Eduskunnan Talousvaliokunnalle

HE 72/2022 vp eduskunnalle laeiksi tilastokeskuksesta annetun lain ja tilastolain muuttamisesta

1. Tausta

Tilastokeskuksen näkökulmasta Tilastokeskusta koskevan lain ja tilastolain uudistamisen tavoitteena on Tilastokeskuksen tietosuojaoikeudellisen aseman selkeyttäminen tutkija- ja aineistopalvelun tarjoajana sekä Tilastokeskuksen tiedonsaantioikeuksien täsmentäminen.

Muuttuva lainsäädäntö luo edellytykset Tilastokeskuksen toiminnan kehittämiselle. Käytännössä muutokset ovat mahdollisia vain, jos samalla turvataan Tilastokeskuksen perustehtävä, tilastotuotanto, jonka avulla syntyy tarvittavat tietovarannot lainsäädännön mahdollistamille uusille palveluille. Tilastotuotannon tietojärjestelmiä ja prosesseja kehitetään jatkuvasti. Tästä huolimatta kehittämisvelka kasvaa, kun järjestelmien elinkaaret päättyvät nopeammin, kuin niitä pystytään korvaamaan.

2. Datahuone - Lainsäädäntö mahdollistaa datahuoneen aineistoinfrastruktuurin toteuttamisen

Kesällä 2022 Valtion taloudellisen tutkimuskeskuksen (VATT) yhteyteen perustettiin datahuone-pilotti, jossa Tilastokeskus toimii yhtenä organisaationa. Datahuoneen tavoitteena on tuottaa valtioneuvostolle päätöksenteon tueksi ajantasaista, rekisteritietoja hyödyntävää ja aikaisempaa tarkempaa tutkimustietoa yhteiskunnallisten ja poliittisten ratkaisujen vaikutuksista.

Pilotin tarkoituksena on siirtää korona-aikana toimineen Tilannehuoneen oppeja kiinteäksi osaksi päätöksenteon tukea ja edistää data-analyysin hyödyntämistä päätöksenteon eri vaiheissa. Pilotin tavoitteena on myös tuoda data-analyysityötä lähemmäs ministeriöiden arkea ja tukea analyysiosaamista ja sen leviämistä ministeriöissä poikkihallinnollisesti.

Tilastokeskus perusti uuden datahuonepilotin työtilat ja toimitti ensimmäiset aineistot Datahuoneessa työskentelevien tutkijoiden käyttöön heinäkuussa 2022. Virallisesti datahuonepilotti käynnistyi syyskuussa 2022. Datahuoneen data-analyysitoimintaa tehdään VATT:n ja HGSE:n yhteistyössä toteuttamilla projekteilla. Projektien tutkimusinfrastruktuurista vastaa Tilastokeskus. Tietoja käsitellään Tilastokeskuksen tietoturvallisessa etäkäyttöympäristössä. Datahuonepilotin ohjausryhmä koordinoi tutkimushankkeita.

Datahuonepilotti päättyy vuonna 2025. Pilotin loppuvaiheessa tarkastellaan kokeilun tuloksia ja tehdään päätökset jatkotoimenpiteistä.

3. Valmisaineistot mahdollistavat laadukkaan ja nopean tutkimuksen

Yksi keskeisiä toimia datahuonetoimintojen tai nopean tutkijapalvelun mahdollistamiseksi on ns. valmisaineistojen koostaminen. Valmisaineistoilla tarkoitetaan jatkuvasti ylläpidettäviä, ajantasaisia ja rajattuja datatiedostoja, joissa suoran tunnistamisen mahdollistavat henkilötiedot on poistettu ja korvattu pseudotunnisteella. Näin yksilön tai yrityksen suora tunnistaminen on estetty. Valmisaineistoihin viitataan myös hallituksen esityksessä HE 72/2022:ssä.

Valmisaineistot muodostavat datakirjaston, josta tutkija voi valita tutkimuksensa kannalta välttämättömät tiedot. Aineistot voidaan luovuttaa tutkijalle nopeasti, koska aineistoja ylläpidetään jatkuvasti ja ne on suojattu valmiiksi siten, että

yksilöiden suora tunnistaminen ei ole mahdollista. Toisin sanoen valmisaineistot ovat ajantasaisia, viimeisimmät lähdetiedot sisältäviä suojattuja data-aineistoja.

Yksi valmisaineistokokonaisuus sisältää ilmiön rajatun osan kannalta tarpeellisen tiedon. Tulkinnan tarpeellisesta kokonaisuudesta tekee aineiston tunteva Tilastokeskuksen asiantuntija. Tutkija voi oman tutkimussuunnitelmansa niin vaatiessa hakea lupaa useampaan valmisaineistoon ja yhdistää ne toisiinsa. Näin hän saa itse muodostettua tarvitsemansa isomman kokonaisuuden.

Tarvittaessa valmisaineistoista poistetaan tietoja, jotka eivät ole välttämättömiä. Näin noudatetaan tietosuoja-asetuksen minimointiperiaatetta. Tutkijalle annetaan vain tutkimuksen kannalta välttämättömät tiedot. Kuitenkin tutkijalle näytetään mitä tietoa ylipäänsä on saatavilla, jotta tutkijalla on mahdollisuus hyödyntää se potentiaali mitä on saatavilla.

Valmisaineistojen avulla Tilastokeskus on voinut kuvata tutkijoille tietovarallisuuden potentiaalia, sitä mitä tietoa on saatavilla. Aikaisempaa laajempi valmisaineistovalikoima parantaisi aineistojen hyödyntämistä tutkimuskäytössä ja samalla se nopeuttaisi tietojen saamista tutkimuskäyttöön.

4. Tutkija- ja aineistopalvelu tulisi nähdä ja rahoittaa osana tutkimusinfrastruktuuria

Tilastokeskukseen tilastotuotannossa syntyvät tietovarannot mahdollistavat eri puolella yhteiskuntaa syntyvien tietojen yhdistelyn. Nyt käsittelyssä oleva hallituksen esitys mahdollistaisi tämän tietovarannon täydentämisen, sillä sen tultua voimaan Tilastokeskuksella olisi oikeus ottaa vastaan ja säilyttää muiden rekisterinpitäjien aineistoja välitettäväksi tutkimuskäyttöön.

Näin syntyvä nykyistä laajempi tietovaranto tuo myös uudenlaisen mahdollisuuden tuottaa valmisaineistoja tutkimuksen käyttöön. Helposti yhdisteltävien ja tutkijoita varten kuvattujen aineistojen tuottaminen vaatii kuitenkin erillistä rahoitusta. Tämä mahdollistaisi sellaisen tiedontuotannon, jota ei aikaisemmin ole tehty.

Nyt käsiteltävä hallituksen esitys (HE 72/2022) ja laki sosiaali- ja terveystietojen toissijaisesta käytöstä (552/2019) tuovat molemmat esille tarpeen käsitellä tutkimuskäyttöön annettuja mikroaineistoja tietoturvallisessa käyttöympäristössä. Tietoturvallisen käyttöympäristön ylläpidosta ja kehittämisestä syntyvät käyttökustannukset olisi huomioitava myös tutkimusrahoituksessa tai palvelun tarjoajien rahoituksessa.

Suomen tietovarallisuuden potentiaalin täysimääräinen hyödyntäminen edellyttäisi, että tutkija- ja aineistopalvelu nähtäisiin ja rahoitettaisiin osana tutkimusinfrastruktuuria. VM:n asettaman ns. Spolanderin työryhmän (Julkishallinnon aineistojen hyödyntämistä selvittävän työryhmä)[1] loppuraportissa vuodelta 2020 suositellaan palvelun hinnoitteluun muutoksia:

> *”Tavoite 9: Maksulliset aineistot ja niiden käyttö on hinnoiteltu käyttäjille siten, että hinta kattaa niiden tarjoamisen rajakustannukset. Tavoitteen saavuttaminen edellyttää rahoitusresurssien uudelleenallokaatiota toimijoiden välillä ja hinnoittelukäytäntöjen muuttamista sekä lainsäädännön uudistamista.”*
>
> *”Tavoite 10: Tutkimusaineistojen tarjonnan infrastruktuuri kuten aineiston tuottajien rekisteriaineistojen kuvaus, aineiston käsittelijöiden valmisaineistojen suunnittelu ja ylläpito, aineistoja koskeva neuvonta ja ohjeistus sekä etäkäyttöjärjestelmien käyttäjäriippumaton ICT-infrastruktuuri rahoitetaan kohdennetulla budjettirahoituksella eikä*

[1] Julkishallinnon aineistojen hyödyntämistä selvittävän työryhmän loppuraportti (valtioneuvosto.fi)

tällaisia infrastruktuurikustannuksia kohdisteta käyttäjille. Tavoitteen saavuttaminen edellyttää rahoitusresurssien uudelleenallokaatiota toimijoiden välillä ja hinnoittelukäytäntöjen muuttamista sekä lainsäädännön uudistamista."

Yhteiskunnan kannalta olisi hyödyllistä, että kansallista tietovarallisuutta hyödynnettäisiin nykyistä enemmän ja ennen kaikkea enemmän kuin mihin tutkimushankkeiden rahoitus markkinaehtoisesti riittäisi. Esitetty tavoite on kannatettava, mutta käytännön kysymykset vaatisivat tarkempaa määrittelyä. Mikäli Spolanderin työryhmän tavoitteet 9 ja 10 eivät johtaisi riittävän alhaisiin hintoihin, voidaan niitä alentaa osoittamalla palvelulle budjetissa hintatuki.

Tätä kirjoitettaessa auditoituja etäkäyttöympäristöjä on jo yhteensä kahdeksan, joista yksi on Tilastokeskuksen etäkäyttöympäristö (Fiona). On odotettavissa, että tulevaisuudessa joudumme pohtimaan millä edellytyksillä Tilastokeskus voi luovuttaa aineistoja muihin auditoituihin etäkäyttöympäristöihin. Mikäli suurten tutkimuslaitosten käyttö siirtyy omiin järjestelmiin, voi tämä vaikuttaa Tilastokeskuksen mahdollisuuteen tarjota palvelua ylipäätään, kun rahoituspohja pienenee. Tällöin kaikilla tutkijoilla ei ole mahdollisuutta hyödyntää Tilastokeskuksen aineistoja, jos ei ole pääsyä etäkäyttöjärjestelmään.

Jotta edellä kuvattu kehityskulku vältetään, tulisi Tilastokeskuksen tutkija- ja aineistopalvelun hinnoittelu määritellä tutkijoille riittävän edulliseksi. Näkemyksemme mukaan, olisi myös kansallisesti järkevää kehittää muutama hyvä kansallinen järjestelmä sen sijaan, että eri toimijat kehittävät kukin omansa.

5. Valtiovarainministeriön asetus Tilastokeskuksen suoritteiden maksullisuudesta päivitettävänä

Tilastokeskuksen suoritteiden maksullisuudesta säädetään valtiovarainministeriön asetuksella. Asetuksen 2 §:ssä määritellään tutkimusaineistojen luovuttaminen liiketaloudellisin perustein hinnoiteltavaksi maksulliseksi suoritteeksi.

Valtiovarainministeriön asetus Tilastokeskuksen suoritteiden maksullisuudesta on parhaillaan päivitettävänä ministeriössä. Vuoden 2023 talousarvioesityksessä on esitetty Tilastokeskuksen tutkijapalveluille 900 000 € hintatukea. Hintatuesta huolimatta tutkija- ja aineistopalvelu tulee olemaan alijäämäinen palvelu, vuonna 2021 alijäämä oli 2,9 miljoonaa euroa.

6. Euroopan tilastoja koskevan asetuksen (EY) N:o 223/2009 uudistamista suunnitellaan

Euroopan tilastoja koskevan asetuksen (EY) N:o 223/2009 uudistamista suunnitellaan. Komissio on käynnistänyt julkisen kuulemisen, joka päättyy 25.10.2022. Asetuksen uudistamisen tavoitteena on mahdollistaa tilastojen nopea tuottaminen käyttäjien kysynnän mukaisesti, lisätä tehokkuutta ja vähentää tietojen toimittajien, myös jäsenvaltioiden, kustannuksia. Tilastokeskuksen näkemyksen mukaan nämä tavoitteet ovat kaikilta osin kannatettavia.

7. Lainsäädännön kehittämistä tulee jatkaa

Tilastokeskusta koskevan lainsäädännön kehittämistyössä on tullut eteen myös sellaisia uusia muutostarpeita, joita ei ehditty valmistelemaan nyt käsillä olevan hallituksen esityksen muotoon. Muutostarpeet koskevat sekä Tilastokeskuksen tehtäviä (esim. Tilastokeskuksen oikeus tarjota "operaattoripalvelua" eli oikeus toimia tietoaineistojen välittäjänä) ja Tilastokeskuksen tiedonsaantioikeuksia.

Tiedon määrä, sen hyödyntämisen mahdollisuudet ja teknologiat kehittyvät tällä hetkellä nopeasti. Tämä edellyttää myös lainsäädännön kehittämistä, jotta pysymme kehityksen vauhdissa mukana.