nostivat esiin ajatuksen siitä, että Helsinki GSE:n Tilannehuone -tyyppinen, ajantasaisia ja yhdistettyä rekisteriaineistoa joustavasti päätöksentekijöiden tarpeiden mukaan analysoiva palvelu olisi tärkeä työkalu myös normaaliaikana. Vuoden 2020 lopussa valtiovarainministeriö tilasi Helsinki GSE:n Tilannehuoneelta selvityksen siitä, 1) tarvitaanko Tilannehuone-tyyppistä palvelua myös normaalioloissa ja jos tarvitaan, 2) miten palvelu voitaisiin normaalioloissa organisoida. Lisäksi sovittiin selvitettävän sitä, miten siirtymä Tilannehuoneesta uuteen organisoitumisen tapaan tehtäisiin ilman, että palvelu missään vaiheessa katkeaa.

Tämä selvitys perustuu yhtäältä vuoden aikana saatuun kokemukseen datan hankinnasta ja hyödyntämisestä sekä kokemuksiin jatkuvasta vuoropuhelusta ja erinomaisesta yhteistyöstä monenlaisten yhteiskunnallisten toimijoiden kanssa ja toisaalta laajasti toteutettuihin haastatteluihin. Olemme haastatelleet niin poliittisia päätöksentekijöitä kuin virkamiehiä ja yksityisen sektorin toimijoita.

Alusta asti on ollut selvää, että koronan aiheuttaman taloudellisen ja yhteiskunnallisen poikkeustilan tarpeisiin rakennettu Helsinki GSE:n Tilannehuone on väliaikainen ratkaisu. Ei ole perusteltua, että normaalioloissa opetukseen ja perustutkimukseen keskittyvä akateeminen yksikkö kantaisi merkittävää vastuuta rutiininomaisesta tilastotuotannosta tai päivittäisestä päätöksenteon tueksi suunnatusta data-analyysistä. Helsinki GSE onkin koko Tilannehuoneen olemassaolon ajan pyrkinyt irrottautumaan sellaisista Tilannehuoneen toiminnoista, jotka eivät akateemiseen yksikköön luontaisesti istu. Mikäli Tilannehuoneen kaltaista toimintaa halutaan jatkaa, on tärkeää, että Helsinki GSE:n hoitamien rutiinien siirtäminen saadaan käynnistetyksi kesällä 2021 ja vastuut on saatu siirretyksi kokonaisuudessaan vuoden 2021 loppuun mennessä.

Tämän selvitystyön aikana käytyjen keskusteluiden perusteella on selvää, että datan hyödyntämisen mahdollisuudet nähdään eri hallinnonaloilla ja hallinnon tasoilla suurina ja merkittävä tarve datapohjaiselle päätöksenteon tuelle tunnistetaan laajasti.

Esitämme

1. **perustettavaksi Datahuoneen**, jolla luodaan päätöksentekijöille uudentyyppinen mahdollisuus hyödyntää reaaliaikaista viranomaistoiminnasta syntyvää dataa monenlaisen päätöksenteon tukena. Datahuoneessa yhdistyvät Helsinki GSE:n Tilannehuoneen esimerkin mukaisesti uudella tavalla aineiston ajantasaisuus, yksilötasoisten aineistojen monipuolinen yhdistely, palvelun nopeus sekä näihin pohjaavan analyysin tuottaminen ja esittely päätöksentekijöille kullekin päätöksentekoprosessille sopivassa muodossa. Käytyjen keskusteluiden perusteella esitämme, että Datahuone olisi rakennettava hyödyntämään Tilastokeskuksen, VATT:in sekä Helsinki GSE:n infrastruktuureja ja osaamista ja että Datahuoneen hallinnollinen ydin rakennettaisiin valtioneuvoston kanslian ja valtiovarainministeriön yhteyteen. Perustelut tälle tavalle organisoitua esitetään alla, kuten myös arviot vaihtoehtoisista toiminnan järjestämisen tavoista.
2. **Tilastokeskuksen tutkijapalveluiden ajantasaistamista** palvelemaan julkisen sektorin ylimmälle päätöksenteolle suunnattujen nopeiden palveluiden Datahuoneen lisäksi laajempaa tieto-, analyysi- ja tutkimustarvetta sekä laajempaa organisaatioiden kirjoa ja päätöksentekijäkuntaa. Tutkijapalveluiden ajantasaistaminen sisältää tutkimuskäytössä olevien aineistojen laajamittaisen muuttamisen nopeasti käyttöön otettaviksi valmisaineistoiksi, näiden valmisaineistojen muuttamisen jatkuvasti ajantasaisiksi ja lupaprosessien sujuvoittamisen sellaiseksi, että aineistojen käyttöön oikeutetut tutkijat saavat aineistot käyttöönsä muutamassa viikossa.

Esityksemme Datahuoneesta ja Tilastokeskuksen tutkijapalveluiden ajantasaistamisesta ovat linjassa 2020 kesäkuussa raporttinsa jättäneen julkishallinnon aineistojen hyödyntämistä selvittäneen työryhmän (http://urn.fi/URN:ISBN:978-952-367-322-9) esitysten kanssa. Työryhmän raportti sisältää yksityiskohtaisia toimenpide-ehdotuksia siitä, miten Tilastokeskuksen tutkijapalveluita tulisi kehittää ja esityksemme rakentuu sen varaan, että ne toteutetaan. Keskitymme alla Datahuoneen perustamiseen liittyviin kysymyksiin.

2. Liitteet

Raportissa on seuraavat liitteet:

- PowerPoint-muotoinen kooste Helsinki GSE:n Tilannehuoneen toimesta vuoden aikana tehdyistä ja käynnissä olevista analyyseista
- Huhtikuun aikana valmistelutyön tueksi tehty tiivis taustamuistio Datahuoneesta
- Helsinki GSE:n, Tilastokeskuksen ja VATT:in näkemykset kunkin organisaation omasta roolista Datahuoneessa resurssitarpeineen. Kommentit pyydetty ko. tiiviiseen taustamuistioon.

3. Datahuone osana keskipitkän aikavälin muutosta

Määrittelemme tavoitteen seuraavasti:

Data on otettava tehokkaasti käyttöön kaikessa päätöksenteossa sellaisissa tilanteissa, joissa tietosuojasäännösten puitteissa datan käytöstä saatava hyöty on suurempi kuin datan käytön kustannukset.

Haastateltujen näkemykset tukevat näkemystämme yllä olevan tavoitteen mielekkyydestä. Käytössä olevan datan määrä ja laatu parantuvat jatkuvasti, datan hyödyntämisen osaaminen kehittyy ja kustannukset laskevat. Näin ollen datan hyödyntämisen kannattavuus kokonaisuudessaan parantuu, mistä puolestaan seuraa, että uusia mahdollisuuksia datan hyödyntämiselle aukeaa jatkuvasti. Datahuone sekä Tilastokeskuksen tutkimuspalveluiden käytettävyyden parantaminen onkin nähtävä prosesseina, jotka vaativat jatkuvaa kehittämistä.

Datan laajamittainen hyödyntäminen päätöksenteon tukena ei tarkoita ainoastaan sitä, että nykyiseen päätöksentekojärjestelmään lisätään yksittäinen dataelementti. Laajamittainen datan tehokas hyödyntäminen päätöksenteossa tarkoittaa todennäköisesti muutoksia niin päätöksentekoprosessiin kuin siinä vaadittaviin osaamistarpeisiinkin.

Datahuoneen perusrakenne on muodostettava sellaiseksi, että Datahuone olemassaolollaan edesauttaa datan hyödyntämisen laajentamista, eikä ainakaan luo tälle luonnolliselle kehitykselle esteitä.

Datan hyödyntämisen esteitä ovat:

- Datainfrastruktuurin puuttuminen tai hyödyntämisen näkökulmasta huono käytettävyys
- Päätöksentekijän näkökulmasta relevanttien datojen tai yksittäisten datan osien puuttuminen
- Päätöksentekijä ei ymmärrä datan mahdollisuuksia tai datan hyödyntämisen rajoitteita
- Data-asiantuntija ei ymmärrä päätöksentekoa ja sitä, mikä kulloinkin on päätöksenteon näkökulmasta kiinnostavaa
- Dataa hyödyntävien organisaatioiden osaamispuutteet
- Asiantuntijoiden pääsy dataan on tarpeettomasti rajoitettu, eli asiantuntijoiden käytettävissä oleva data ei ole niin ajantasaista tai informaatiosisällöltään laaja-alaista kuin se voisi olla.

Näiden ongelmien ratkaisemiseksi esitämme, että välittömästi käynnistyvässä ensimmäisessä vaiheessa, ensimmäisen vuoden aikana

- keskitettyyn Tilastokeskuksen hallinnoimaan datainfrastruktuuriin tehdään toiminnallinen päivitys siten, että kaikki datojen käyttöön oikeutetut tutkijat saavat nykyistä ajantasaisemmat ja monipuolisemmat datat selkeästi nykyistä nopeammin. Tämä edesauttaa datan hyödyntämisen skaalautumista ja luo pohjan datan käytölle päätöksentekojärjestelmässä.[1]

[1] Kaksi ensimmäistä toimenpidettä sisältyvät julkishallinnon aineistojen hyödyntämistä selvittäneen työryhmän (http://urn.fi/URN:ISBN:978-952-367-322-9) toimenpidesuosituksiin. Työryhmän kaikki 11 suositusta on perusteltua toteuttaa.

- tehdään valmisteilla olevat lainsäädännölliset muutokset, joilla selkeytetään tutkijapalveluiden roolia Tilastokeskuksen toiminnassa ja laajennetaan Tilastokeskuksen mahdollisuuksia kerätä aineistoja tutkimuskäyttöön.
- muutetaan Helsinki GSE:n Tilannehuoneen data-allas pysyväksi toiminnoksi. Data-allas sisältää tietosuojasäännösten sallimissa rajoissa kaiken saatavissa olevan datan ajantasaisena. Data-altaan sisältö (kenties vähäisin poikkeuksin) on tarjolla tutkijapalvelun lupaprosessin kautta kaikille tutkijoille ja muille käyttäjille jatkuvasti ja tiheällä frekvenssillä päivittyvinä valmisaineistoina.
- jatketaan ja vakinaistetaan Helsinki GSE:n Tilannehuoneen prosessi uuden datan hankinnasta ja siirretään siitä vastuu myöhemmin esitettävällä tavalla Datahuoneelle.
- pystytetään data-allasta hyödyntävä, ylintä päätöksentekoa palveleva, nopealla aikataululla ja tiiviissä vuorovaikutuksessa päätöksentekijöiden kanssa toimiva asiantuntijainfrastruktuuri. Tällä saadaan yhtäältä parannettua päätöksenteon laatua ja toisaalta jatkuvasti demonstroitua päätöksentekijöille datan käytön mahdollisuuksia ja ongelmia. Dataa uudella tavalla päätöksenteossa hyödyntävät päätöksentekijät saavat myös ensikäden tietoa datan hyödyntämisen mahdollisuuksista ja osaavat näin ollen ohjata datan hyödyntämisen kehittämistä oikeaan suuntaan. Vastaavasti data-asiantuntijat oppivat ymmärtämään paremmin, miten datalla ja siihen perustuvalla analyysillä voidaan tukea parhaiten tukea päätöksentekoa.

Ensimmäisen, vuoden kestävän, vaiheen aikana käynnistetään

- perusaineistojen määrällinen ja laadullinen parantaminen.
- hallinnonalakohtaiset hankkeet, joissa selvitetään ja toimeenpannaan datan hankinnan laajentaminen ja nopeuttaminen sekä systematisoidaan datan hyödyntämistä päätöksenteossa.
- käynnistetään dataosaajien rekrytoinnit ministeriöihin siten, että jokaisessa ministeriössä on rekrytointien jälkeen vähintään yksi Datahuoneen sisällä työskentelyyn kykenevä virkamies. Näin edesautetaan myös ministeriöiden osaamista Datahuoneen hyödyntämisessä sekä joudutetaan uuden datan hankintaa.
- luodaan tietosuojasäännösten sallimissa rajoissa mahdollisimman kattava avoimen rajapinnan tilastotuotanto. Tämä korvaa joiltain osin Tilastokeskuksen maksullista tilastotuotantoa.

Erityisen huolellisesti on keskipitkällä aikavälillä mietittävä, mikä osa data-analyysistä on perusteltua tehdä päätöksentekoyksikön sisällä, esimerkiksi ministeriöissä omien data-asiantuntijoiden toimesta, mikä osa on perusteltua tehdä keskitetysti Datahuoneen toimesta ja mikä osa analyysistä kannattaa ostaa ulkopuolisilta toimijoilta. Selvältä kuitenkin vaikuttaa, että muutoksen alkuvaiheessa, kun hajautettua kapasiteettia ei vielä juurikaan ole tarjolla, muutoksen käynnistäminen edellyttää riittävää panostusta keskitettyyn analyysikapasiteettiin. Myöhemmin, hajautetun kapasiteetin kasvaessa, keskitetyn yksikön keventäminen voi olla perusteltua.

4. Datahuoneen ensimmäisen vaiheen palvelut päätöksentekijöille

Datahuoneen ensimmäisen vaiheen palvelut voidaan jakaa suoraan ylimmille päätöksentekijöille suunnattuihin palveluihin sekä datainfrastruktuurin merkittävän parantumisen kautta kaikille julkisen ja yksityisen sektorin päätöksentekijöille suunnattuihin parantuneisiin palveluihin.

Molemmissa osissa tärkeänä elementtinä ovat Tilastokeskuksen tekemät merkittävät parannukset tutkijapalveluiden kautta käyttöön tarjottavassa datassa.

4.1 Datahuoneen suorat palvelut ylimmille päätöksentekijöille

Koska Helsinki GSE:n Tilannehuone jo tuottaa niitä palveluita, joita Datahuone tulee laajemmassa mitassa tuottamaan, on luontevaa rakentaa Datahuone Helsinki GSE:n Tilannehuoneen infrastruktuuria hyödyntäen. Näin varmistetaan myös se, ettei palveluiden tuotantoon tule katkosta. Ensimmäisessä vaiheessa Datahuone luo Helsinki GSE:n Tilannehuoneen mallin mukaisesti päätöksentekijöille uudentyyppisen mahdollisuuden hyödyntää viranomaisten toiminnasta syntyvää lähes reaaliaikaista dataa monenlaisen päätöksenteon tukena. Datahuoneessa yhdistyvät Helsinki GSE:n Tilannehuoneen toimintatavan mukaisesti aineiston ajantasaisuus, yksilötasoinen aineistojen yhdistely, palvelun nopeus sekä näihin pohjaavan analyysin tarjoilu päätöksentekijöille kullekin päätöksentekoprosessille sopivassa muodossa. Koronan taloudellisia vaikutuksia tutkimaan perustettu Helsinki GSE:n Tilannehuone toimii Datahuoneen prototyyppinä ja suuri osa Tilannehuonetta varten tehdystä työstä ja tavasta organisoitua on suoraan hyödynnettävissä Datahuoneen rakentamisessa.

Datahuoneen palvelut:

1. Nykyistä paremmin päätöksentekijöiden aikatauluihin mukautuva **pikaselvitystuotanto** siten, että helpoimpiin kysymyksiin kyetään vastaamaan erittäin nopeasti, jopa tunneissa ja hiukan monimutkaisempiin kysymyksiinkin muutamissa päivissä tai viikoissa. Data-analyysin lisäksi keskeinen osa uutta palvelua on analyysin tulosten läpikäynti yhdessä päätöksentekijöiden kanssa.
2. Dataa tehokkaasti hyödyntävän päätöksenteon yksittäisten tapausten **yhteiskehittäminen** kunkin päätöksenteon kannalta relevantin päätöksentekotahon kanssa.
3. Nykyistä ketterämmin päätöksentekijöiden tarpeisiin vastaava ja nopeasti kehittyvä **joustava tilastotuotanto**, jonka osana on keskeisiin aineistoihin tietosuojasäännösten puitteissa tarjottava avoin rajapinta.

Datahuoneen palvelut suunnataan ministeriöille ja niiden alaisille organisaatioille, eduskunnan valiokunnille, muille julkisen sektorin toimijoille, kuten kunnille sekä soveltuvin osin myös yksityiselle sektorille. Perustamisen jälkeen palvelutoiminnan mahdollisesti kysyntävetoisesti laajetessa siitä aiheutuvat muuttuvat kustannukset kerätään yhteistyökumppaneilta.

Käymme seuraavaksi läpi Datahuoneelle suunnitellut, yllä luetellut palvelut.

1. Pikaselvitykset

Datahuone tarjoaa keskeisille päätöksentekijöille mahdollisuuden datan lähes reaaliaikaiseen hyödyntämiseen. Käytännössä tämä tarkoittaa yksinkertaisissa kyselyissä muutaman tunnin vasteaikaa, jolloin dataa voidaan hyödyntää osana nopeatempoistakin neuvotteluprosessia. Lyhyellä vasteajalla on mahdollista tehdä yksinkertaista kuvailevaa analyysiä valmiista aineistosta. Käytännössä näissä tapauksissa tuotetaan yksinkertaisia kuvaajia ilman tekstimuotoista selitystä.

Datahuoneen tuottaa muutaman päivän tai viikon vasteajalla edellistä monimutkaisempaa analyysiä, jossa voidaan yhdistellä olemassa olevia aineistoja ja mahdollisesti hyödyntää yksinkertaisia malleja. Muutaman viikon tai kuukauden vasteajalla on mahdollista hankkia ja yhdistellä uutta dataa ja tehdä edellisiä syvällisempää analyysia. Hyvin lyhyellä toimitusajalla

tehtäviä analyysejä ei ole mahdollista kilpailuttaa, vaan ne on perusteltua tuottaa keskitetysti julkisen sektorin omana toimintana.

Tärkeä pikaselvitysten onnistumisen edellytys on päätöksentekijöiden ja Datahuoneen asiantuntijoiden saumaton vuorovaikutus alkaen analyysitarpeen ja datan mahdollisuuksien molemminpuolisesta ymmärtämisestä aina tulosten avaamiseen päätöksentekijöille. Kaikki tämä tapahtuu suullisesti tiiviissä ja jatkuvassa dialogissa.

2. Datan hyödyntämisen yhteiskehittäminen

Kaikki päätöksentekoprosessien dataa hyödyntävät kehittämismahdollisuudet eivät mahdu tilastotuotannon, pikaselvitysten tai muiden erilaisten selvitysten teettämisen piiriin. Tällaisten ongelmien ratkaisussa erityistä huomiota on kiinnitettävä mahdollisuuksien tunnistamiseen, dataosaajien ja päätöksentekijöiden vuorovaikutuksessa tapahtuvaan ongelman ratkaisuun ja ratkaisun mahdolliseen skaalaamiseen muille samasta ongelmasta kärsivälle.

Datahuone yhtäältä etsii itse tällaisia mahdollisuuksia ja on aktiivisesti yhteydessä päätöksentekijöiden hanke-ehdotusten muodossa. Toisaalta Datahuone on avoin erilaisten päätöksentekotahojen yhteydenotoille. Yhteiskehittäminen luo paitsi lisäarvoa parantuneiden Datahuoneen palveluiden kautta, myös sitä kautta, että näin kehitetään uusia aineistoja, jotka liitetään osaksi kaikille tutkijoille avoimia Tilastokeskuksen valmisaineistoja. Avaamalla tällaiset aineistot laajemmin tutkijoiden käyttöön varmistetaan niistä saatavan yhteiskunnallisen hyödyn kasvu.

Datan hyödyntämisen yhteiskehittämistä ei suurimmassa osassa tapauksia ole perusteltua tehdä keskitetysti. Datahuoneen tärkein rooli onkin yhteiskehittämisen mahdollisuuksien etsimisessä ja tutkijoiden ja päätöksentekijöiden kohtaamisen edistämisessä ja mahdollisten dataan liittyvien ratkaisujen edistämisessä.

3. Joustava tilastotuotanto

Datahuoneen joustavan tilastotuotannon tavoitteena on julkaista kerran viikossa jatkuvasti sisällöllisesti kehittyvä yhteiskunnan eri sektoreiden tilasta. Raportti perustuu erilaisten käyttäjien tarpeille ja hyödyntää monimuotoisesti Datahuoneen datoja. Sitä laajennetaan, kun uusia datalähteitä saadaan ja tarpeita ilmenee. Datahuone kartoittaa uusia joustavan tilastoinnin tarpeita päätöksiä tekevien tahojen kanssa.[2]

Toisena jatkuvan tilastotuotannon palveluna Datahuone tarjoaa riittävällä tavalla tiivistetyn rekisteriaineiston julkisesti jaettavaksi, jolloin palvelun käyttäjä voi tehdä itse esimerkiksi vertailuita.[3] Näiden palveluiden osalta tulee mietittäväksi työnjako Tilastokeskuksen ja Datahuoneen välillä.

[2] Esimerkiksi kuntien verotulojen ennakointia on haitannut se, että kunnat eivät ennen kunkin vuoden huhtikuuta ole saaneet tietoa kuntalaisten edellisenä vuona saamista palkkatuloista. Tulorekisteristä tämä tieto on nopeasti ja luotettavasti kuntakohtaisesti poimittavissa ja julkaistavissa. Datahuone edesauttaa näiden matalalla roikkuvien hedelmien poimimista.

[3] Esimerkkinä toiminnalle toimii VATT:in Tomi Kyyrän Tilannehuoneessa tekemä työkalu: https://tkyyra.shinyapps.io/earnings_fi/

4.2 Datahuonehankkeen seurauksena muodostuvan palvelumarkkinan kautta syntyvät palvelut päätöksentekijöille

Datahuonehankkeen keskitetty toiminta ja maan ylintä päätöksentekoa varten luotu koneisto luovat rungon datan hyödyntämisen kehittämiselle. Tätä keskitettyä datan hyödyntämistä keskipitkällä ja pitkällä aikavälillä todennäköisesti merkittävämpi yhteiskunnallinen hyöty syntyy parantuneen datainfran tarjoamisesta aukeavien mahdollisuuksien kautta.

Tavoitteena on, että ensimmäisessä vaiheessa laadullisesti ja määrällisesti parantuneet datat saadaan viedyksi Tilastokeskuksen tutkijapalveluiden nopeutuneen protokollan kautta Tilastokeskuksen sääntöjen puitteissa tarjolle kaikille tutkijoille ja sitä kautta tukemaan laajasti monenlaista päätöksentekoa eri sektoreilla ja päätöksenteon tasoilla. Parantuneesta datainfrastruktuurista pääsee tutkijapalveluiden säännösten puitteissa hyödyntämään julkisen sektorin toimijoiden lisäksi myös yksityinen sektori.

5. Datahuoneen taloudelliset vaikutukset

Datahuone luo merkittäviä julkisen talouden tehostamismahdollisuuksia niin lyhyellä, keskipitkällä kuin pitkälläkin aikavälillä.

Datan hyödyntämisen julkistaloudelliset hyödyt voidaan jakaa seuraaviin kategorioihin:

- Taloudellisesti merkittävien päätösten tarkentumisesta johtuvat suorat säästöt.

  Esimerkki 1: Koronakriisin aikaisten kustannustukien taustalla olevien laskelmien tekeminen ei olisi onnistunut tai olisi merkittävästi vaikeutunut ilman Tilannehuoneen olemassaoloa ja sinne luotua ajantasaista dataa, jossa yhdistettiin eri lähteistä tulevat aineistot.

  Esimerkki 2: Yleisesti uskotaan, että korona on heikentänyt oppimista eri koulutusasteilla. Koulutuksesta vastaava ministeri on esittänyt satojen miljoonien eurojen panostamista ongelman korjaamiseen. On todennäköistä, että eri väestöryhmiin koronan vaikutukset ovat hyvin erilaisia. Datahuonetta hyödyntämällä voidaan nykytilannetta merkittävästi tarkemmin arvioida oppimisvajeen suuruus eri väestöryhmissä sekä suunnata nykytilannetta tarkemmin päätetty tuki sellaisille ryhmille, jossa oppimisvajetta on. Tällöin tavoiteltuun lopputulokseen päästään aiottua pienemmällä summalla.

- Yhteiskunnallisesti merkittävien päätösten suorat vaikutukset.

  Esimerkki: Jatketaan edellistä esimerkkiä oppimisvajeesta. Koulutus parantaa yksilön tuottavuutta. Sellainen oppimisvajeen korjaaminen, jota ei ole tunnistettu, on vaikeaa. Yhteiskunta saa taloudellista etua, mikäli kyetään tunnistamaan nämä ryhmät ja reagoimaan oikein.

- Epävarmuuden pieneneminen.

  Esimerkki: Nykytilanteessa kuntien kyky ennakoida verotulojensa kehitystä on huono. Datahuone voisi pienin kustannuksin tarjota kuntakohtaisen palkkatuloennusteen kuntien päätöksentekijöille. Tällainen epävarmuuden pieneneminen luo nykyistä paremman pohjan kuntien taloudenpidolle.

- Valmistelun kustannusten laskeminen

  Mikäli valmistelun laatu halutaan pitää vakiona, hyvin toimiva Datahuone voi tehostaa valmistelua ja laskea valmistelun kustannuksia eri päätöksenteon tasoilla.

Datapohjaisen päätöksenteon tuen toiminta ja kustannukset voidaan jakaa mielekkäästi kahtia: keskitettyyn ja hajautettuun.

Keskitetysti on perusteltua tehdä esimerkiksi:

- Teknisen infrastruktuurin ylläpito ja kehitys.
- Datan ylläpito, uuden datan hankinta ja lupien hallinnointi.
- Pikaselvitysvalmiuden ylläpito ja pikaselvitysten laatiminen sekä tähän liittyvä priorisointi.
- Tilastotuotannon ylläpito ja kehitys.
- Ulkoinen viestintä, esimerkiksi median palveleminen.
- Kokonaisuuden kehittäminen ja toimijoiden välinen koordinaatio.

Nämä kustannukset ovat luonteeltaan **kiinteitä** ja ne on perusteltua kattaa suoraan julkisista varoista. Suurin osa niistä (teknisen infrastruktuurin ja datan ylläpito) palvelevat julkisen datan saatavuuden parantamista yleisesti, eivät vain Datahuonetta. Yksinomaan Datahuoneeseen liittyvät kustannukset ovat hyvin maltilliset.

Näiden palveluiden määrät päätetään hallinnollisesti. Ei ole poissuljettua, etteikö tätä kiinteää kustannusta jakamaan saataisi valtion lisäksi muitakin merkittäviä datainfran hyödyntäjiä, kuten esimerkiksi kuntia. Merkittävin taloudellinen hyöty Datahuoneesta saadaan tilanteessa, jossa Datahuoneen infrastruktuuri saatetaan kaikkien relevanttien tahojen käyttöön, jolloin dataa hyödyntävän päätöksenteon hyödyt voidaan saattaa laajemmin eri toimijoiden ulottuville.

Toinen osa kustannuksista muodostuu **muuttuvista** kustannuksista, jotka on perusteltua kattaa asiakasmaksuilla:

- Joustavan tilastotuotannon lisäpalvelut.
- Hitaampien selvitysten ja laajempien yhteiskehittämishankkeiden suunnittelu ja toteuttaminen sekä näistä hankkeista nousevat muuttuvat kustannukset.
- Merkittävästi nykyistä ajantasaisemman ja nopeammin käyttöön otettavan datan tarjoaminen kaikille relevanteille tutkijoille ja muille dataan oikeutetuille tahoille.

Näiden palveluiden määrät on skaalattava kysynnän mukaan.

6. Datahuoneen toimintojen organisointi

Datahuoneen organisoinnin suunnittelussa tulee ottaa sekä lyhyt että pitkä aikaväli huomioon.

Kuten yllä olevasta on käynyt ilmi, Datahuoneessa tarvitaan useita kyvykkyyksiä. Käymme nämä seuraavaksi läpi ja mikäli mahdollista, esitämme vaihtoehtoisia ratkaisuja jokaiseen. Sen jälkeen kuvaamme vaihtoehtoiset siirtymäpolut nykytilasta uuteen toimintamalliin.

6.1. Datahuoneen edellyttämien kyvykkyyksien organisointivaihtoehdot

1. **Datainfrastruktuuri**. Tehtyjen haastatteluiden ja saatujen kokemusten perusteella näyttää selvältä, että Tilastolaki pienin, jo valmisteilla olevin, muutoksin antaa hyvän pohjan Datahuoneen datainfrastruktuurin rakentamiselle. Tilastokeskus olisi luonteva sijoituspaikka datainfrastruktuurille. Käytäntö on ollut hyvin toimiva myös Helsinki GSE:n Tilannehuoneen tapauksessa. Luontevia muita ratkaisumalleja datainfrastruktuurin tarjoajiksi ei ole. Finndata on vastaperustettu ja luomassa käytänteitään; se on myös erikoistunut terveyssektorin aineistoihin, toisin kuin Tilastokeskus. Ministeriöillä itsellään ei ole vaadittavaa kyvykkyyttä.

2. **Kyky tuottaa aineistoihin perustuvia kevyitä analyyseja erittäin nopealla (tuntien ja päivien sisällä) että syvällisempiä analyysejä nopealla (muutama viikko) aikataululla.** Tällä hetkellä tämä kyky on vain Helsinki GSE:n Tilannehuoneella, mutta valtion organisaatioista VATT:lla on mahdollisuus ottaa vastuu Datahuoneelle tulevasta, sektorirajat ylittävästä nopeiden analyysien tuottamisvastuusta. Tulevaisuudessa työnjakoa keskitetysti tehtävän analyysin, ministeriöiden itselleen hankkiman analyysiosaamisen ja tämän osaamisen puitteissa tekemän analyysin sekä kilpailun pohjalta tilattavien tutkimusten ja selvitysten välillä tulee kehittää. Vaihtoehto VATT:lle olisi kehittää yksinomaan ministeriöiden omaa kykyä tuottaa tällaisia analyysejä. Tässä näemme kuitenkin kolme ongelmaa: Ensinnäkin, ministeriöillä ei tällä hetkellä ole tällaista kyvykkyyttä, joten toiminnan jatkuvuus vaarantuisi. Toiseksi, hajauttamalla toiminta ministeriöihin heikentyisi sektorirajat ylittävä ymmärrys aineistojen luomista mahdollisuuksista tukea päätöksentekoa. Kolmanneksi, ministeriöihin hajautettu analyysivastuu jättäisi muut hallinnon tahot selkeästi heikompaan asemaan. Näkemyksemme onkin, että paras tie eteenpäin on kytkeä VATT Datahuoneen toimintoihin samalla, kun ministeriöiden analyysi- ja analyysinhyödyntämiskykyä kehitetään.

3. **Kyky tuottaa aineistoihin perustuvia tilastoja**. Tällä hetkellä tähän kykenevät erityisesti Tilastokeskus, mutta myös Helsinki GSE:n Tilannehuone. Jatkossa päävastuu olisi luontevaa siirtää Datahuoneelta Tilastokeskukselle ainakin tilastotuotantoon menevien aineistojen osalta. Muiden aineistojen osalta VATT on vaihtoehto.

4. **Toiminnan organisointikyky ja kyky varmistaa keskusteluyhteys päätöksentekijöiden ja data-asiantuntijoiden välinen yhteistyö** niin analyysien teossa, tulosten hyödyntämisessä kuin analyysien ja aineistojen kehittämisessä. Helsinki GSE:n Tilannehuone on luonut toimintatapoja yhteistyöhön eri ministeriöiden kanssa, mutta on selvää, että toiminnan laajentaminen edellyttää vankempaa organisatorista osaamista. Vaihtoehtoina tässä ovat lähinnä VNK ja VM. VNK:n puolesta puhuu erityisesti tarve varmistaa sektorirajat ja hallinnon eri tasot ylittävä yhteistyö. VM:n puolesta puhuu sen ylivaalikautinen toimeenpanokyky ja ministeriön virkamiesten omasta osaamisesta kumpuava ymmärrys datan hyödyntämisestä. Luontevin ratkaisu voisikin olla, että VNK ja VM yhdessä ottavat organisatorisen vastuun kantaakseen.

5. **Kyky varmistaa toiminnan laatu ja riippumattomuus**. Yksilötasoisen yhteiskunnallisen datan hyödyntäminen Datahuoneen tavoin on kansainvälisesti katsoen uraa uurtavaa. Datan analysoinnissa ja analyysien tulkinnassa on merkittävä riski väärintulkinnoille ja muille virheille. Sekä tähänastinen kokemus että haastattelut tukevat sitä näkemystä, että Datahuoneen toimintojen kehittäminen edellyttää vahvaa ja jatkuvaa akateemisen

osaamisen kytkentää toimintaan ja sen kehittämiseen. Tämä suoraan päätöksentekoa tukevaan analyysitoimintaan kytkeytyvä kehittäminen on syytä erottaa siitä akateemiselle yhteisölle luontaisesti lankeavasta vastuusta ja mahdollisuudesta käyttää parantuneita tilastoaineistoja tutkimukseen. Edellytyksiä sille, että akateeminen taho voi tällaisen vastuun ottaa ovat sekä kyky (ja halu) sektorirajat ylittävään kysymyksenasetteluun, että riittävä yksikön koko. Ainoa tällainen yksikkö Suomessa on VM:n, OKM:n ja Suomen Pankin myötävaikutuksella Aalto-yliopiston, Helsingin yliopiston ja Hankenin yhtistyössä perustama taloustieteen tutkimuksen ja opetuksen valtakunnallista vastuuta kantava Helsinki GSE. Vaihtoehto sille, että vastuu akateemisesta laadunvarmistuksesta, ja riippumattomuuden takaamisesta annetaan jollekin yksikölle, on antaa se joukolle akateemisia tutkijoita, jotka valitaan tehtävään määräajaksi. Tässä ratkaisussa on kaksi huonoa puolta: yhtäältä valtiovallan kyky akateemisten kyvykkyyksien arviointiin on huono, toisaalta ratkaisu henkilöpohjaisena ei takaisi jatkuvuutta tai akateemisen yhteisön laajempaa sitoutumista. Toimittaessa institutionaalisessa yhteydessä Datahuoneen vastuut voidaan kytkeä kiinteäksi osaksi Helsinki GSE:n akateemista ydintoimintaa.

6.2. Vaihtoehtojen vertailua

Datahuoneen esikuvana toimivan Helsinki GSE:n Tilannehuoneen käynnistysprosessissa keväällä 2020 selvitettiin nopealla aikataululla erilaisia oikeudellisia, hallinnollisia ja teknisiä vaihtoehtoja Tilannehuoneen perustaksi. Ennakkoarvioista poiketen näiden selvitysten tulos oli, että Tilannehuone on oikeudellisesti, hallinnollisesti ja teknisesti mahdollinen. Käynnistämisprosessin aikana ei löydetty muita oikeudellisesti, hallinnollisesti ja teknisesti mahdollisia ratkaisuja.

Tiivistetysti, Helsinki GSE:n Tilannehuone on organisoitu seuraavasti:

1. Tilannehuone on tutkimusprojekti Tilastokeskuksen tutkijapalveluissa. Sillä on tavallista tutkimushankkeita selvästi
    a. laajemmat ja ajantasaisemmat aineistot käytössä
    b. enemmän tutkijoita
    c. paremmat ja nopeammat keinot saada analyysien tulokset ulos tutkijapalveluista
    d. tutkijapalveluiden erikoispalvelut (esim. oma sähköpostiyhteys ongelmatilanteita varten, nopeutettu käyttölupien käsittely, nopeutettu aineistojen saanti)
2. Tutkijapalvelut (yhdessä CSC:n kanssa) ovat allokoineet sekä työvoimaa että tietokonekapasiteettia Tilannehuonetta varten
3. GSE:n ja VATT:n tutkijat muodostavat neljä tutkimusryhmää, joilla on omat vastuualueet tilastotuotannosta ja analyyseistä. Tutkimusryhmät ovat
    a. Kotitaloudet
    b. Terveys
    c. Työmarkkinat
    d. Yritykset
4. GSE:n palkkaama projektipäällikkö hallinnoi Tilannehuoneen käytännön toimintaa ja ohjaa tutkimusapulaisia. Tutkimusapulaiset ovat GSE:n opiskelijoita, jotka on palkattu Tilannehuoneeseen. GSE:n muut projektipäälliköt osallistuvat Tilannehuoneen toimintaan tarpeen mukaan (esim. lupa- ja aineistoasiat).
5. GSE:n ja VATT:n professorit muodostava ohjausryhmän, joka johtaa ja kehittää Tilannehuoneen toimintaa.

Selvitystyön perusteella summaamme näkemyksemme vaihtoehdoista seuraavasti:

- Emme näe mahdolliseksi, että edes keskipitkällä aikavälillä Datahuone voisi, tai Datahuoneen olisi mielekästä, toimia merkittävästi Tilannehuoneesta poikkeavalla teknisellä tai oikeudellisella pohjalla.
- Emme myöskään näe lyhyellä aikavälillä vaihtoehtoa VATTille keskitetyn palveluntuotannon järjestämisestä. Keskipitkällä aikavälillä, ministeriöiden ja muiden hajautettujen kyvykkyyksien lisääntyessä, keskitetyn palvelutuotannon, ja siten myös VATT:in rooli voi olla perusteltua uudelleen arvioida. On kuitenkin selvää, että rajanveto keskitetysti tilattavien toimien ja kilpailun kautta hankittavien tutkimusten ja selvitysten välillä on harkittava huolellisesti, ja sitä tulee kokemusten myötä tarkistaa niin, että Datahuone luo tarpeettomasti monopoliasemaa ydintoimijoilleen tai kilpailuetua toiminnoissa jotka olisi mielekästä kilpailuttaa.
- Emme näe keskipitkällä aikavälillä mahdolliseksi sitä, että Suomeen syntyisi Helsinki GSE:tä vastaava valtakunnallinen akateeminen taloustieteen toimija, joka voisi ottaa Datahuoneessa Helsinki GSE:lle suunniteltuja vastuita. Huomionarvoista on, että datainfrastruktuurin ydin on esityksessämme suunniteltu sellaiseksi, että kaikki akateemiset ja ei-akateemiset tutkimusyksiköt ovat Helsinki GSE:n kanssa samalla viivalla hyödyntämässä Tilastokeskuksen muutoksen vuoksi syntyneitä parantuneita aineistoja. Keskeistä akateemisen yhteisön itsensä ja sen yhteiskunnallisen vuorovaikutuksen näkökulmasta on, että Tilastokeskuksen tutkijapalvelut uudistetaan tässä muistiossa esiin tuotujen esitysten mukaisesti. Kuten tässä muistiossa esitetään, tärkeää on, että myös myöhemmin Datahuoneelle luotavat uudet aineistot (kenties vähäisin poikkeuksin) automaattisesti siirretään osaksi Tilastokeskuksen tutkimuspalveluiden jatkuvasti päivittyviä valmisaineistoja. Näin taataan se, että Datahuoneen ydintoimijoille ei synny kannustinta toimia portinvartijana suhteessa aineistojen käyttöön ja kaikilla asiaankuuluvat kriteerit täyttävillä tutkijoilla on yhdenvertainen pääsy aineistoihin.
- Emme näe epäselvyyttä siinä, tarvitaanko valtionhallinnossa sisällä olevaa hallinnollisesti Datahuonetta johtavaa toimijaa. Haastatteluissa on esitetty erilaisia näkemyksiä siitä, millainen rooli VNK:lla ja VM:llä toiminnassa pitäisi olla.

6.3. Hahmotelma datahuoneen organisaatiosta ja vastuista

Edellisen perusteella summaamme näkemyksemme Datahuoneen toivottavasta organisoinnista seuraavasti.

Datahuoneen johtoryhmä

- kokoonpano: 3 Helsinki GSE:n nimeämää jäsentä, joista yksi toimii puheenjohtajana, 1 VNK:n nimeämä jäsen, 1 VM:n nimeämä jäsen ja 1 OKM:n nimeämä jäsen.
- Tilastokeskuksen ja VATT:in edustajat toimivat johtoryhmän pysyvinä asiantuntijoina.
- vastaa toimeksiantojen priorisoinneista
- päättää ja valvoo rahoituksen käyttöä

Datahuoneryhmä

- ryhmään kuuluu kolme täysipäiväistä valtioneuvoston kansliaan ja valtiovarainministeriöön sijoitettua virkamiestä
  - toiminnanjohtaja, joka toimii myös johtoryhmän sihteerinä
  - ministeriö- ja muista asiakassuhteista vastaava henkilö
  - tuotannosta vastaava henkilö

- Datahuoneryhmään kuuluvien VNK:n ja VM:n virkamiesten lisäksi sitä tukevat Tilastokeskuksen ja VATT:in nimeämät vastuuhenkilöt sekä tarvittavilta osin Helsinki GSE:n edustaja
- vastaa Datahuoneessa tapahtuvan datatyön ja päätöksenteon yhteistoiminnan edistämisestä
- osallistuu datan hankintaan

Helsinki GSE

- tekee Datahuoneen dataa hyödyntäen akateemista tutkimusta
- tekee ja edistää yksityiskohtaisia ja yleisiä aloitteita Datahuoneen kehittämiseksi
- tukee Datahuoneen toiminnan ja tutkimuksen korkeaa laatua ja toimii osaltaan laadun ja riippumattomuuden varmistajana

Tilastokeskus

- vastaa teknisestä infrastruktuurista ja datainfrastruktuurista Datahuoneen tilaamassa laajuudessa
- laajentaa voimakkaasti kaikille tutkijoille tarjottavia valmiita ja jatkuvasti päivittyviä tilastoaineistoja, jolloin ajantasaisia aineistoja saadaan Datahuoneen ohella muidenkin tutkijoiden käyttöön nykyistä kattavammin ja nopeammin.
- vastaa Tutkijapalvelun nykyisen toimintamallin mukaisesti datan hankinnasta, datapäivityksistä ja lupien hallinnoinnista
- tuottaa datoja Datahuoneeseen sovitun rahoituksen puitteissa
- vastaa datan pseudonymisoinnista (aineistojen tunnisteellisten tietojen korvaamisesta niin, että data on edelleen yhdisteltävissä) ja muista tietoturvan vaatimista menettelyistä
- tuotettu palvelu on Tilastokeskuksen maksullista palvelutuotantoa ja rahoitustarve kattaa toiminnan kustannukset. Myös muut Tilastokeskuksen asiakkaat voivat hankkia samanlaista palvelua.

VATT

- vastaa datapohjaisesta selvitystuotannosta
- vastaa datan käytön yhteiskehittämisestä päätöksentekijöiden kanssa

6.4. Siirtymäaikataulu nykyjärjestelystä Datahuoneeseen

GSE on sitoutunut jatkamaan Tilannehuoneen toimintaa 31.5.2021 asti ja on parastaikaa varmistamassa, että aineistoluvat mahdollistavat toiminnan jatkamisen vuoden loppuun tai 31.5.2022 asti. GSE ei kuitenkaan pysty vastaamaan toiminnan pyörittämisestä täysimittaisesti 31.5.2021 jälkeen, vaan vastuita toiminnoista on pystyttävä siirtämään muille tahoille niin, että GSE:n rooli on viimeistään 1.1.2022 alkaen korkeintaan yllä hahmoteltu kehittämis- ja laadunvarmistustehtävä.

Tämä edellyttää sitä, että

a. luodaan uusi Tilannehuonetta peilaava valitun virkamiehen vastuulla toimiva tutkimusprojekti Tilastokeskuksen tutkijapalveluihin, johon kytkeytyvät Datahuoneen rutiineista vastaavat VATT:in ja ministeriöiden työntekijät
b. Helsinki GSE:n olemassa olevaa Tilannehuonehanketta jatketaan ja ryhtyy toteuttamaan ko. hankkeen puitteissa Datahuoneessa itselleen suunniteltuja tehtäviä

c. Vastuiden siirto kaikista nykyisistä Tilannehuoneen toiminnoista käynnistyy kesällä 2021 ja toteutetaan vaihe vaiheelta, kuitenkin niin, että kaikki vastuut ovat siirtyneet vuoden loppuun mennessä.