Datahuone – taustamuistio 15.4.2021

Ratkaistava ongelma
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 hyvinkin suuria päätöksiä tehdään hyvin nopealla aikataululla.

Suomessa on poikkeuksellisen laadukas julkisten rekisteriaineistojen perusinfrastruktuuri. Suuressa osassa päätöksentekoprosesseista ei kuitenkaan ole mahdollisuutta juurikaan hyödyntää näitä aineistoja. Tämä koskee erityisesti nopeita päätöksentekotilanteita. Merkittäviä ongelmia ovat aiheuttaneet esimerkiksi niin viiveet käytössä olevien aineistojen päivityksissä kuin pitkät lupaprosessit.

Yksi datan hyödyntämisen vaikeus piilee siinä, että data-asioita tuntemattoman päätöksentekijän on vaikea tunnistaa datan hyödyntämisen mahdollisuuksia ja toisaalta datasta voidaan kuvitella olevan apua sellaisissa tilanteissa, joihin datan hyödyntäminen ei sovellu. Sama pätee myös toisin päin: data-asiantuntijan on usein vaikea syvällisesti ymmärtää päätöksentekoa ja sitä, mikä kulloinkin on päätöksenteon näkökulmasta kiinnostavaa. Tämän esityksen tarkoituksena on osaltaan tarvittavan dialogin edistäminen.

Helsinki GSE:n Tilannehuone on raapaissut ajantasaisten rekisteriaineistojen hyödyntämisen pintaa ja eri sektoreilla useilla esimerkeillä osoittanut, että nykyistä merkittävästi nopeampi, rikkaampi ja laadukkaampi datan hyödyntäminen niin hitaissa kuin nopeissakin päätöksentekotilanteissa on mahdollista.

Datahuone
Datahuone luo päätöksentekijöille uudentyyppisen mahdollisuuden hyödyntää viranomaisten toiminnasta syntyvää lähes reaaliaikaista dataa monenlaisen päätöksenteon tukena. Datahuoneessa yhdistyvät uudella tavalla 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:

- Nykyistä ketterämmin päätöksentekijöiden tarpeisiin vastaava ja nopeasti kehittyvä **joustava tietotuotanto**, jonka osana on keskeisiin aineistoihin tietosuojasäännösten puitteissa tarjottava avoin rajapinta.
- 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.
- Dataa tehokkaasti hyödyntävän päätöksenteon yksittäisten tapausten **yhteiskehittäminen** kunkin päätöksenteon kannalta relevantin päätöksentekotahon kanssa.

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.

Tilastokeskuksen tutkijapalvelun toimintaa laajennetaan siten, että se pystyy vastaamaan Datahuoneen muiden toimijoiden määrittelemiin aineistotarpeisiin. Keskeisin etenemissuunta on niin sanottujen valmisaineistojen lisääminen ja ajantasaistaminen. Valmisaineistot ovat kaikkien tutkijoiden käytettävissä ja siten toimintamalli palvelee Datahuoneen ohella koko tutkijayhteisöä. Datahuoneen perustaminen on siis osa laajempaa Tilastokeskuksen tutkijapalveluiden uudistamista ja monipuolistamista. **Tavoitteena on, että jatkossa laajat**

ajantasaiset rekisteriaineistot ovat kaikkien tutkijoiden käytössä Tilastokeskuksen tutkijapalveluiden kautta.

Tutkijapalvelun laajentaminen kiirehtii tilastolainsäädännön uudistamista. Nykylainsäädännön perusteella Tilastokeskuksella on oikeus ottaa vastaa rekisteri- ja muita tietoja tilastojen laatimiseksi. Keskeinen haaste on selkeyttää Tilastokeskuksen oikeutta ottaa vastaan ja muokata rekisteriaineistoja myös muihin tietotarpeisiin kuten esitetyn Datahuoneen käyttöön. Tilastolainsäädännön valmistelu on joka tapauksessa käynnistymässä VM:ssä kevään aikana. Datahuoneen, sekä Tutkijapalvelun laajemminkin, tarpeet on mahdollista huomioida tässä yhteydessä. Uuden toimintamallin ja muiden enemmän teknisten tarpeiden huomioiva hallituksen esitys voidaan antaa syysistuntokauden alkuun mennessä.

Hahmotelma datahuoneen organisaatiosta ja vastuista
Datahuoneessa keskeisessä roolissa olisivat:

- 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 kaikille tutkijoille tarjottavia valmiita tilastoaineistoja, jolloin ajantasaisia aineistoja saadaan Datahuoneen ohella muidenkin tutkijoiden käyttöön kattavammin ja nykyistä 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

Resursointitarve
Resursointitarve jakautuu käynnistysvaiheeseen ja jatkuvaan toimintaan. Resursointitarve jakautuu seuraavasti:

VNK/VM yhteinen datahuoneryhmä
Jatkuva kustannus: 300k€/vuosi

Helsinki GSE
Jatkuva kustannus: 500k€/vuosi

Tilastokeskus
Investointi (kulut kahden ensimmäisen vuoden aikana): 1,9M€/vuosi
Jatkuva kustannus: 950k€/vuosi

VATT
Investointi (kulut kahden ensimmäisen vuoden aikana): 600k€/vuosi
Jatkuva vuosikustannus kahden ensimmäisen vuoden jälkeen: 450k€/vuosi

Yhteensä
Kaksi ensimmäistä vuotta: 3,3M€/vuosi
Pysyvä rahoitus kahden ensimmäisen vuoden jälkeen: 2,2M€/vuosi

Liitteet
Helsinki GSE:n suunnitelma
Tilastokeskuksen suunnitelma
VATT:in suunnitelma

Lisätietoja:
Oskari Nokso-Koivisto
0440450732