Tilastokeskus

13.4.2021

Datahuoneen mahdollinen perustaminen Tilastokeskuksen näkökulmasta

Yleistä

Tilastokeskus suhtautuu myönteisesti esitykseen Datahuoneen perustamiseksi. Helsinki GSE:n perustama Talouden tilannehuone on osoittautunut toimivaksi konseptiksi koronaepidemian vaikutuksia arvioitaessa, ja on nähtävissä, että vastaavaa tietotuotantoa ja siihen perustuvaa analyysia tarvittaisiin yhteiskuntapoliittisen päätöksenteon tueksi normaalioloissakin.

Tilastokeskus on pidemmällä aikavälillä valmis ottamaan tehtäväkseen tietopohjaista päätöksentekoa tukevaa tietotuotantoa myös laajemmin kuin sille perinteisesti kuuluvan tilastotuotannon. Toiminnan laajentaminen merkittävästi nykyisestä Tilastokeskuksen tutkijapalvelusta voidaan toteuttaa tekemällä tarvittavat muutokset tilastolakiin ja lakiin tilastokeskuksesta sekä osoittamalla tehtävään rahoitusta. Lisäksi toiminnan laajentaminen edellyttää, että Tilastokeskuksessa harkitaan, missä määrin toiminta tulee eriyttää perinteisestä tilastotuotannosta ja missä määrin on mahdollista hyödyntää ilmeisiä synergioita tilasto- ja muun tietotuotannon välillä.

Tilastokeskukselle luonteva ja myös kansainväliseen sääntelyyn soveltuva rooli on tietojen kerääminen ja vastaanottaminen (nykyään lähestulkoon kokonaan perustuen olemassa oleviin rekisteriaineistoihin) sekä niiden käsittely ja analysointi suoraan käytettäviksi tilastoiksi ja muiksi tiedoiksi. Samaten Tilastokeskuksen mandaattiin sopii aineistojen käsittely, suojaaminen ja saattaminen tutkijoiden käyttöön. Keskeistä on huomioida, että Tilastokeskus ei näe mahdollisena laajentaa Tilastokeskuksen tehtäviä tutkimustoimintaan tai miltään osin politiikkasuosituksien valmisteluun.

Tilastokeskuksen arvion mukaan Datahuoneen toiminnan aloittaminen on tehtävissä nykyisen Tutkijapalvelun toimintaa laajentamalla ilman, että edellä kuvattuun mahdolliseen laajempaan mandaattiin otetaan kantaa. Toiminnan volyymin kasvu ja aineiston käyttöön liittyvien hallinnollisten viiveiden lyhentäminen puoltaisi kuitenkin vahvasti sitä, että toiminta huomioitaisiin Tilastokeskusta koskevassa laissa. Tilastokeskuksen arvion mukaan tarvittavat muutokset tilastokeskusta koskevaan lakiin olisi toteutettavissa siten, että tarvittava HE olisi annettavissa syysistuntokauden alussa.

Laajemmat muutokset Tilastokeskuksen tehtäviin ja tiedonsaantioikeuksiin vaativat kuitenkin laajempaa valmistelua. Samassa yhteydessä on syytä odottaa EU-lainsäädännön valmisteilla olevia muutoksia (esim. Data Governance Act), jotka saattavat tulla voimaan noin ensi vuodenvaihteessa. Tässä laajemmassa valmistelussa voi ottaa huomioon kokemukset nyt mahdollisesti syntyvän Datahuoneen toiminnasta. Laajempi lakiuudistus voitaisiin mahdollisesti saada eduskunnan käsittelyyn 2022.

Esitetty hallintomalli

Tilastokeskus ottaa kantaa Datahuoneen hallintomalliin ainoastaan oman toimintansa näkökulmasta. Tilastokeskuksen näkökulmasta seuraavat periaatteet on huomioitava toiminnassa:

1. Datahuoneelle tuotettava palvelu on Tilastokeskuksen maksullista palvelutuotantoa, ja Tilastokeskus saa toiminnasta kustannuksia vastaavan rahoituksen. Myös muut Tilastokeskuksen asiakkaat voivat hankkia samanlaista palvelua.
2. Palvelu toteutetaan Tilastokeskuksen tutkijapalvelu-konseptilla, joka on tarjolla myös muille tutkijoille ja tiedonkäyttäjille.
3. Tutkijapalvelun toiminnan laajetessa se pidetään edelleen erillään käsitteellisesti ja tarpeellisessa määrin myös toiminnallisesti tilastotuotannosta, jota säätelee kansainvälinen normisto ja mm. tilastolaki.
4. Tutkijapalvelun toiminnan laajentaminen ja aineistotuotannon nopeuttaminen puoltaa toiminnan huomioimista Tilastokeskusta koskevassa laissa. Tilastoaineistojen tietojen luovuttaminen tutkimus- ja muuhun käyttöön ei ole sallittua ennen tilastojen julkistamista. Lainmuutos mahdollistaisi kuitenkin rekisteriaineistojen vastaanottamisen tutkijapalveluita varten samaan aikaan kun se vastaanotetaan tilastointia varten.
5. Tilastokeskus varautuu vastaamaan aineistokysyntään laajentamalla valmisaineistojensa saatavuutta. Tämä ja sen vaatimat tietojärjestelmämuutokset ovat laaja investointi, johon Tilastokeskus voi ryhtyä vain, jos se saa varmuuden rahoituksesta laajennettavaksi esitetylle toiminnalle. Siksi linjaukset Datahuone-toiminnan rahoituksesta ovat olennainen edellytys toiminnan vaatimien investointien toteuttamiselle.

Arvioita kustannuksista

Seuraavassa on eritelty investointikustannukset ja pysyvät kustannukset, joiden avulla on mahdollista rakentaa ja ylläpitää Tilastokeskuksen toiminta nopeaan tietotuotantoon soveltuvassa Datahuoneessa. Hinnat kattavat Datahuoneen infrastruktuurin rakentamisesta ja pyörittämisestä aiheutuvat menot Tilastokeskuksen osalta.

Datahuoneessa tarvittavien aineistojen käsittelyä varten voidaan ottaa käyttöön Tilastokeskuksen Fiona-etäkäyttöjärjestelmä, joka on suojattujen tutkimusaineistojen käyttöön tarkoitettu tietoturvallinen käyttöympäristö. Fiona-etäkäyttöjärjestelmä on maksullista palvelutoimintaa, ja sen kautta on mahdollista tarjota ympäristö tutkimushankkeiden käyttöön hinnaston mukaisesti.

Datahuoneen nopea toiminta mahdollistetaan lisäämällä merkittävästi rekisteri- ja muita aineistoja niin sanotuiksi valmisaineistoiksi. Tämä vähentää tapauskohtaista räätälöintityötä

merkittävästi, ja nopeuttaa aineistojen käyttöön saamista. Tavoitteena on lisätä valmisaineistojen määrää nykyisestä noin 60 valmisaineistosta 100 valmisaineistoon. Samalla parannettaisiin palvelua kaikille tutkijoille tarjoamalla käyttöön entistä kattavampi ja ajantasaisempi valmisaineistokirjasto.

Datahuoneelle tarjottava palvelu ei ole tilastotuotantoa. Juridisen statuksen selkeyttämiseksi, tutkijapalveluiden tehtävä tulee määritellä Tilastokeskuksen tehtäväksi lakiin tilastokeskuksesta. 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 HE voidaan antaa syysistuntokauden alkuun mennessä.

Tilastokeskuksen Datahuoneelle tuottama palvelu on maksullista palvelua, mikä mahdollistaa sen, että jos joku toinen Tilastokeskuksen asiakas haluaisi tilata Tilastokeskukselta palveluita vastaavan kaltaiseen datahuoneeseen, tämä tulisi pienentämään Datahuoneen vuosimaksun suuruutta. Alla Tilastokeskuksen alustava kustannusarvio investointivaiheen kustannuksista ja pysyvistä kustannuksista.

Datahuoneen investointivaiheen kustannukset Tilastokeskuksessa

- **Tutkimusaineistojen muodostaminen tilastojulkaisuiden arkistoitavista perusaineistoista**, prosessin perustaminen, 1.000.000 €
    - **Kuvaus:** Prosessi mahdollistaa metatiedoilla rikastettujen yksikkötason tilastoaineistojen viemisen tutkimusaineistoksi heti tilastojulkistuksen jälkeen.
- **Valmisaineistojen lisääminen**, kansalliset rekisterit ja muut aineistot valmisaineistoiksi, 650.000 €
    - **Kuvaus:** Kansallisten rekistereiden vieminen valmisaineistoksi tutkijoiden käyttöön. Tämä edellyttää muutosta lakiin tilastokeskuksesta. Kattaa noin 40 uutta valmisaineistoa. Ei kata kustannuksia, joita rekisterinpitäjät mahdollisesti veloittavat rekisterin aineistojen toimittamisesta Datahuoneen käyttöön.
- **Prosessinhallintajärjestelmän käyttöönotto**, 250.000 €
    - **Kuvaus:** Tutkijapalveluiden toimeksiannon työnkulun hallinta käyttölupahakemuksen vastaanottamisesta toimitukseen.
- **Infrastruktuurin kehittäminen**, 500.000 €
    - **Kuvaus:** Tutkijoita varten luotavan oman käyttöliittymän rakentaminen Tilastokeskuksessa käytössä olevaan Metsy-metatietojärjestelmään. Rajapinnan rakentaminen valmisaineistokuvausten viemiseksi aineistokatalogiin. Valmisaineistoprosessin automatisointi kuten mm. valmisaineistojen suojaus, sisällön poiminta ja tarkistuksia sekä aineistojen siirtoja eri järjestelmien välillä.

Yhteensä vuosien 2022 ja 2023 Datahuoneen investoinnit Tilastokeskuksessa 2.400.000 €

Pysyvät kustannukset, eli investointien aiheuttama vuosittainen kustannus 950.000 €. Tämä sisältää 7 henkilöä sekä muita kustannuksia. Tehtävät ja muut kustannukset tarkemmin alla.

- Uusien valmisaineistojen ylläpito, neuvonta ja päivittäminen tarvittavalla syklillä (nopeimmat viikoittain, harvemmin päivittyvät vuosittain).

- Aineistojen ja muuttujien kuvaaminen.
- Uusien rekistereiden hankinta, toimitus ja lupien hallinnointi
- Tekninen infrastruktuuri ja datainfrastruktuuri
- Datahuoneen työntekijöiden käytön edellyttämät kapasiteetti- ja resurssimaksut
- Datahuoneen käytössä olevien ohjelmistojen lisenssimaksut

Yllä kuvatussa Datahuoneen toimintamallissa ja siihen liittyvissä kustannusarvioissa ei ole otettu huomioon Datahuoneen mahdollista vaikutusta Tilastokeskuksen muihin maksullisen toiminnan tuloihin. Viime vuosina Tilastokeskuksen maksulliset tulot ovat olleet 7-9 miljoonaa euroa. Näistä valtionhallinnon osuus on 4 miljoonaa € ja paikallishallinnon 1,1 miljoona €. Datahuoneen vaikutusta Tilastokeskuksen talouteen, maksullisen palvelutoiminnan kautta, ei ole käytettävissä olevassa aikataulussa pystytty arvioimaan.

Edellä kuvatut investoinnit valmisaineistoihin parantaisivat kaikkien tutkijoiden saamaa palvelua, ja ratkaisusta hyötyisi siten Datahuoneen palveluiden käyttäjien ohella myös tavanomainen tutkimus. Edellä kuvatut laskelmat rahoitustarpeesta eivät kuitenkaan kata Datahuoneen ulkopuolisten tutkijoiden lisenssi-, kapasiteetti- tai lupakäsittelymaksuja. On oletettavaa, että helposti käyttöön saatavien valmisaineistojen määrän lisääntyessä niiden käyttö laajenee, ja tämä tulisi huomioida tutkimusrahoituksessa. Tilastokeskus on valmis osallistumaan myös näiden kustannusten arviointiin.

Datahuoneen perustamista koskevissa keskusteluissa on viitattu mahdollisuuteen tulevaisuudessa laajentaa aineistojen saatavuutta avoimena datana ja samalla ilman käyttäjiltä perittäviä maksuja. Tilastokeskus korostaa, että edellä esitetyissä kustannusarvioissa ei ole otettu tätä mahdollista etenemissuuntaa huomioon, mutta on valmis aloittamaan tätä koskevat selvitykset mm. tarvittavan budjettirahoituksen määrän arvioimiseksi.