

Ajankohtaista tutkijapalveluista

Tilastokeskus



Ajankohtaista tutkijapalveluista – News from Research Services Lokakuu/ October 2024

For English see below

Tilastokeskus valmistelee tutkimusaineistojen tietosuojaan liittyviä muutoksia, joiden tavoitteena on vahvistaa kansalaisten ja organisaatioiden tietojen suojaa Tilastokeskuksen aineistoilla tehtävissä tutkimuksissa. Muutokset vaikuttavat käyttöluvien myöntämisen ja ne tulevat käyttöön vuonna 2025. Haluamme suunnitella muutokset yhdessä tutkijayhteisön kanssa.

Etäkäyttöympäristön ja aineistojen hinnat nousevat vuonna 2025. Syynä hintojen korotukseen on se, että tutkijapalvelun tulot eivät kata palvelun tuottamisen kustannuksia. Aineistojen ja FIONA-etäkäyttöjärjestelmän hinnat julkaistaan marraskuun loppuun mennessä ja käyttöluvien hinnat vahvistetaan vuodenvaihteeseen mennessä.

Keväällä tekemämme [asiakaskyselyn](#) perusteella tyytyväisyys palveluun on parantunut. Tutkijapalvelun hakemuskäsittely on pysynyt kohtuullisen sujuvana. Laajojen aineistohakemusten käsittely on kuitenkin hieman hidastunut aiempaa tarkemman tietosuojan arvioinnin vuoksi. Laajat hakemukset pääsevät käsittelyyn tällä hetkellä noin kolmen kuukauden päästä hakemuksen saapumisesta. Normaalien hakemusten odotusaika käsittelyyn ottamiseen on tällä hetkellä noin neljä viikkoa.

Tilastokeskuksen tutkimusaineistojen tietosuoja vahvistuu

Tilastokeskus suunnittelee muutoksia tutkimusaineistojen luovuttamisen käytäntöihin. Muutosten tavoitteena on parantaa tietosuojaa.

Muutokset on tarkoitus ottaa käyttöön syksyllä 2025. Suurin muutos koskee näillä näkymin FIONA-hankkeille luovutettavien tutkimusaineistojen rajaamista tutkimusten kannalta välttämättömiin tietoihin. Käyttäjä- ja aineistomääriltään suurien FIONA-hankkeiden kohdalla mietitään myös ratkaisuja hankkeiden tietosuojan parantamiseksi.

Järjestimme muutokseen liittyen keskustelutilaisuuden tutkijoille 17.9.2024. Tilaisuuden yhteenvedossa ja esityksessä on tarkempia tietoja suunniteltavista muutoksista.

[Tutkimusaineistojen tietosuoja vahvistuu yhteenveto \(pdf\)](#)

[Tutkijatilaisuus esitys \(pdf\)](#)

Tutkijapalveluiden pseudonymisointi uudistuu

Tilastokeskus on uudistamassa tilastoaineistojen pseudonymisointia. Tässä yhteydessä myös tutkijapalveluiden pseudonymisointiratkaisu muuttuu. Uusi pseudonymisointimenetelmä sopii paremmin erilaisten tunnisteiden suojaamiseen ja parantaa tutkimusaineistojen teknistä tietoturvaa.

Pseudonymisoinnin uudistaminen näkyy FIONA-käyttäjille uudenlaisia pseudotunnisteina. FIONA-etäkäyttöhankkeet saavat väliaikaisesti käyttöönsä tunnisteiden välisen avaimen uusien pseudotunnisteiden käyttöönottoa varten. Avaimet varmistavat sujuvan siirtymän uusien tunnisteiden käyttöön.

Ratkaisu on parhaillaan pilotointivaiheessa ja uudet pseudotunnisteet on tarkoitus ottaa käyttöön valmisaineistoissa ja räätälöidyissä aineistoissa vuoden 2025 alusta alkaen. Lisätietoa muutoksesta ja siirtymäajasta julkaistaan myöhemmin syksyllä FIONA News - uutiskirjeessä.

Tutkijapalveluiden hinnat ja laskutus 2025

Valmistelemme tutkijapalveluiden kautta saatavilla olevien aineistojen ja FIONA-etäkäyttäjärjestelmän hintojen ja käyttäjämaksun korotuksia. Korotusten suuruudet eivät ole vielä tiedossa. Myös muita maksuja voidaan tarkistaa.

Hinnankorotukset johtuvat siitä, etteivät tutkijapalveluiden tulot kata palvelun tuottamisesta syntyneitä menoja. Aineistojen ja FIONA-etäkäyttäjärjestelmän hinnat julkaistaan marraskuun loppuun mennessä. Käyttölupien hinnat määräytyvät asetuksessa Tilastokeskuksen suoritteiden maksullisuudesta, joka annetaan vuodenvaihteeseen mennessä valtiovarainministeriöstä.

Viestimme hintojen korotuksista tarkemmin heti kun mahdollista, viimeistään marraskuun lopussa.

Valmisaineistovalikoima laajenee

Valmisaineistovalikoimamme laajenee uusin moduulein. Väestön ennakkotiedot ja korkeakoulujen hakurekisteritiedot valmistuvat loppuvuoden 2024 aikana. Korkeakoulujen opiskelijoiden historiatietoja sisältävä valmisaineisto valmistuu keväällä 2025.

Väestön ennakkotiedot ([FOLK VAEN](#)) sisältää ennakkotietoja henkilöistä, joilla on kotipaikka Suomessa. Uuden aineiston avulla tutkijat saavat mahdollisimman ajankohtaisia tietoja väestöstä ja sen rakenteesta. Aineisto sisältää muun muassa demografiatietoja (ikä, sukupuoli, siviilisäätö, karkeistettuja kansalaisuustietoja), tietoja lapsiluvusta ja olinpaikasta sekä suojatut rakennus- ja huoneistotunnukset. Ennakkotiedot ovat alustavia tietoja, ja edelleen ainoastaan FOLK-moduulit sisältävät lopullista, korjattua ja tarkastettua väestötietoa.

Korkeakoulujen hakurekisteritiedot (EDUC_HAREK) sisältää kalenterivuositain korkeakoulutukseen hakeneiden tietoja. Tietoja on yliopistohauista vuodesta 1992 ja ammattikorkeakouluhauista vuodesta 1996 alkaen. Valmisaineisto perustuu Opetushallitukselta saatuihin tietoihin ja täydentää mahdollisuuksia tutkia koulutukseen hakeutumista.

Korkeakoulujen opiskelijat historiatiedot (EDUC_OPISKHIST) täydentää koulutustietoihin keskittyvää EDUC-valmisaineistoperhettä Tilastokeskuksen arkistoista työstetyillä

korkeakoulujen historiatiedoilla. Uusi valmisaineisto kokoaa yhteen korkeakoulujen vanhat henkilötason opiskelijarekisterit vuosilta 1968–1972 sekä 1975–1995. Moduuli sisältää yksikkötasoista tietoa aikanaan kirjoilla olleista opiskelijoista, koulutuksista ja tutkinnoista.

Tulorekisterin palkkatiedoista on loppuvuonna valmistumassa uusi helppokäyttöisempi *Tulorekisterin summatut palkkatiedot* -valmisaineisto (TAX_SUMINCOMES).

Laajennamme lisäksi valmisaineistokirjastoamme kevään 2025 aikana KOSKI-koulutustietovarantoon kootuilla, eri koulutusasteiden opintosuoritus-, opiskeluoikeus- ja arvosanatiedoilla. Julkaisemme myös positiivisen luottotietorekisterin kuluttajaluottotietoja sekä Maahanmuuttoviraston oleskelulupatietoja sisältäviä valmisaineistoja. Tiedotamme aineistojen tarkemmista sisällöistä ja julkaisuajankohdista myöhemmin.

Valmisaineistokirjastomme laajentuu seuraavien kuukausien aikana myös palveluaineistoilla eli keskeisillä väestöä ja elinoloa kuvaavilla haastatteluaineistoilla. Tällaisia aineistoja ovat esimerkiksi työolo-, työvoima- ja kulutustutkimusaineistot.

Tilastokeskuksen henkilötiedonkeruista tehtyjen palveluaineistojen, kuten tulonjakotilaston palveluaineiston hinnoittelu muuttuu niin, että aineiston hinta vastaa muiden valmisaineistojen hinnoittelua. Aineistot ovat käytettävissä tietoturvallisen FIONA-etäkäyttöjärjestelmän kautta, mikä helpottaa yhdistelyä Tilastokeskuksen muihin rekisteritietoihin. Käyttölupaa palveluaineistoihin voi hakea muiden aineistojen tapaan sähköisessä [lupapalvelussamme](#).

Datahuoneen käyttöön uusia aineistoja

Viime aikoina olemme koonneet eri rekistereistä tilastotarkoituksiin kerättäviä tietoja rakentamisesta ja asumisesta sekä henkilöverotuksesta Datahuoneen tutkimusryhmien käyttöön. Tutkimuskäyttöön on nyt saatavissa mm. rakennus- ja huoneistorekisterin vuositietoja, kuukausittain päivittyviä tietoja uudis- ja korjausrakentamisen luvista, kiinteistöverotietokanta ja energiatodistusaineisto. Verottajan lähteistä on saatavissa pseudonymisoituja tietoja asuntolainoista, kotitalousvähennyksistä ja perintö- ja lahjaveroista. Myös muut tutkimushankkeet voivat hakea tietoja käyttöönsä. Lisäksi Datahuoneelle on tänä vuonna toimitettu tietoja maksuhäiriömerkinnöistä Oikeusrekisterikeskuksen luvalla.

Datahuone on tietopohjaista päätöksentekoa edistävä VATT:n yksikkö, joka toimii kiinteässä yhteistyössä Tilastokeskuksen ja Helsinki GSE:n kanssa. Datahuone hyödyntää Tilastokeskuksen ja muiden lupaviranomaisten aineistoja valtioneuvoston ajankohtaisiin tietotarpeisiin. Suurin osa Datahuoneen analyyseistä ja selvityksistä toimitetaan suoraan ministeriöille. Datahuone on julkaissut alkusyksystä tutkimuksia mm. asumistukiuudistuksesta ja työperäisesti Suomessa oleskelevien työttömyydestä. [Lue lisää Datahuoneen verkkopalvelusta.](#)

Datahuonetoiminnon toimintamallit uudistuvat

Uudistamme Tilastokeskuksen datahuonetoiminnon toimintamalleja siten, että ne ovat tehokkaampia ja vastaavat tietosuojalainsäädännön vaatimuksiin.

Erityisenä kehityskohteenä on kaksiportainen hankemalli lupamenettelyineen ja teknisine ratkaisuihineen. Uudella mallilla pystytään rajaamaan tutkimushankkeen aineisto-oikeuksia ja

kapasiteettia etäkäytössä nykyistä joustavammin ja kohdistamaan käyttöä selkeämmin tarpeen mukaan. Projektissa automatisoidaan samalla tutkimusaineistojen päivittämistä.

Projekti pilotoi myös valmisaineistojen muodostamista Datahuoneelle (VATT) hankituista aineistoista tietovarantopohjaisesti. Pidemmän tähtäimen tavoitteena on, että tulevaisuudessa Datahuonekin voisi yhä useammin ottaa käyttöönsä valmisaineiston ja Datahuoneelle räätälöityjä aineistoja saataisiin laajempaan käyttöön. Projektiin valittuja pilottiaineistoja ovat väestön ennakkotiedot, Opetushallituksen Koski-koulutustiedot ja Maahanmuuttoviraston oleskeluluvat.

Julkisten aineistojen vieminen FIONA-hankkeelle

FIONA-etäkäyttöhankkeelle on mahdollista viedä myös Tilastokeskuksen tai muun tahon tuottamia julkisia aineistoja. Julkisetkin tiedot voivat muuttaa tutkimushankkeen tietosisältöä siten, että muutos voi edellyttää tietosuojan kokonaisarviointia. Koska Tilastokeskuksen tulee varmistaa henkilötietojen sekä liike- ja ammattisalaisuuden suojan toteutuminen, tulee käyttöluvapäätöstä tehtäessä olla tieto etäkäyttöhankkeella olevasta aineistokokonaisuudesta. Tämän vuoksi myös julkiset aineistot liitetään pääsääntöisesti tutkimushankkeen käyttöluvan aineistoluetteloon.

Käytännössä siis suoria tunnisteita tai paikkatietoa sisältävien sekä esimerkiksi ruutu- tai postinumerotasoisten ulkopuolisten julkisten aineistojen lisääminen FIONA-hankkeelle saattaa edellyttää käyttöluvan päivittämistä, kun taas esimerkiksi luokitukset ja tietyt kuntakohtaiset tiedot voi lähtökohtaisesti viedä hankkeelle ilman muutoksia käyttöluvaan. Myös Tilastokeskuksen ja muiden tilastoviranomaisten julkaisemia tilastotietoja etäkäyttöhankkeelle voi viedä ilman käyttöluvan päivittämistä.

Lupapalveluun on lisätty tuote ”Ulkopuolisen julkisen aineiston lisäys FIONAan”, jotta myös julkisista lähteistä kerätyn aineiston lisääminen käyttöluvalle olisi mahdollisimman suoraviivaista.

FIONAn uutisia

Statan pakettijakelu muuttuu siten, että hankkeet voivat itse ottaa käyttöön paketit Fionassa. Jakelun piiriin kuuluvat Boston College Statistical Software Components (SSC) -arkistosta asennettavat paketit. Toiminnallisuus otetaan käyttöön ja ohjeistetaan Fionan laajuisesti joulukuun huoltokatkossa 17.12.2024.

Tarjottaviin ohjelmistoversioihin tulee tarkennuksia vuoden 2025 aikana. Tarkoitus on rajata nykyisten sovellusten tuettujen versioiden määrää, jotta ylläpitotyön määrä pysyisi kohtuullisena. Rajauksen jälkeen kullakin hankkeella on vain yksi tuore versio sovelluksista käytössä. Muutoksesta tiedotetaan ja ohjeistetaan tarkemmin myöhemmin.

Aiemmat tarjotut CentOS-Linux-koneet on nyt korvattu Ubuntu-käyttöjärjestelmällä. Samassa yhteydessä aiempia ongelmia on korjattu, kuten päälle jätettyyn sessioon palaaminen.

Laajennamme Fionasta ulos siirrettävien tulosten tarkastuksessa käytetyn pistotarkastusmallin käyttöönottoa loppuvuoden aikana, kun olemme saaneet ulkoisten aineistotoimittajien hyväksynnät mallin käytölle.

Syksyn vinkki

Tutkijapalveluiden tuoreimmat uutiset päivitetään [tutkijapalveluiden verkkosivujen Ajankohtaista-osioon](#). FIONA-uutisia päivitetään myös FIONAn etusivulla avautuvaan FIONA News -uutiskirjeeseen.

News from Research Services October 2024

Statistics Finland is preparing changes related to the data protection of research data, which aims to strengthen the protection of data of citizens and organisations in studies made with Statistics Finland's data. The changes will affect the granting of user licences and they will be introduced in 2025. We want to plan the changes together with the researcher community.

The prices of the remote access environment and data will rise in 2025. The reason for the price increases is that the income from the Research Services does not cover the costs of producing the service. The prices of data and the FIONA remote access system will be published by the end of November and the prices of user licences will be confirmed by the end of the year.

The [customer survey](#) (in Finnish only) we made in the spring shows that satisfaction with the service has grown. The processing of applications by the Research Services has remained relatively smooth. However, the processing of extensive data applications has slowed down slightly due to more detailed assessment of data protection than before. Extensive applications are currently processed in about three months from the reception of the application. The time taken to process normal applications is at the moment around four weeks.

Data protection of Statistics Finland's research data is strengthened

Statistics Finland is planning changes to the practices of releasing research data. The aim of the changes is to improve data protection.

The changes are meant to be introduced in autumn 2025. The biggest change apparently concerns limiting research data released to FIONA projects to data necessary for the studies. Solutions for improving project data protection are also considered for FIONA projects with large user and data volumes.

In relation to the change, we arranged a discussion event for researchers on 17 September 2024. The summary and presentation of the event contain more detailed information on the planned changes.

[Data protection of research data strengthened: summary \(PDF\)](#) (in Finnish only)

[Researcher event presentation \(PDF\)](#) (in Finnish only)

Pseudonymisation is renewed in the Research Services

Statistics Finland is in the process of renewing pseudonymisation of statistical data. In this context, the pseudonymisation solution of the Research Services also changes. The new pseudonymisation method is better suited for the protection of different identifiers and improves the technical data security of research data.

The renewal of pseudonymisation is visible to FIONA users as new types of pseudo identifiers. FIONA remote access projects will have temporary access to the key between identifiers for the adoption of new pseudo identifiers. The keys ensure a smooth transition to the use of new identifiers.

The solution is currently in the piloting phase and the new pseudo identifiers are meant to be adopted in ready-made and tailored data from the beginning of 2025. Further information about the renewal and transition period will be published later in the autumn in the FIONA News newsletter.

Prices and invoicing of the Research Services in 2025

We are preparing increases in prices and user fees for data available through the Research Services and for the FIONA remote access system. The sizes of the increases are not yet known. Other charges may also be reviewed.

The price increases are due to the fact that the income from the Research Services does not cover expenses generated from service production. The prices of the data and the FIONA remote access system will be released by the end of November. The prices of user licences are determined in the decree concerning the charge criteria of the performances of Statistics Finland, which is issued by the Ministry of Finance by the end of the year.

We will provide more detailed information about the price increases as soon as possible, at the latest by the end of November.

Ready-made data selection will expand

Our selection of ready-made data will expand with new modules. Preliminary population data and application register data for higher education institutions will be completed during late 2024. Ready-made data containing history data on higher education students will be completed in spring 2025.

The *preliminary population data* ([FOLK VAEN](#)) contain preliminary data on persons with domicile in Finland. With the help of the new data researchers obtain as up-to-date data as possible on the population and its structure. The data include demographic data (age, gender, marital status, less detailed nationality data), data on the number of children and location and protected building and dwelling identifiers. Preliminary data are initial data and still only the FOLK modules contain final, corrected and checked population data.

The *application register data for higher education institutions* (EDUC_HAREK) contain data on applicants to higher education by calendar year. Data are available on applications to universities from 1992 onwards and to universities of applied sciences from 1996 onwards. The ready-made data are based on data obtained from the Finnish National Agency for Education and complements the possibility to examine entrance to education.

The *history data on higher education students* (EDUC_OPISKHIST) supplement the EDUC ready-made data family focusing on education data with the historical data on higher education compiled from Statistics Finland's archives. The new ready-made data combines the old individual-level student registers of universities from 1968 to 1972 and 1975 to 1995. The module contains unit-level data on enrolled students, education and degrees at the time.

Towards the end of the year, new, more user-friendly ready-made data on the *Incomes Register's summed wages and salaries data* (TAX_SUMINCOMES) will be completed based on the Incomes Register's wage and salary data.

We will also extend our ready-made data library during spring 2025 with data on completion of studies, study rights and grades at different levels of education compiled into the KOSKI education database. We will also publish consumer credit data from the positive credit register and ready-made data containing the Finnish Immigration Service's residence permit data. More detailed contents of the data and publication times will be provided later.

During the next months, our ready-made data library is also expanded with service data files, that is, key interview data describing the population and living conditions. These include data on working conditions, labour force and the Household Budget Survey.

The pricing of service data files made from Statistics Finland's personal data collections, such as the service data of the income distribution statistics, changes so that the price of the data corresponds to that of other ready-made data. The data are available through the data secure FIONA remote access system, which makes combining them with Statistics Finland's other register data easier. Like other data, a licence to use service data can be applied for [in our electronic licensing service](#).

New data available to the Data Room

Recently we have compiled data on construction and housing and personal taxation collected from different registers for statistical purposes for the use of the Data Room research groups. Annual data from the building and dwelling register, monthly updated data on permits for newbuilding and renovation building, the real estate tax database and energy certificate data are now available for research use. Pseudonymised data on housing loans, tax credits for household expenses and inheritance and gift taxes are available from the Tax Administration's sources. Other research projects can also apply for data for their use. In addition, data on payment default entries have been submitted to the Data Room this year with the permission of the Legal Register Centre.

The Data Room is a VATT unit that promotes knowledge-based decision making and works in close cooperation with Statistics Finland and Helsinki GSE. The Data Room makes use of the data of Statistics Finland and other licensing authorities for the Government's current information needs. Most of the Data Room analyses and reports are supplied directly to the ministries. In early autumn, the Data Room has published studies on the housing allowance reform and the unemployment of persons staying in Finland for work-related reasons. [Read more in the Data Room online service](#). (in Finnish only)

Operating models of the Data Room function are being renewed

We are renewing the operating models of Statistics Finland's Data Room function so that they are more efficient and comply with the requirements of data protection legislation.

A special development target is a two-tier project model with its licence procedures and technical solutions. The new model will allow us to limit the research project's data rights and capacity in remote access use more flexibly than at present and to attribute use more clearly according to need. The project will also automate the updating of research data.

The project also piloted the formation of ready-made data from materials obtained for the Data Room (VATT) on the basis of data resources. The long-term objective is that in future, the Data Room could increasingly use ready-made data and data tailored for the Data Room could be available for wider use as well. Pilot data selected for the project include

preliminary population data, the Finnish National Agency for Education's Koski education data and the Finnish Immigration Service's residence permits.

Exporting public data to a FIONA project

Public data produced by Statistics Finland or some other party can also be exported to a FIONA remote access project. Even public data may change the data content of the research project so that the change may require an overall assessment of data protection. Because Statistics Finland must ensure the protection of personal data and business and professional secrets, information on the data set included in the remote access project must be known when making the user licence decision. For this reason, public data are also mainly attached to the list of data files in the research project's user licence.

In practice, adding public data containing direct identifiers or geographic information and, for example, external public data on grid or postal code level to a FIONA project may therefore require updating of the user licence, while classifications and certain municipality-specific data can, as a rule, be exported to the project without changes to the user licence. Statistical data published by Statistics Finland and other statistical authorities can also be exported to the remote access project without updating the user licence.

The product "Addition of external public data to FIONA" has been added to the licensing service so that adding data collected from public sources to the user licence would be as straightforward as possible.

FIONA news

Stata's package distribution will change in such a way that the projects are able to take into use the packages in FIONA. The distribution includes packages installed from the Boston College Statistical Software Components (SSC) archive. The functionality will be introduced and guidelines will be issued for the whole of FIONA during the December maintenance break on 17 December 2024.

Specifications will be made to the software versions available during 2025. The aim is to limit the number of supported versions of the existing applications so that the volume of maintenance work remains reasonable. After the limitation, each project has only one recent version of the applications in use. Information and instructions on the change will be issued later.

The previously provided CentOS-Linux machines have now been replaced by the Ubuntu operating system. Simultaneously, earlier problems have been corrected, such as returning to a session left on.

We will extend the introduction of the random checking model used in the checking of outputs to be exported from FIONA towards the end of the year once we have received the approval of external data suppliers for the model's use.

Tip of the autumn

The latest news of the Research Services are updated in the [Topical section of the Research Services' website](#). FIONA news are also updated in the FIONA News newsletter opening on the FIONA front page.

