

Ajankohtaista tutkijapalveluista

Tilastokeskus 



Tiedote FIONAn käyttäjille pseudonymisoinnin muutoksista tutkijapalveluissa 1.4.2025

Announcement to FIONA users regarding changes in pseudonymisation in Research Services on April 1, 2025

For English see below

Tilastokeskuksen tutkijapalveluiden pseudonymisointiratkaisu uudistuu. Uusi pseudonymisointimenetelmä sopii paremmin erilaisten tunnisteiden suojaamiseen ja parantaa tutkimusaineistojen teknistä tietoturvaa. Henkilöaineistot yksilöivä shnro pysyy toistaiseksi ennallaan, sen tulevista muutoksista tiedotetaan erikseen.

Uudistus toteutetaan 1.4.2025. Silloin FIONAn valmisaineistoissa vanhalla menetelmällä tuotetut pseudotunnisteet on korvattu uudella menetelmällä tuotetuilla tunnisteilla.

Samaan aikaan FIONAan viedään yleisimmistä pseudotunnisteista linkkitauluja, jotka yhdistävät vanhalla ja uudella tavalla tuotetut pseudotunnisteet. Näillä linkkitauluilla käyttäjät voivat suojata itse uudelleen W-levylle tekemänsä tiedostot, joissa on vielä käytössä vanhat pseudotunnisteet. Linkkitaulut ovat käytössä 1.10.2025 saakka, jonka jälkeen vanhoja tunnisteita ei enää tulisi käyttää. Linkkitaulut on tarkoitettu vain tälle siirtymäajalle, eikä niitä tule kopioida hankkeen W-levylle.

Koska uusittavien aineistojen määrä ja koko on hyvin suuri, aloitamme uusien aineistoversioiden toimittamisen FIONAan jo hyvissä ajoin ennen käyttöönottoa. Uudet tiedostot ovat samannimisiä kuin vanhat tiedostot. Kun uusia aineistoja toimitetaan FIONAn valmisaineistokansioihin, vanhat tiedostot viedään väliaikaisesti alikansioon "vanha_suojaus". Käyttöönoton yhteydessä nämä alikansiot poistetaan.

Uudistus edellyttää FIONAn käyttäjiltä tiettyjä toimia. Suurimpana muutoksena useimpien pseudotunnisteiden nimet muuttuvat, joka aiheuttaa koodeihin muutoksia. Jos aineistoa on tarve käyttää ennen kuin uudet pseudotunnisteet on otettu käyttöön, pitää vanhat tiedostot viedä ”vanha suojaus” -aliansioon.

Räätälöityjen ja itse muodostettujen aineistojen suojausta varten tutkijapalvelut toimittaa FIONAan linkkitauluja, jotka yhdistävät vanhalla ja uudella tavalla tuotetun aineiston. Jos valmisaineiston pseudotunnisteista laaditut linkkitaulut eivät riitä omien ja räätälöityjen tiedostojen suojaamiseen, voi tutkijapalveluita pyytää muodostamaan lisää tarvittavia uusia tauluja. Kun omat aineistot on uudelleensuojattu, käyttäjä voi pyytää niiden siirtämistä W-levyltä D-levylle.

Uuden ja vanhan pseudotunnisteen erottaa muuttujanimestä ja sisällöstä. Lähtökohtaisesti vanhan muuttujan nimi on muotoa stunnus ja uuden muotoa tunnus_s, mutta tunniste voi joissain tapauksissa olla myös eriniminen. Sisältö on kummassakin mielivaltaisen näköinen merkkisarja, mutta vanhassa sen pituus on 20 merkkiä sisältäen numeroita sekä ISOJA kirjaimia väliltä A-F, uudessa tunnisteessa on 36 merkkiä sisältäen numeroita, pieniä kirjaimia a-f sekä väliviivoja.

Uudistus on suuri ja pyrimme mahdollisimman saumattomaan muutokseen. Valitettavasti kaikkia häiriöitä huhtikuun ensimmäisen ja toisen viikon aikana ei voida sulkea täysin pois ja tällä voi olla vaikutusta aineistojen käytettävyyteen. Kehotamme huomioimaan tämän mahdollisuuksien mukaan tutkimusaineistojemme käytön suunnittelussa. Teemme parhaamme, jotta häiriöt olisivat mahdollisimman vähäisiä ja lyhytaikaisia.

Kiitos ymmärryksestänne ja yhteistyöstänne.

Tutkijapalvelut

Announcement to FIONA users regarding changes in pseudonymisation in Research Services on April 1, 2025

The pseudonymisation solution of Statistics Finland’s Research Services is being renewed. The new pseudonymisation method is better suited for pseudonymising various identifiers and improves the technical security of research data. The shnro that identifies personal data will remain unchanged for the time being, and any future changes will be announced separately.

The update will be implemented on April 1, 2025. At that time, the existing datasets in FIONA will be replaced with new versions where the pseudo identifiers produced by the old method are replaced with those produced by the new method.

Simultaneously, link tables of the most common pseudo identifiers will be introduced in FIONA, connecting pseudo identifiers produced by the old and new methods. With these link tables, users can re-pseudonymise their files on the W-drive that still use the old pseudo identifiers. The link tables will be available until October 1, after which the old identifiers should no longer be used. The link tables are intended only for this transition period and should not be copied to the project's W-drive.

Due to the large number and size of the datasets to be updated, we will start delivering the new dataset versions to FIONA well in advance of the implementation. The new files will have the same names as the old ones, and as new datasets are delivered to FIONA's ready-made dataset folders, the old ones will be temporarily moved to a subfolder named "vanha_suojaus". These subfolders will be deleted upon implementation.

The update requires certain actions from FIONA users. The biggest change is that the names of most pseudo identifiers will change, which will cause changes in program codes. If there is a need to use the dataset between the delivery of the new dataset and the implementation, the previously described "vanha_suojaus" subfolder path must also be added to the program code.

For the pseudonymisation of tailored and self-created datasets, link tables will be delivered to FIONA by Research Services, connecting datasets produced by the old and new methods. If the link tables created for the pseudo identifiers in the ready-made datasets are not sufficient to pseudonymise your own and tailored files, you can request Research Services to create the necessary new tables. Once your own datasets have been re-pseudonymised, you can request their transfer from the W-drive to the D-drive.

The difference between the new and old pseudo identifiers can be identified by the variable name and content. Generally, the old variable name is in the form "stunnus" and the new one in the form "tunnus_s", but the identifier may also have a different name in some cases. The content in both is an arbitrary-looking string of characters, but in the old one, its length is 20 characters, including numbers and UPPERCASE letters from A-F, while in the new identifier, it is 36 characters, including numbers, lowercase letters a-f, and hyphens.

The update is significant, and we naturally strive for the smoothest possible transition. Unfortunately, it is not possible to completely avoid all disruptions during the first and second weeks of April, and this may affect the usability of the datasets. We encourage you to consider this in your planning for the use of our research data. We will do our best to keep disruptions to a minimum and as short as possible.

Thank you for your understanding and cooperation.

Researcher Services

