

# Tutkijapalveluiden säännöt ja ohjeet

## Sisällys

1. Johdanto .....	1
2. Salassapitovelvoite.....	2
3. Etäkäyttöjärjestelmän säännöt .....	2
3.1 Järjestelmään kirjautuminen .....	2
3.2 Järjestelmän käyttäminen .....	2
3.3 Resurssit .....	3
3.4 Ylläpito .....	3
4. Tutkimuslaboratorion säännöt .....	3
4.1 Henkilökortin ja kulkuavaimen käyttö.....	3
4.2 Tutkimuslaboratoriossa työskentely.....	4
5. Tietojen suojaaminen ja tulosteiden tarkistusmenettely .....	4
5.1 Tulosteiden tietosuojavaatimukset.....	4
5.2 Tutkimustulosten tarkistusmenettely käytännössä .....	6
6. Tulosten julkistaminen .....	7
7. Hankkeen päätyttyä.....	8
8. Sanktiot .....	8
9. Vastuut .....	8
Liitteet.....	9

## Yhteystiedot

Tutkijapalvelut:  
Päivystäjä  
Valtteri Valkonen  
Satu Nurmi  
Marianne Johnson

puh. 029 551 2758  
puh. 029 551 3431  
puh. 029 551 2926  
puh. 029 551 3777

[tutkijapalvelut@tilastokeskus.fi](mailto:tutkijapalvelut@tilastokeskus.fi)

Mikrosimulointi:  
Sanni-Sandra Hellman  
Sampo Lappo

puh. 029 551 3210  
puh. 029 551 3922

[mikrosimulointi@tilastokeskus.fi](mailto:mikrosimulointi@tilastokeskus.fi)

## 1. Johdanto

Tilastokeskuksen Tutkijapalvelut muodostaa yksikötason aineistoja tieteellisiin tutkimuksiin ja tilastollisiin selvityksiin. Aineistojen käyttö voi tapahtua Tilastokeskuksen hallinnoiman FIONA-etäkäyttöpalvelun kautta ja tutkimuslaboratoriossa tai Tilastokeskuksen ulkopuolella anonymisoidulla otosaineistolla. Tilastokeskus asettaa yksikötason aineiston käyttöluvan ehdoksi että aineiston käyttäjä sitoutuu noudattamaan tutkijapalveluiden sääntöjä ja ohjeita. Salassapitovelvoitteeseen, tietojen suojaamiseen, tulosten julkistamiseen ja hankkeen päättymiseen liittyvät säännöt ja ohjeet koskevat kaikkea yksikötason aineistojen tutkimuskäyttöä. Aineistojen etäkäytölle ja tutkimuslaboratoriokäytölle on omat säännöt, ks. luvut 3 ja 4. Sekä etäkäytössä että tutkimuslaboratoriokäytössä noudatetaan tutkimustulosten tarkistusmenettelyä, ks. luku 5. Tutkijapalvelun etäkäyttöjärjestelmän käytännöt

koskevat myös SISU-mikrosimulointimallin etäkäyttöä. Poikkeuksena mikrosimuloinnissa noudatetaan erilaista käytäntöä tutkimustulosten tarkastamiselle.

## 2. Salassapitovelvoite

Tilastokeskuksen yksikkötason aineistojen käyttö edellyttää käyttö lupaa. Tietoja saa käyttää ainoastaan käyttöluvan saanut henkilö ja vain päätöksessä hyväksyttyyn tarkoitukseen. Tietojen kohteita ei saa yrittää tunnistaa aineistosta. Allekirjoittamalla salassapitositoumuksen tutkija sitoutuu siihen, ettei ilmaise eikä käytä omaksi hyödyksi tutkimuksen yhteydessä tietoonsa saamia, salassa pidettäviksi säädettyjä tietoja (tutkimusaineistoon sisältyvät yksikkötason henkilö- ja yritystiedot). Vaitiolovelvollisuus koskee myös tilastotuotantoon liittyviä atk-ohjelmistoja sekä ohjeistoja koskevia tietoja, joiden ilmitulo vaarantaa tietosuojan (tilastolaki 280/2004: 24 §). Vaitiolovelvollisuus on voimassa sopimuksen päätyttyäkin.

## 3. Etäkäyttöjärjestelmän säännöt

### 3.1 Järjestelmään kirjautuminen

Etätyöpöytäyhteys tutkijan työasemalta Tilastokeskuksen palvelimelle avataan verkkopalvelun kautta (Liite B). Käyttäjä kirjautuu verkkopalveluun käyttäjätunnuksella ja salasanalla, jotka ovat henkilökohtaisia ja niitä tulee säilyttää huolellisesti. Erityisesti käyttäjätunnukset ja salasanat tulee säilyttää erillään.

Käyttäjä tunnistetaan lisäksi tekstiviestikoodilla. Järjestelmä lähettää käyttäjän kännykkänumeroon ns. flash-viestin. Tähän numeroon liittyviä puhelimia sekä SIM-kortteja tulee säilyttää huolellisesti.

Etäkäytön tulee tapahtua yhteyden avaamisesta sopimuksen tehneen asiakkaan tiiloista jos ei muusta ole erikseen sovittu Tilastokeskuksen kanssa. Etäyhteyden saa ottaa vain asiakkaan määrittämistä IP-osoitteista. Tilastokeskus tarkistaa, että ne ovat käyttöehtojen mukaisia.

Etätyöaseman tietoturvapäivitysten sekä virus- ja palomuurisuojausten on oltava ajantasaisia.

### 3.2 Järjestelmän käyttäminen

Tutkimushankkeille varataan työkansio, jossa hankkeen materiaali säilytetään. Luku oikeudet aineistokansioihin annetaan hankekohtaisen tai SISU-mallin käyttöluvan mukaisesti. SISU-mallissa jokaiselle käyttäjälle varataan oma työkansio, johon käyttäjä voi halutessaan antaa oikeudet muille mallin käyttäjille. Mikrosimulointimallin käyttäjät voivat myös yhteisen Forum-kansion kautta jakaa tiedostoja etäkäyttöympäristössä muiden käyttäjien kesken.

Järjestelmää tulee käyttää vain käyttöluvassa mainittuun tarkoitukseen, eikä sen toimintaa saa haitata millään tavoin. Järjestelmän käytöstä kerätään lokitietoja laskutusta, järjestelmän ylläpitoa ja valvontaa varten<sup>1</sup>. Järjestelmän käytöstä laskutetaan tutkijapalvelun hinnaston mukaisesti.

<sup>1</sup> Rekisteriseloste TK-00-473-08

Etäyhteys on tarkoitettu vain asiakkaan nimeämän, aineistoon käyttöluvan saaneen, tutkijan käyttöön. Tästä syystä etälaite on lukittava, jos sen äärestä poistutaan ja laite jää valvomatta. Järjestelmästä tulee kirjautua ulos, kun etäyhteydellä ei enää työskennellä.

Aineistoa on käsiteltävä siten, ettei suojattava tieto joudu sivullisen tietoon. Suojausta voi parantaa muun muassa työtilojen järjestelyillä sekä näytön suojauksella. Käyttöluvallista aineistoa ei saa paljastaa millään tavalla henkilöille, joilla ei ole aineistoon käyttö lupaa.

Ohjelma- ja tulostiedostot kulkevat ainoastaan Tilastokeskuksen henkilökunnan kautta sähköpostitse osoitteesta [tutkijapalvelut@tilastokeskus.fi](mailto:tutkijapalvelut@tilastokeskus.fi). Tutkimustulokset tarkistetaan tietosuojan varmistamiseksi (ks. luku 5). Mikrosimuloinnissa tuotettujen tulosteiden tarkastusmenettely eroaa tästä.

### 3.3 Resurssit

Kullekin tutkimushankkeelle varataan levytilaa 300 GB. SISU-mikrosimulointimalin käyttäjille levytilaa on varattu 40 GB / käyttäjä. Jos hanke tai SISU-mallin käyttäjä tarvitsee levytilaa enemmän, tästä tulee käydä erillisiä neuvotteluita Tutkijapalveluiden kanssa. Järjestelmässä käytettävissä olevat laskenta- ja ohjelmistoresurssit ovat rajalliset ja niiden saatavuus riippuu muiden kirjautuneiden käyttäjien lukumäärästä. Päällekkäisten aineistojen ja välitiedostojen tallentamista ei suositella. Yhtäaikaista raskaita ajoja pyydetään välttämään. Yksittäiselle asiakkaalle tai hankkeelle voidaan tarjota omia laskenta- ja ohjelmistoresursseja erillisen sopimuksen mukaisesti.

### 3.4 Ylläpito

Tilastokeskus vastaa etäkäyttöjärjestelmän ylläpidosta virka-aikana. Ongelmista voi ilmoittaa sähköpostitse osoitteeseen [tutkijapalvelut@tilastokeskus.fi](mailto:tutkijapalvelut@tilastokeskus.fi) tai soittamalla tutkijapalveluiden päivystyspuhelimeen 029 551 2758. Ongelmatilanteissa käyttäjän tulee ensisijaisesti ottaa yhteyttä etäkäyttö sopimuksessa nimettyyn laitoksen yhteyshenkilöön. Tilastokeskus ei vastaa ohjelmistojen käytöstä.

Tilastokeskuksella on oikeus sulkea järjestelmä ylläpidollisista syistä. Ylläpitoon liittyvistä käyttökatkoista informoidaan etäkäytön avaussivun informaatiotiedotteessa sekä lähettämällä sähköpostiviesti organisaatioiden yhteyshenkilöille.

## 4. Tutkimuslaboratorion säännöt

### 4.1 Henkilökortin ja kulkuavaimen käyttö

Tutkimuslaboratoriossa työskentelevälle tutkijalle annetaan kuvallinen henkilökortti, jota on käytettävä aina liikuttaessa Tilastotalossa. Kortin puuttuessa on otettava yhteys Tilastokeskuksen henkilökuntaan.

Tutkimuslaboratoriossa voi työskennellä vain virka-aikana klo 8.00–16.15, ellei toisin ole sovittu.

Tutkijalle annetaan kulkuavain Tilastotalossa liikkumista varten. Tilastokeskukseen saavuttaessa leimataan kellokortilla "sisään" ja lähtiessä "ulos". Tauolle lähtiessä

leimataan vastaavasti "ulos" ja palatessa "sisään". Unohtuneet leimaukset ilmoitetaan sähköpostiosoitteeseen [tutkijapalvelut@tilastokeskus.fi](mailto:tutkijapalvelut@tilastokeskus.fi).

## 4.2 Tutkimuslaboratoriossa työskentely

Tutkimuslaboratoriossa työskentelevä tutkija kirjautuu FIONA-etäkäyttöjärjestelmään tutkimuslaboratorion työasemalta. Tutkijalle annetaan henkilökohtainen käyttäjätunnus ja salasana, jotka tulee säilyttää huolellisesti. Tutkimuslaboratoriossa työskentelevien tutkijoiden tulee noudattaa FIONA-työskentelyyn samoja sääntöjä ja ohjeita kuin omilta työasemiltaan järjestelmään kirjautuvien (ks. luku 3).

Työasema on lukittava, jos siltä poistutaan tilapäisesti, ja kone suljettava päivän päättyessä. Työasemalle voidaan jättää ajoja päälle poikkeustapauksissa erillisen sopimuksen mukaan. Tutkijapaikan varaukset voi hoitaa puhelimitse (029 551 3493 ja 2926) tai sähköpostitse ([tutkijapalvelut@tilastokeskus.fi](mailto:tutkijapalvelut@tilastokeskus.fi)). Tutkijapaikasta veloitetaan tutkijapalvelun hinnaston mukaisesti.

Ohjelma- ja tulostiedostot kulkevat ainoastaan tutkijapalveluiden henkilökunnan kautta sähköpostitse osoitteesta [tutkijapalvelut@tilastokeskus.fi](mailto:tutkijapalvelut@tilastokeskus.fi). Tutkimustulokset tarkistetaan tietosuojan varmistamiseksi (ks. luku 5).

## 5. Tietojen suojaaminen ja tulosteiden tarkistusmenettely

Salassapitovelvoitteen mukaan tutkijan on pidettävä huolta siitä, että tutkimustuloksissa ei ole yksikkötason tietoja tai mahdollisuutta niiden paljastumiseen. Tutkijapalveluilla on käytössä tutkimustulosteiden tarkistusmenettely, jolla halutaan varmistaa tietosuojan toteutuminen tutkijan aineistosta tuottamissa tulosteissa. Tutkijan tulee huolehtia, että tarkastukseen lähetettävät tulosteet täyttävät kappaleessa 5.1 luetellut tietosuojavaatimukset. Tulosteiden tulee olla selkeästi tulkittavia. Taulukoissa solukohtaisten havaintojen lukumäärän tulee olla näkyvissä, kuin myös estimaattien ja tunnuslukujen laskennassa käytettyjen havaintojen lukumäärät. Tarvittaessa Tilastokeskuksen henkilökunta antaa lisätietoa tietojen suojaamiseen liittyen.

### 5.1 Tulosteiden tietosuojavaatimukset

Seuraavassa käydään tarkemmin läpi erityyppisten tulosteiden tietosuojavaatimuksia.

#### *Frekvenssi- ja määrätaulukot*

Tutkijan julkaisemien tulosteiden tulee noudattaa tietosuojan osalta Tilastokeskuksen linjauksia taulukkomuotoisten tietojen suojaamisesta (liite A). Pääsääntönä yri-  
tystietojen suojaamisessa on, että kussakin solussa tai ryhmässä on oltava aina vähintään 3 (painottamatonta) havaintoa. Poikkeuksena Global Value Chains/International Sourcing -kyselyaineistojen osalta kynnysrajana on 5 havaintoa.

Kynnysarvosäännön rinnalla tulee soveltaa dominanssisääntöä<sup>2</sup> (1,75) tuoreissa yri-  
tystiedoissa (15 kk viiteajankohdasta). Poikkeuksena Palvelujen ulkomaankauppa -

---

<sup>2</sup> Tietoa dominanssisäännöstä liitteen A kappaleessa 3 Taulukkoaineistojen tietosuojaan liittyvä käsittely ja menetelmät

Päivitetty 21.4.2017

aineistojen yhteydessä dominanssisääntöä on sovellettava aina myös vanhojen tietojen osalta.

Toimipaikkatason tietoja suojatessa on varmistettava myös yritystason suojaus, eli kussakin solussa on oltava toimipaikkoja vähintään kolmesta eri yrityksestä. Samoin konsernitason suojaus on otettava huomioon yritysaineistoissa, joissa on tietoja konsernisuhteista. Hyödyketiedoissa (teollisuuden tuotantotilaston tuotteet sekä aineet ja tarvikkeet) yritysten lukumäärä on luottamuksellinen tieto kaikissa tuotantonimikkeissä.

Henkilötietojen suojaamisessa noudatetaan solukohtaista kynnysarvoa 3 ja kiinnitetään erityistä huomiota taulukoitavien muuttujien arkaluonteisuuteen. Yhdistetyissä työntekijä-työnantaja-aineistoissa on suojattava sekä henkilö- että yritystaso, eli kussakin taulukon solussa on oltava työntekijöitä vähintään kolmesta eri yrityksestä. Taulukkomuotoisissa yritystilastoissa esiintyviin ammatinharjoittajatietoihin sovelletaan samoja suojauskäytäntöjä kuin muihinkin yritystietoihin.

*Joidenkin erityisesti muiden viranomaisten luovuttamien, aineistojen suojaussäännöt voivat poiketa yllä mainituista. Nämä poikkeavat suojaussäännöt kirjataan käyttö lupaan..*

### Erilaiset jakaumatunnusluvut

Maksimi ja minimi liittyvät yleensä yhteen havaintoon. Jos tämä havainto on mahdollista tunnistaa, ei maksimia tai minimiä voi julkaista.

Jakaumapisteet (pl. minimi ja maksimi) muodostavat erikoistapauksen taulukosta, jossa solufrekvenssejä vastaavat jakaumapisteiden väliin jäävien havaintojen lukumäärät. Mikäli nämä lukumäärät ylittävät kynnysarvon 3, voidaan jakaumapisteet julkaista.

Moodi voidaan julkaista, mikäli (lähes) kaikki havainnot eivät saa samaa arvoa.

Keskiarvo, muut suhdeluvut ja jakaumatunnuslukujen korkeammat momentit (esim. varianssi) voidaan julkaista, mikäli niiden laskennassa on käytetty vähintään kolme havaintoa.

### Muut numeeriset tulostetyypit

Indeksipisteluvut, korrelaatiokertoimet ja testisuureet (t, F,  $X^2$ , yms.) voidaan yleensä julkaista, mikäli laskennassa on käytetty tarpeeksi (vähintään 10) havainnoja.

Regressiomallin kokonaisuudessaan voi julkaista, mikäli mallin taustalla on riittävästi havainnoja ja malli ei kuvaa aikasarjaa yhteen yritykseen/henkilöön perustuvista havainnoista. Mallin yksittäisiä kertoimia voidaan yleensä aina julkaista.

### Kuvat

Numeeristen tulosteiden tapaan myöskään kuvat eivät saa paljastaa yksittäisen havaintoyksikön tietoja. Aineistoista piirretyt kuvat ovat sallittuja, jos yksittäinen kuvapiste ei voi paljastaa sen taustalla olevaa yksittäistä havaintoa.

Kuvat viedään tarkistettavaksi taulukoiden tapaan selkeästi ja täsmällisesti dokumentoituna. Tarkistettavaksi soveltuvia kuvaformaatteja ovat esimerkiksi:

Päivitetty 21.4.2017

**Bittikarttaformaatit**

PNG (Portable Networks Graphics)  
BMP (Bitmap)  
JPEG (Joint Photographic Experts Group)  
TIFF (Tagged Image File Format)

**Vektoriformaatit**

EPS (Encapsulated PostScript)  
PS (PostScript)  
PDF (Portable Document Format)  
SVG (Scalable Vector Graphics)  
WMF/EMF (Windows Metafile)

Stata-ohjelmassa yllä mainittuja kuvaformaatteja pystyy luomaan graph export -komennolla. SPSS-ohjelmassa kuvaformaatin saa valita Export output -toiminnossa. R-ohjelmassa tietoa piirtofunktioista saa komennolla help(grDevices). Tiettyt kuvatyypit, kuten Statan gph-tiedostot, tallentavat lähtökohtaisesti kuvan piirtämiseen käytetyn aineiston, jonka vuoksi ne eivät välttämättä sovellu ulos siirrettäviksi.

Yksittäisten havaintoyksiköiden tietojen tunnistaminen on vaikeampaa, jos kuva on piirretty otosaineiston pohjalta. Monet yritysaineistot ovat kuitenkin kokonaisaineistoja, jolloin yksittäisten yritysten tietoja voi tunnistaa kuvista helpommin esimerkiksi ääriarvojen perusteella. Vaikka joku aineisto ei ole kokonaisaineisto kaikista Suomen yrityksistä, saattaa se olla sitä suppeamman osajoukon, kuten tietyn toimialan, suhteen.

Pylväsdiagrammit ja muut luokitellun aineiston esittämiseen käytetyt kuvat ovat tyypillisesti sallittuja julkaistaviksi, kunhan kussakin luokassa on riittävästi havaintoja. Tällainen aineisto voidaan yleensä esittää myös taulukkomuodossa ja siihen voidaan soveltaa suoraan Tilastokeskuksen taulukkoaineistojen suojausohjeita (liite A).

Jakaumakuvat sisältävät toisinaan poikkeavia havaintoja tai ääriarvoja, jotka voivat paljastaa havaintoyksikön tietoja. Tasoitetut tai riittävän karkealla asteikolla esitetyt jakaumat, histogrammit ja kertymäfunktiot ovat sallittuja. Ohjelmien piirtofunktiot merkitsevät usein automaattisesti mm. laatikkokuvaajiin (box plot) poikkeavat havainnot, jotka pitää jättää yleensä pois julkaistavista kuvista.

Hajontakuvia käytetään tyypillisesti kahden jatkuvan muuttujan arvojen esittämiseen, jonka vuoksi se on tietosuojaan kannalta edellisiä kuvaajia hankalampi. Hajontakuvien kohdalla tulee kiinnittää erityistä huomiota aineiston luonteeseen mm. otoksen koon, tiedon arkaluonteisuuden ja poikkeavien havaintojen esiintymisen kannalta.

Selkeästi kiellettyjä kuvatyyppejä ovat muun muassa poikkeavien havaintoyksiköiden arvoja esittävät kuvat tai hajontakuvat, joista pystyy päättämään esimerkiksi alansa suurimman yrityksen tietoja.

## 5.2 Tutkimustulosten tarkistusmenettely käytännössä

Tutkijapalveluiden etäkäytössä ja mikrosimulointimallin etäkäytössä noudatetaan toisistaan poikkeavia tarkistusmenettelyjä. Etä- ja tutkimuslaboratoriokäytössä tuotetut tutkimustulokset tarkistetaan ennen tietojen luovuttamista eikä etäympäristöstä voi siirtää tiedostoja omalle työasemalle itse, vaan tiedonsiirto tapahtuu erillisellä

pyynnöllä sähköpostin kautta. SISU-mikrosimuloinnissa puolestaan käyttäjä siirtää tutkimustuloksia sisältäviä tiedostoja suoraan omalle työasemalle ilman etukäteen tehtävää tarkistusta.

### *Tarkistusmenettely tutkimushankkeissa*

Etä- ja tutkimuslaboratoriokäytössä tuotetut tutkimustulokset tarkistetaan tietosuojan varmistamiseksi. Kaikkien tarkistukseen vietävien tiedostojen tulee täyttää samat kriteerit kuin julkaistaviksi tarkoitettujen taulukoiden ja kuvioiden. Esimerkiksi log-tiedostoista tulisi pyytää tarkistettavaksi vain välttämättömät tai julkaistaviksi tarkoitetut osiot. Tiedostot viedään tarkistukseen kopioimalla ne tarkistuskansioon (...out) ja lähettämällä tarkistuspyyntö sähköpostiosoitteeseen tutkijapalvelut@tilastokeskus.fi. Tulostiedostojen määrä ja koko tulee pitää kohtuullisena (enintään 2 Mb).

Tietosuojatarkistuksen jälkeen tulokset lähetetään tutkijan sähköpostiosoitteeseen. Tarkistamiseen on varattava 1-2 työpäivää.

### *Tarkistusmenettely mikrosimuloinnissa*

Mikrosimuloinnin etäympäristöstä sallitaan tutkimustulosten ulossaanti ilman etukäteen tehtävää tarkistusta. Näin ollen tutkija voi siirtää tutkimustuloksia sisältäviä tiedostoja etäympäristöstä omalle työasemalleen.

Jokaisella käyttäjällä on etäkäyttöympäristössä oma henkilökohtainen Mail-sähköpostikansio, jonka kautta tiedostoja voidaan siirtää käyttäjän omalle työasemalle. Siirto tapahtuu kopioimalla halutut tiedostot (User-, Forum- tai Admin-kansioista) käyttäjän henkilökohtaiseen Mail-kansioon. Noin kahden minuutin kuluttua kopiointista tiedosto siirtyy automaattisesti sekä käyttäjän henkilökohtaiseen että Tilastokeskuksen mikrosimuloinnin (mikrosimulointi@tilastokeskus.fi) sähköpostiin. Jokaisesta Mail-kansioon kopioidusta tiedostosta tulee erillinen sähköpostiviesti, jonka liitteenä on kansioon kopioitu tiedosto. Sähköpostissa ilmenee siirrettävän tiedoston nimi ja lähetyspäivämäärä. Lähetettävän tiedoston koko voi olla korkeintaan 1 megatavua (Mb).

Tilastokeskus tarkistaa siirretyt tiedostot mikrosimuloinnin sähköpostista jälkikäteen. Tutkija on velvollinen noudattamaan siirrettäviin tietoihin liittyen tutkijapalveluiden etäympäristöä koskevia ohjeita ja pelisääntöjä. Näitä ovat mm. siirrettävien tutkimustulosten suojaaminen, tiedostojen koko- ym. rajoitukset, tietojen tulokkavuus ja tulosten julkaiseminen (luku 6). *Aivan erityistä huomiota on kiinnitettävä siihen, ettei mikrosimulointiympäristöstä siirrettävissä tiedoissa ole vahingossakaan yksilötason tietoja tai mahdollisuutta niiden paljastamiseen.*

## **6. Tulosten julkistaminen**

Tutkija sitoutuu julkistamaan tutkimustulokset vain sellaisessa muodossa, ettei niistä voi tunnistaa yksittäisen yrityksen tai henkilön tietoja. Tämän varmistamiseksi tutkimusraportit ja -julkaisut voidaan vaatia tarkistettavaksi ennen tulosten julkistamista. Tutkijan tulee huomioida, että tietosuojatarkistukseen on varattava riittävästi aikaa (noin 1-2 viikkoa). Tutkimustulosten tarkistamisesta kannattaa sopia etukäteen tutkijapalvelujen henkilöstön kanssa.



Tuloksia julkistettaessa Tilastokeskus on mainittava tiedon lähteenä.

## 7. Hankkeen päätyttyä

Etäkäyttö- ja tutkimuslaboratoriahankkeissa syntynyt materiaali poistetaan hankkeen tai SISU-mallin käyttöoikeuden päätyttyä, jollei säilytyksestä ole tehty erillistä sopimusta. Tilastokeskuksen ulkopuolelle luovutettu tietoaineisto on palautettava/hävitettävä käyttöluvan voimassaolon päätyttyä. Kaikki aineistosta otetut kopiot ja välitiedostot on myös hävitettävä. Hävittämisestä on ilmoitettava Tilastokeskukselle.

## 8. Sanktiot

Jos tutkija tai asiakas toimii vastoin etäkäytöstä tehtyjä sopimuksia tai ohjeita, etäkäyttöyhteys katkaistaan. Yhteys avataan uudelleen, jos asiakas esittää hyväksyttävän kirjallisen perustelun rikkomuksen syistä ja toimenpiteet, joilla rikkomus jatkossa estetään.

Salassapitovelvollisuuden rikkomisesta rangaistaan rikoslain 38 luvun 1 ja 2 §:n mukaan. Salassapitovelvoite koskee myös tutkijoita. Rangaistus on sakko tai enintään vuosi vankeutta.

## 9. Vastuut

Jokaisen tutkimusaineistojen etäkäyttösopimuksen solmineen laitoksen tulee nimetä hallinnollisista ja teknisistä asioista vastaavat yhteyshenkilöt. Yhteyshenkilöiden tulee tukea laitoksen uusia etäkäyttäjiä, seurata Tilastokeskuksen tiedotuksia etäkäytöstä ja osallistua Tilastokeskuksen järjestämiin koulutustilaisuuksiin sekä yhdessä Tilastokeskuksen kanssa etsiä vastuksia ongelmatilanteisiin. Yhteyshenkilöt perehdyttävät aineistoja käsittelevät tutkijat etäkäyttöympäristöön, jotta voidaan varmistaa, että tutkijoilla on riittävät tekniset valmiudet aineistojen käsittelyyn eivätkä he tahattomasti vaaranna järjestelmän toimivuutta tai aineistojen tietosuojaa. Yhteyshenkilöiden tulee ilmoittaa Tilastokeskukselle kaikista organisaatiossaan tapahtuvista etäkäytön kannalta olennaisista tietoturvaan tai tietosuojaan vaikuttavista muutoksista.

Henkilö, joka allekirjoittaa laitoksen/organisaation puolesta sopimuksen tutkimusaineistojen etäkäytöstä, on vastuussa laitoksella FIONA-järjestelmässä työskentelevistä tutkijoista. Hän vastaa myös siitä, että tutkijoiden työskentely tapahtuu Tilastokeskuksen hyväksymistä tiloista ja että niissä voidaan käsitellä aineistoja siten, ettei aineistojen tietosuoja vaarannu. Vastuuhenkilö hyväksyy myös laitoksen osalta uudet etäkäyttäjät allekirjoittamalla heidän etäkäyttösitoumuksensa. Hän myös ilmoittaa välittömästi Tilastokeskukselle yhteys- ja vastuuhenkilöiden muutoksista.

Tutkijat vastaavat siitä, että noudattavat salassapitositoumusta sekä tämän dokumentin ohjeistuksia. Tutkijoiden tulee informoida Tutkijapalveluita projektin loppumisesta ja työpaikan vaihdoista. Työpaikan vaihdon yhteydessä tulee toimittaa Tutkijapalveluille uusi etäkäyttösitoumus.



## ***Liitteet***

LIITE A: Ohjeet tutkimusaineistoista muodostettujen taulukkoaineistojen suojaamiseksi.

LIITE B: Ohjeet etäkäyttöjärjestelmään kirjautumiselle

## *Liite A: Ohjeet tutkimusaineistoista muodostettujen taulukkoaineistojen suojaamiseksi*

### *1. Ohjeiden tarkoitus*

Nämä ohjeet on laadittu Tilastokeskuksen omien taulukkomuotoisten yritys- ja henkilötietojen suojausohjeiden pohjalta (TK-00-270-13 ja TK-00-271-13). Näiden ohjeiden avulla Tilastokeskus haluaa edistää vastuuntuntoisia tapoja toimia tietosuoja-asioissa ja helpottaa voimassaolevien lakien ja tilastoeettisten periaatteiden soveltamista tutkimusaineistosta tuotettujen tilastotaulukoiden julkaisemisessa.

Jokaisen Tilastokeskuksen tutkimusaineistoa käyttävän ja taulukkomuotoisia yritys- tai henkilötietoja julkaisevan tutkijan tulisi perehtyä näihin ohjeisiin sekä niiden taustalla oleviin lakeihin ja tilastoeettisiin periaatteisiin.

### *2. Ohjeiden soveltaminen*

Tilastolaissa (280/2004) Tilastokeskuksen tilastotarkoituksiin keräämien tietojen luovutuksesta tutkimuskäyttöön säädetään pykälässä 13. Kyseisen pykälän perusteissa todetaan seuraavaa<sup>3</sup>:

”Tietoja luovutettaessa olisi henkilötietojen sekä liike- ja ammattisalaisuutta koskevien tietojen suoja varmistettava tapauskohtaisesti käytännön toimenpiteillä kuten esimerkiksi edellyttämällä riittäviä tietoturvamenettelyjä sekä huolehtimalla tarvittavan tietojen käyttöä koskevan valvonnan ja seurannan järjestämisestä. – – Koska tieteellisen tutkimuksen lopputulokset ovat yleensä julkisia, tulisi niiden julkaisemisen yhteydessä lisäksi aina erikseen varmistaa, ettei tutkimuksen perustana olleita yksittäisiä tilastoyksiköitä olisi mahdollista tunnistaa tutkimuksen julkisista lopputuloksista.”

Edellisen nojalla on siis huolehdittava tiedonantajien yksityisyyden sekä liike- ja ammattisalaisuuden suojasta niiden tieteellisten tutkimusten lopputuloksissa, joissa aineistona on käytetty Tilastokeskuksen luovuttamia tietoja. Tutkijan tulee ottaa huomioon nämä suojausohjeet suunnitellessaan ja laatiessaan taulukkomuotoisia julkaisuja tutkimustuloksistaan. Tilastokeskus myös osaltaan tarkistusmenettelyn kautta huolehtii tietosuojan toteutumisesta.

### *3. Taulukkoaineistojen tietosuojaan liittyvä käsitteistö ja menetelmät*

Taulukkoaineistolla tarkoitetaan tilastoa, jossa yksikkötason aineistoa on aggregoitu ja järjestetty taulukkomuotoon. Taulukkomuotoisten henkilötilastojen tilastoyksikkö on esimerkiksi yksityinen henkilö, perhe, kotitalous tai asuntokunta. Taulukkomuotoisten yritystilastojen tilastoyksikkö on esimerkiksi yritys, toimipaikka tai konserni. Taulukkomuotoisissa yritystilastoissa esiintyviin ammatinharjoittajatietoihin sovelletaan samoja suojauskäytäntöjä kuin muihinkin yritystietoihin.

Taulukkoaineisto voi olla joko

- **frekvenssitaulukko**, joissa soluavrot ovat soluun kuuluvien tilastoyksiköiden lukumääriä, tai

---

<sup>3</sup> Hallituksen esitys eduskunnalle laeiksi tilastolain sekä maaseutuelinkeinotilastoista annetun lain 2 ja 3 §:n muuttamisesta (HE 154/2012).

Päivitetty 21.4.2017

- **määrätaulukko**, joissa soluarvot ovat jonkin taulukoitavan muuttujan (esim. liikevaihdon) summia, keskiarvoja tai muita vastaavia tunnuslukuja, tai
- edellisten yhdistelmä, jossa on näkyvissä sekä solufrekvenssit että määrätietoja.

Määrätaulukot ovat yritystietojen esittämisessä selvästi yleisempiä, kun taas henkilötietoja esitetään useammin frekvenssitaulukkoina.

Taulukkoaineistoon kohdistuu **paljastumisriski**, jos jokin taulukossa esiintyvä tilastoyksikkö on vaarassa paljastua. Solukohtaisen paljastumisriskin määrittelyssä käytetään **herkkyysääntöjä**. Yleisimmät herkkyysäännöt ovat

- **kynnysarvosääntö**, jonka mukaan solu on herkkä, mikäli siihen kuuluu vähemmän kuin ennalta määrätyn kynnysarvon verran tilastoyksiköitä.
- **dominanssisääntö** eli **(n,k)-sääntö**, jonka mukaan solu on herkkä, mikäli sen  $n$  suurinta tilastoyksikköä muodostavat vähintään  $k$  prosenttia solun kokonaisarvosta. Dominanssisääntöä käytettäessä tulee parametrit  $n$  ja  $k$  määritellä numeerisesti tilastoitavien yritysten tasapuolisen kohtelun varmistamiseksi.
- **p%-sääntö**, jonka mukaan solu on herkkä, mikäli solun kokonaisarvon perusteella laskettu suurimman tilastoyksikön arvon estimaatti poikkeaa enintään  $p$  prosenttia oikeasta arvosta.

Herkkyysääntöjä voidaan käyttää myös useampaa rinnakkain. Tällöin solu on herkkä, jos se yhdenkin käytetyn herkkyysäännön mukaan on herkkä.

Taulukoiden suojausmenetelmänä käytetään yleensä **peittämistä** tai **luokituksen muuttamista**.

- Peittämiseen kuuluu ensisijainen, paljastumisriskissä olevien solujen peittäminen ja toissijainen peittäminen. Toissijaisella peittämisellä varmistetaan, ettei taulukon rivi- tai saraketotaalien avulla pystytä paljastamaan ensisijaisesti peitettyjen solujen arvoja.

Peittäminen voidaan tehdä myös rivikohtaisesti. Jos taulukon johonkin rivitotaaliin kuuluu vain pieni määrä tilastoyksiköitä (vähemmän kuin käytetty kynnysarvo), peitetään kyseinen rivi kokonaisuudessaan huomioimatta sen eri soluissa olevien tilastoyksiköiden lukumäärää.

- Luokitusta muuttamalla taulukosta pyritään hävittämään paljastumisriskissä olevat solut yhdistämällä niitä sisältävät luokat muihin taulukon luokkiin. Luokituksen muuttaminen tarkoittaa usein käytännössä koko luokituksen karkeistamista.

Taulukoiden suojausmenetelmänä voidaan käyttää myös paljastumisriskissä olevien solujen arvojen muuttamista. Tällaisia menetelmiä ovat esimerkiksi pyöristäminen ja alkuperäisen soluarvon korvaaminen likimääräisellä satunnaisluvulla.

#### 4. Suositukset taulukkomuotoisten henkilötietojen suojaamiseen

Taulukkoaineiston sisältämien henkilöiden, yksittäisten perheiden tai asuntokuntien paljastumisriski on aina arvioitava taulukoita suunniteltaessa ja ennen tietojen julkaisemista. Tietosuojatoimenpiteet tulee kohdistaa niin, että paljastumisriski on riittävän pieni, mutta suojauksen myötä aineistosta ei menetetä tarpeettomasti tietoa. Ensisijaisesti tulee huomioida tilastoyksikön oikeus tietosuojaan, mutta samalla tulee muistaa yhteiskunnan ja kansalaisten oikeus yhteiskunnallista päätöksentekoa ja suunnittelua varten tarvittavaan luotettavaan tilastotietoon.

#### 4.1 Taulukkoon kohdistuvan paljastumisriskin arvioiminen

Taulukkoon kohdistuva paljastumisriski määrittelee tietosuojatoimenpiteiden tarpeellisuuden. Paljastumisriskiä määriteltäessä huomioidaan

- **taulukon arkaluonteisia tietoja sisältävät muuttujat,**
- **pienet solu- ja luokkafrekvenssit (kynnysarvosääntö),**
- perusjoukon suuruus,
- muuttujien määrä sekä
- tilastointialueen koko ja sijaintitarkkuus.

Tietojen suojaustarvetta arvioitaessa saattaa olla myös tärkeää se, ovatko taulukon arvot suhteellisia vai absoluuttisia, koskevatko tiedot yhtä vuotta vai ovatko tiedot useamman vuoden summia tai keskiarvoja ja käytetäänkö tilastoinnissa perusjoukkoa tiettyä väestön erityisryhmää (esim. ulkomaalaiset, rikoksentekijät, poliisit, työttömät tai korkeatuloiset).

Paljastumisriskiä arvioitaessa on kiinnitettävä erityistä huomiota arkaluonteisiin tietoihin, joita ovat esimerkiksi henkilötietolain mukaan arkaluonteiset tiedot<sup>4</sup>, jotka kuvaavat henkilön

- rotua tai etnistä alkuperää,
- yhteiskunnallista, poliittista tai uskonnollista vakaumusta tai ammattiliittoon kuulumista,
- rikollista tekoa, rangaistusta tai muuta rikoksen seuraamusta,
- terveydentilaa, sairautta tai vammaisuutta taikka henkilöön kohdistettuja hoitotoimenpiteitä tai niihin verrattavia toimia,
- seksuaalista suuntautumista tai käyttäytymistä tai
- sosiaalihuollon tarvetta tai hänen saamiaan sosiaalihuollon palveluita, tukitoimia ja muita sosiaalihuollon etuuksia,

sekä muut arkaluonteiset tiedot, kuten

- kuolinsyy,
- kieli, kansalaisuus, syntyperä tai synnyinmaa,
- tulot, velat ja varallisuus sekä
- pääasiallinen toiminta, ammattiasema, harvinainen ammatti tai muu sosioekonomista asemaa kuvaava muuttuja.

Jos paljastumisriski on olemassa, tulee taulukossa olevat henkilötiedot suojata riittävästi.

#### 4.2 Suojaussuosituksiset

Taulukkoon kohdistuva paljastumisriski kohdistuu aina tiettyihin soluihin. Määrittämällä nämä solut tai niitä sisältävät luokat ja suojaamalla ne sopivalla suojausmenetelmällä voidaan taulukon paljastumisriski laskea hyväksyttävälle tasolle.

##### *Kynnysarvon käyttö*

Paljastumisriskin sisältävät eli niin sanotut herkat solut taulukossa määritellään kynnysarvon avulla. Kynnysarvon tarkemmassa määrittelyssä on huomioitava samoja asioita kuin paljastumisriskin arvioinnissa. Kynnysarvoksi suositellaan rivikohtaisesti sovellettaessa vähintään 10 tilastoyksikköä ja solukohtaisesti sovellettaessa vähintään 3 tilastoyksikköä.

---

<sup>4</sup> Henkilötietolaki (523/1999), 11 §.

### *Suojausmenetelmät*

Solujen (ja samalla taulukon) suojaaminen voidaan toteuttaa muuttamalla taulukon rakennetta, peittämällä yksittäisiä soluarvoja tai kokonaisia rivejä, tai muuttamalla soluarvoja esimerkiksi pyöristämällä.

Jos taulukossa on paljon herkkiä soluja tai hervät solut ovat keskittyneet muutamaaan luokkaan, suojaaminen kannattaa tehdä taulukon rakennetta muuttamalla. Taulukon rakennetta voidaan muuttaa muuttujien luokituksia muuttamalla tai muuttujien määrää kontrolloimalla.

Taulukon muuttujien määrän tulee olla sitä vähäisempi mitä pienempien alueiden tilastoista on kyse. Pienaluetilastoissa tulee myös välttää usean muuttujan ristiintaulukointeja ja julkaista mieluummin suoria jakaumia.

Jos taulukossa on vain vähän herkkiä soluja, suojaus on suositeltavaa toteuttaa peittämällä. Peittäminen voidaan tehdä joko solukohtaisesti tai rivikohtaisesti eli peittämällä rivin kaikki solut.

### *Täydentäviä suosituksia*

**Moniulotteiset taulukot.** Jos henkilötietoja sisältävässä taulukossa on 3 tai useampia muuttujia, joista vähintään yksi muuttuja on arkaluonteinen tai aluetaso maakuntaa pienempi, on paljastumisriski hyvin todennäköinen.

**Tilastointialueen koko.** Suurempia alueita koskevissa tilastoissa, kuten maakunta- tai suuraluekohtaisissa tilastoissa, harvoin tarvitaan minkäänlaisia tietosuojatoimenpiteitä. Näissäkin tilastoissa perusjoukko ja luokitukset tulisi kuitenkin valita niin, ettei taulukkoon tule tarpeettomasti pieniä solufrekvenssejä.

Tilastointialueen koon pienetessä myös tilastoitavien yksiköiden määrä vähenee. Kuntapohjaiset tilastotkin voivat sisältää paljastumisriskin silloin kuin alueen perusjoukko on pieni. Kuntaa pienempien alueiden tilastoinnissa paljastumisriski on aina mahdollinen.

**Otostilastot.** Otanta vaikuttaa paljastumisriskiin. Riski on suurempi, jos tilastotiedot on tuotettu käyttäen aineistona koko perusjoukkoa, kuin jos otosaineistosta on asetelmapainojen avulla tuotettu koko perusjoukkoa koskevia estimaatteja. Kuitenkin myös otosaineistoista muodostettujen henkilötaulukoiden paljastumisriski tulee arvioida. Tietosuojan ja estimaattien laadun varmistamiseksi tulee otoksiin perustuvissa tilastoissa käyttää kynnsarvoa solujen herkkyyden määrittelyssä. Kynnsarvo voi kuitenkin olla pienempi, kuin vastaavassa, koko perusjoukon tietoihin perustuvassa tilastossa käytettäisiin.

**Tapaus tilastoyksikkönä.** Tilastoitaessa tapauksia (esim. rikostapaukset, liikenneonnettomuudet) ei välttämättä ole kyse henkilötilastoista. Tapaustilastoihin on kuitenkin sovellettava henkilötilastojen suojaussuosituksia, mikäli tilastosta on mahdollista tunnistaa yksittäinen henkilö tai paljastaa tämän ominaisuuksia.

## **5. Suositukset taulukkomuotoisten yritystietojen suojaamiseen**

Taulukkomuotoisten yritystilastojen suojaaminen voidaan toteuttaa eri tavoin riippuen erilaisista tilastojen laadintaan liittyvistä tekijöistä. Seuraavassa esitetään kolmiportainen yritystietosuojan porrastus, johon kaikki tietosuojan toteuttamistavat ovat ryhmiteltävissä:

1. Tilanteessa, jossa vain tarkka paljastuminen on arka asia, riittää **kynnysarvosäännön** käyttö. Kynnysarvosääntö on oletussääntö. Kynnysarvon tulee aina olla vähintään 3.
2. Kun likimääräinen yritystiedon paljastuminen on arka asia, on herkkyyssääntönä käytettävä **dominanssisääntöä** tai **p%-sääntöä**. Dominanssi- tai p%-sääntön käyttö tulee kuitenkin rajata vain tuoreisiin tilastotietoihin ja rinnalla on aina käytettävä myös kynnysarvosääntöä. Tieto on tuore niin kauan kuin sen paljastuminen vaikuttaa markkinatilanteeseen tai yksittäisen yrityksen toimintaan. Aikaraja tiedon tuoreudelle ja dominanssisäännön käytölle on 15 kuukautta viiteajankohdasta. Vanhempien tietojen osalta tulee käyttää kynnysarvosääntöä.
3. Jos suojaus on mahdollista toteuttaa **salaamalla tiedonantajien identiteetti ja lukumäärä**, tämä on suositeltavaa. Esimerkkinä tästä ovat otosaineistoista lasketut estimaatit, joiden yhteydessä ei julkaista tietoa otokseen kuuluvista tilastoyksiköistä.

Tilastotaulukkoa ei tarvitse suojata, jos siihen ei kohdistu paljastumisriskiä tai sen sisältämät tiedot on lailla säädetty julkisiksi. Esimerkiksi valtion ja kunnallisten viranomaisten toimintaa ja julkisten palvelujen tuottamista kuvaavat tiedot ovat pääosin julkisia.<sup>5</sup> Yli 25 vuotta vanhat yrityksiä koskevat tiedot ovat myös julkisia.<sup>6</sup>

#### Täydentäviä suosituksia

**Aikaulottuvuus.** Yritystietojen suojausmenettelyistä päätettäessä tulee aikaulottuvuus ottaa huomioon. Yritystietojen relevanssi vähenee merkittävästi, mitä enemmän aikaa on kulunut tilaston viiteajankohdasta julkaisuajankohtaan. Hyvin ajankohtaiset lyhyen aikavälin tilastot tulee suojata myös likimääräiseltä paljastumiselta.

**Solufrekvenssit.** Tietosuoja vaarantamatta pienetkin solufrekvenssit voidaan määritä taulukossa julkaista, vaikka varsinainen soluarvo olisi suojattu. Toisaalta tilastoyksiköiden lukumäärän salaaminen on soveltuviissa olosuhteissa keino, jolla itse taulukoitavan muuttujan tiedot saadaan julkisuuteen.

**Toimipaikkatilastot.** Toimipaikkatilastoja suojatessa on varmistettava myös yritystason suojaus. Solun paljastumisriskiä määriteltäessä on kiinnitettävä huomiota sekä solussa olevien toimipaikkojen lukumäärään, että eri yritysten lukumäärään, joille toimipaikat kuuluvat.

**Otoksiin perustuvat tilastot.** Otanta ei välttämättä yksin ole riittävä suojausmenetelmä, sillä otokseen pohjautuvassa tilastossa suurimmat yritykset ovat yleensä mukana ja useimmiten suurin mielenkiinto ja paljastumisriski kohdistuu juuri suuryrityksiin. Myös otosaineistoista muodostettujen yritystaulukoiden paljastumisriski tulee arvioida. Paljastumisen mahdollistavat estimaatit ja indeksipisteluvut on suojattava. Estimaatti tai pisteluku voi mahdollistaa paljastumisen, jos sen laskennassa on käytetty vain muutaman yrityksen tietoja. Tällöin myös estimaatin luotettavuus kärsii.

<sup>5</sup> Tilastolaki (280/2004), 12 §.

<sup>6</sup> Laki viranomaisten toiminnan julkisuudesta (621/1999), 31 §. Jos kyseessä on yksittäinen ammatinharjoittaja (henkilötieto), niin salassapitoaika on 50 vuotta ko. henkilön kuolemasta.

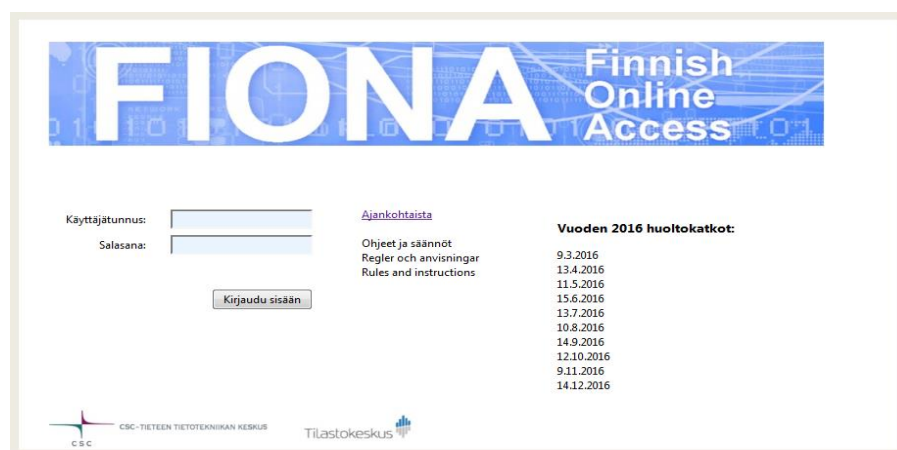
## Liite B: Kirjautuminen ja työskentely etäkäyttöjärjestelmässä

Tämä dokumentti kuvaa etäkäyttöympäristöön kirjautumista ja palvelun käyttöä. Kirjautuminen etäkäyttöympäristöön poikkeaa hieman riippuen siitä mitä nettiselainta tutkija käyttää. SISU-mikrosimulointimallin etäkäyttöympäristön työskentelyohjeet on kuvattu mallin käsikirjassa.

### 1. Etäkäyttöympäristöön kirjautuminen käyttäen IE tai Firefox-selainta

1. Kirjautumissivu löytyy Tilastokeskuksen tutkijapalveluiden internet-sivuilta (Tilastokeskus -> Tuotteet ja palvelut -> Tutkimuspalvelut -> Mikroaineistot -> Etäkäyttö) <http://tilastokeskus.fi/tup/mikroaineistot/etakaytto.html>
2. Seuraavaksi valitaan kirjautuminen  
- FIONA, CSC-ympäristö. Kaikissa uusissa projekteissa työskentely tapahtuu FIONA-järjestelmässä.

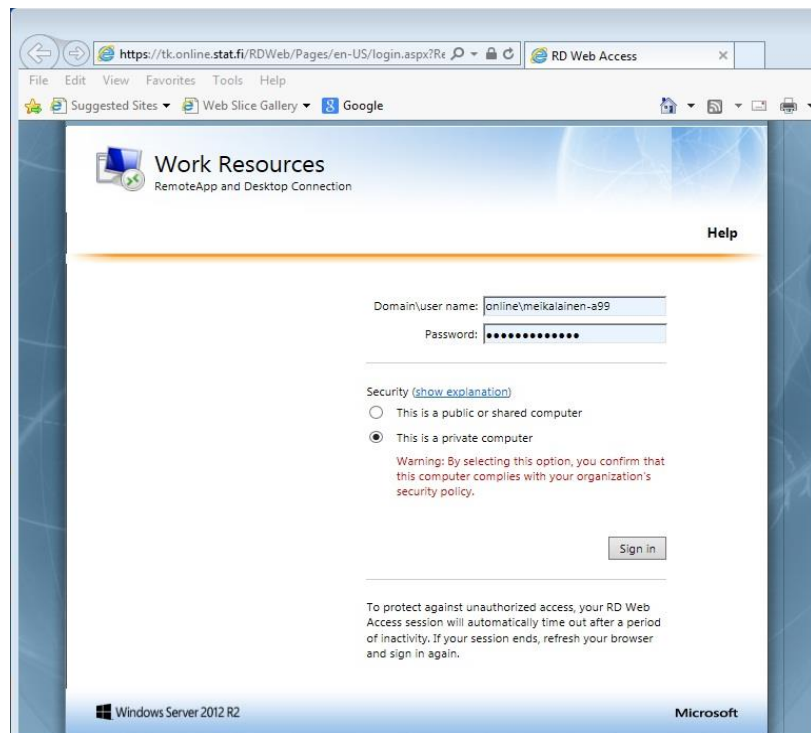
Yhteyden voi myös ottaa suoraan osoitteeseen <https://fiona.stat.fi>



3. Ensimmäisellä käyttökerralla hyväksytään (Allow) lisäosan asennus This webpage wants to run the following add-on: "Microsoft Remote Desktop Services Web Access Control..."
4. Kirjautumisessa käytetään vain yhtä käyttäjätunnusta. Kirjaudu antamalla domain -nimen ( online) sekä käyttäjätunnuksesi. Tunnus kirjoitetaan käyttäen hankenumeroa esim. Online\meikalainen-a99. Rastita "This is a private computer" ennen Sign in -painikkeen painamista.



Päivitetty 21.4.2017

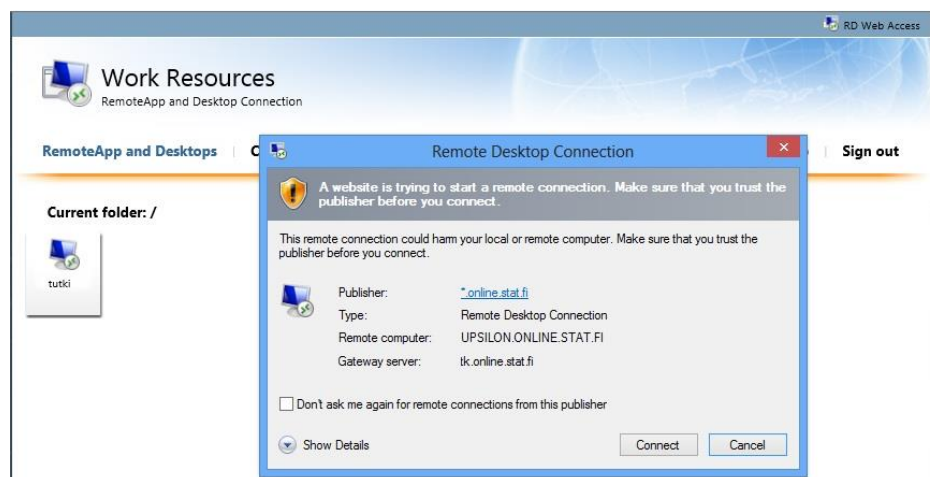


5. Work Resources –sivulle tulee näkyviin palvelinryhmät. Tutki -kuvakkeesta pääsee etäkäyttöpalvelimille Alfa, Beta, Gamma, Delta, Zeta ja Eta. Niissä on 4 CPU:ta/palvelin.

Niille, jotka ovat ilmoittaneet tarvitsevansa SAS ohjelmistoa, näkyy myös kuvake Tutki-SAS. Kuvakkeen takaa löytyvät etäkäyttöpalvelimet, joissa on mm. SAS 9.4 (Penta, Heksa, Hepta ja Okta. 4 cpu/palvelin).

Mikrosimuloijat pääsevät Mikrosim –kuvakkeesta mikrosimuloinnin etätyöpöydille.

Varoitusikkuna varoittaa tuntemattomasta palvelimesta Upsilon, joka on istuntojen kuormanjakaja. Voit laittaa rastin kohtaan Don't ask me again... ja painaa Connect.

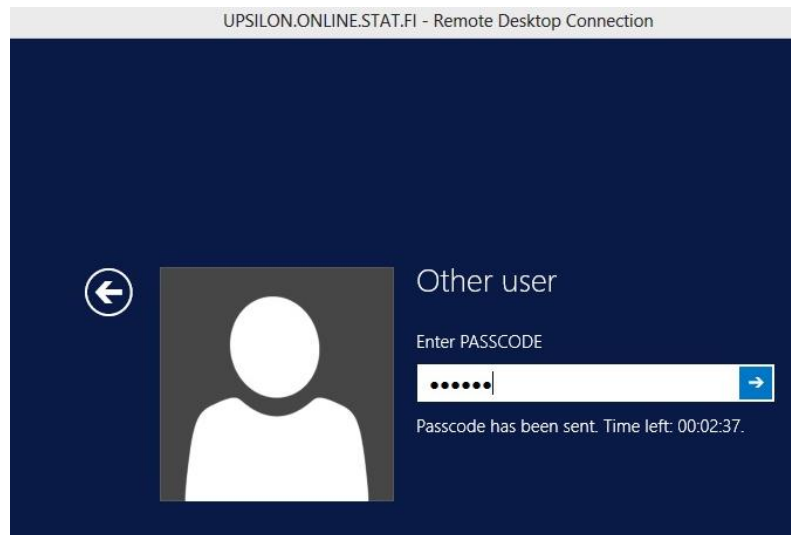


6. Jos käytät Firefox-selainta, tulee tässä kohdassa ensimmäisellä yhteydenotto-kerralla ylimääräinen kysymys:

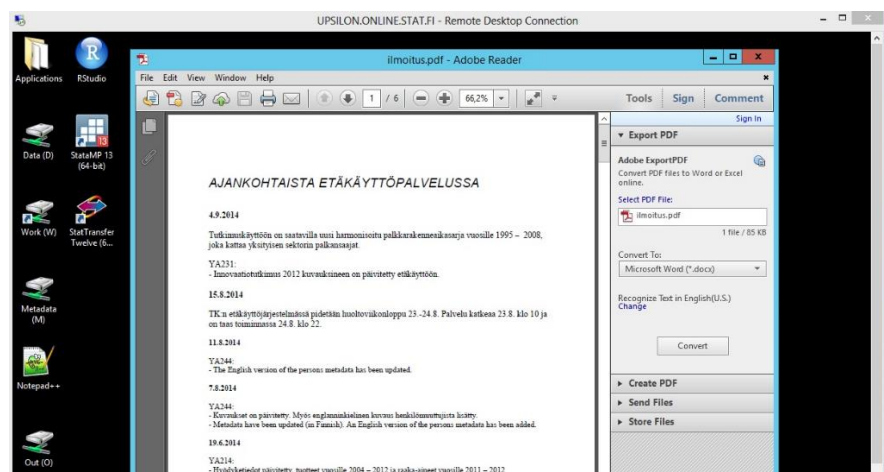
Päivitetty 21.4.2017

Olet avaamassa tiedostoa cpub-tutki-tutki-CmsRdsh.rdp. Mitä tiedostolle tehdään? Valitaan Aava ohjelmalla: Remote Desktop Connection (oletus) ja laite- taan rasti kohtaan ”Tee näin oletuksena tämän tyyppisille tiedostoille.”

7. Jos käytät Firefox-selainta, tulee tässä kohdassa käyttäjätunnuksen kysely uudelleen. Käytetään samaa hanketunnusta kuin kohdassa 4.
8. Seuraavaksi ilmestyy SMS Passcode –tunnistautuminen. Syötä ”Enter Passcode” ruutuun GSM-numeroosi tullessa viestissä oleva kertakäyttösalasana.



9. Tämän jälkeen avautuu etätyöpöytä.



## 2. Yleiset työskentelyohjeet tutkimusaineistoilla

Etäkäyttöjärjestelmä toimii kymmenellä Windows-palvelimella. Käyttäjät ohjataan kirjautumisvaiheessa sille palvelimelle, jolla on kirjautumishetkellä vähemmän kuormaa. Työskentely tapahtuu vastaavasti kuin tavallisella Windows työasemalla. Projektit on nimetty juoksevin numeroin alkaen esimerkkiprojektista a01.

Kaikki materiaali säilytetään työhakemistossa W:\a01. Hakemistosta otetaan varmuuskopio kerran vuorokaudessa, jolla pyritään turvaamaan tiedot levy-palvelimen rikkoutumisen varalta. Käyttäjille näkyvät seuraavat levyt:

Päivitetty 21.4.2017

D: Data	sisältää tutkijapalveluiden valmisaineistoja, lukuoi- keudet riippuvat käyttöluvasta
M: Metadata	sisältää aineistokuvauksia
O:\a01 Out	hakemisto tulosten siirtoa varten
W:\a01 Work	työhakemisto

Järjestelmän kuormittaminen saattaa haitata muiden työskentelyä. Palvelin hidastuu huomattavasti, jos järjestelmästä loppuu muisti (swap-muistin käytön vuoksi). Tämän vuoksi käyttäjien tulee välttää turhaa muistin kulutusta.

Ongelmatilanteissa otetaan yhteys oman organisaation yhteyshenkilöön. Yhteyshenkilöt vastaavat ohjelmistojen käyttötuesta. Tilastokeskuksen tuki vastaa vain ohjelmistojen ja käyttäjätunnuksien toimivuudesta.

### Aineistot ja ohjelmistot

Tutkimusaineistot ovat SAS 7 -muodossa (.sas7bdat). Aineistojen voi siirtää halua-  
maansa muotoon Stat/Transfer-ohjelmalla. Järjestelmässä on käytettävissä seuraavat ohjelmat:

R  
Python Anaconda + komponentit  
Stata  
SAS  
SPSS Statistics (rajattu käyttäjien lukumäärä)  
Octave  
Stat/Transfer  
Graphviz  
Kdiff  
Libre Office (tekstinkäsittely ja taulukkolaskenta)  
Lyx, Latex, MikTeX  
Notepad++  
Pnet, Pajek  
Rstan, Rtools, Rstudio, TramineR (liittyvät R-ohjelmiin)  
Strawberry  
QGIS

Tilastokeskus ei tarjoa tukea käytössä oleville sovellusohjelmille.

### Stata-ohjelmat

Järjestelmään on asennettu Stata 12, Stata 13 ja Stata 14. Ohjelman 64-bittisyys tarkoittaa, että se pystyy lukemaan suuria aineistoja muistiin. Liian suuri muistin varaus johtaa kuitenkin järjestelmän hidastumiseen, jonka vuoksi Statalle saa varata enintään 4000 MB muistia.

### Työskentelyn lopettaminen

Log Off

Kun lopetat työskentelyn työpäivän päätteeksi, kirjaudu järjestelmästä ulos. Uloskirjautuminen vapauttaa järjestelmän resurssit.

Valitse Logoff-pikakuvake työpöydältä (tai start-valikosta).

Päivitetty 21.4.2017

Sulje selainikkuna, jossa kirjautumissivu näkyy.

#### Disconnect

Jos pidät lyhyen tauon työskentelyssä, voidaan etäyhteys katkaista ilman ulos-kirjautumista. Samaan istuntoon voi palata kirjautumalla uudelleen. Yhteys katkaistaan sulkemalla etäyhteysikkuna (tai start-valikosta Shut Down ja Disconnect). Disconnect jättää istunnon auki, jonka vuoksi resurssit eivät vapaudu (ohjelmistolisenssit ja käyttäjälle varattu muisti). Tämän vuoksi pitää aina muistaa tehdä Log Off työpäivän päätteeksi.

Etätyöpöytäyhteys lukittuu, jos se on käyttämättömänä 10 minuuttia. Yhteys katkaistaan, jos se on käyttämättömänä 30 minuuttia. Ohjelmia ei suljeta. Avaamalla yhteyden uudelleen päästään jatkamaan työskentelyä.

#### Tulosten siirto

Kaikki tutkimustulokset siirretään järjestelmästä ulos ylläpitäjien toimesta. Tulosten tarkistukseen viemisessä tulee noudattaa erityistä harkintaa. Tulosten viemistä tarkistettavaksi pienissä erissä tulee välttää. Tarkistukseen tarkoitettujen taulukoiden ja kuvioden tulisi olla sisällöltään siinä muodossa, missä ne on tarkoitus julkaista.

Huolehdi, että noudatat taulukointisääntöjä (katso säännöt liitteineen).

Dokumentoi tulostaulukoinnit huolellisesti tarkastettaviin tiedostoihin. Jokaisen taulukon pitää olla itsensä selittävä samaan tapaan kuin journali artikkeleissa. Havaintojen lukumäärän tulee näkyä kussakin ryhmässä.

Siirrä tiedostot kansioon O:\a01 (esimerkkiprojekti)

Lähetä tulosten siirtopyyntö [tutkijapalvelut@tilastokeskus.fi](mailto:tutkijapalvelut@tilastokeskus.fi) (mainitse projektin järjestysnumero).

Tulokset lähetetään tarkastuksen jälkeen tutkijan sähköpostiosoitteeseen.