



Tilastokeskus 

TILASTOKESKUKSEN TIETOTILINPÄÄTÖS 2018

Raportti henkilötietojen käsittelystä
28.5.2019

Pääjohtajan katsaus

Tilastokeskuksen tehtävä on tuottaa luotettavaa ja ajantasaista tietoa asiakkaiden hyödyksi ja yhteiskunnalliseen keskusteluun. Toimintamme rakentuu eri lähteistä saatavien tietojen ja niiden käsittelyn varaan. Työmme laadukkaan lopputuloksen kannalta on olennaista, että käytössämme on hyvin hallintoituja tietoaaineistoja ja henkilöstö, joka osaa niitä käsitellä ja hyödyntää.

Nyt käsillä oleva tietotilinpäätös on järjestyksessä toinen Tilastokeskuksessa laadittu tietotilinpäätös. Tämä vuoden 2018 raportti täydentää ensimmäistä vuoden 2017 toiminnasta kertovaa raporttia. Tässä tilinpäätöksessä keskityimme kertomaan erityisesti siitä, mitä tietosuoja ja tietoturva varmistavia toimenpiteitä Tilastokeskuksessa tehtiin vuonna 2018. Osana tietotilinpäätöstä on myös mittaristo, johon on koottu keskeisiä kuvaajia henkilötietojen käytöstä tilastotuotannossa.

Lainsäädäntö on ohjannut tilastoviranomaisen tietojen käsittelyä jo pitkään, ja tietojen asianmukaiselle käsittelylle on olemassa pitkät perinteet. Oman lisänsä lainsäädäntöön toi tilinpäätösvoimalla astunut yleinen tietosuoja-asetus. Vuonna 2018 jatkoimme toteutusprojektia, jonka tarkoituksena oli tehostaa entisestään henkilötietojen tietosuojalainsäädännön mukaista käsittelyä Tilastokeskuksessa.

Vuoden 2018 aikana järjestettiin eri rooleissa työskenteleville tilastokeskuksellaisille tietojen käsittelyssä ja hallinnassa tarvittavaa lisäkoulutusta. Henkilöstö osallistui laajasti muun muassa tietosuoja-asioista järjestettyihin infoihin. Tilinpäätösvoimalla panostettiin myös tietoturvaan kehittämällä esimerkiksi käyttövaltuushallintaa ja parantamalla tuotantojärjestelmien lokitietojen keräämistä ja seuranta.

Kansikuva: Shutterstock
Taitto: Tarja Soikkeli

Tietoja lainattaessa lähteenä on mainittava Tilastokeskus.

© 2019 Tilastokeskus

Tiedon käyttö lisääntyy yhteiskunnassa entisestään, ja tiedon varaan halutaan rakentaa Suomelle hyvää tulevaisuutta. Pyrkimys ilahduttaa ja saa kannatusta myös Tilastokeskuksesta. Tavoitteiden saavuttamiseksi tarvitsemme yhteistyötä laajasti yli organisaatorajojen ja ratkaisuja, jotka auttavat meitä käyttämään olemassa olevia tietovarantoja hyödyksi uusin tavoin, eettiset näkökulmat huomioiden. Näiden yhteisten asioiden eteen haluamme Tilastokeskuksessa tehdä töitä nyt ja tulevaisuudessa.

Tilastokeskuksen pääjohtaja

Marjo Bruun



Sisällys

Pääjohtajan katsaus	3
1 Tiivistelmä	7
2 Tilastokeskuksen strategia	7
3 Kokonaisarkkitehtuuri	8
4 Tietoarkkitehtuuri	10
5 Muuttuvan tietosuojalainsäädännön vaikutus Tilastokeskuksen toimintaan	12
6 Henkilöstön osaamisen varmistaminen	13
7 Tietoturvan hallinta	14
8 Tiedonhankinta	17
8.1 Henkilötietojen elinkaari	17
8.2 Tiedonkeruurekisteri	19
8.2.1 Hallinnolliset ja yksityisen sektorin aineistot	20
8.2.2 Suorat tiedonkeruut	20
8.3 Pseudonymisointi tietoaineistojen vastaanotossa	21
8.4 Verkkoharavointi uutena tiedonkeruumenetelmänä	22
8.5 Henkilötiedonkeruiden kehittäminen vuosina 2015–2019	23
8.6 Koski-palvelun käyttöönottoon valmistautuminen	24
8.7 Tulorekisterin käyttöönottoon valmistautuminen	25
9 Henkilötietovarannon uudistamishanke (STIINA)	26
10 Yritys- ja toimipaikkarekisteri	27
11 Avoin data	28
11.1 Tilastokeskuksen tuottama avoin data	28
11.2 Tilastokeskuksen hyödyntämä avoin data	29
12 Toimeksiantoina tehtävä henkilötietojen käsittely	29
13 Asiakas- ja sidosryhmätiedot	31
14 Asianhallinta ja arkistointi	32
14.1 Asianhallinta ja analoginen asiakirja-aineisto	32
14.2 Tilastoaineiston arkistointi	33
15 Mittarit	33

15.1	Tilastot	33
15.2	Tilastokeskuksen tilastotuotannossaan käyttämien aineistojen vastaanotto.	35
15.3	Aineistojen vastaanoton tavat vuonna 2018	35
15.4	Saapuvien aineistojen frekvenssit 2018	36
15.5	Aineistotoimitukset Tilastokeskukseen tilastotuotannon käyttöön vuonna 2018	36
15.6	Arkaluonteisia*/erityisiä tietoryhmiä** kuvaavia tietoja	37
15.7	Suorat tiedonkeruut	38
15.9	Vuonna 2018 toteutetut suorat henkilöperustiset tiedonkeruut . . .	39
15.10	Henkilö- ja kotitalousperustiset suorat tiedonkeruut vuonna 2018 otoskoon mukaan	39
15.11	Henkilötieto ja suoralla tiedonkeruulla kerättävät tiedot vuonna 2018	40
15.12	Arkaluonteinen/erityisiin tietoryhmiin kuuluva tieto suorassa tiedonkeruussa vuonna 2018.	41
	15.12.1 Haastattelujen määrät vuosina 2017–2018	41
15.13	Tilaustutkimukset	42
	15.13.1 Toimeksiantojen lukumäärä vuosittain: asiakasrahoitteiset tilaustutkimukset vuosina 2010–2018.	42
	15.13.2 Tilaustutkimukset 2018 otoskoon mukaan	42
15.14	Arkistointi	43
	15.14.1 Arkistoitujen tilastoaineistojen lukumäärät vuosina 2001–2018	43
15.15	Tutkimusaineistot	44
	15.15.1 Tutkimusaineistojen käyttöluvat ja luovuttaminen.	44
	15.15.2 Tutkimusaineistojen käyttöluvapahakemukset ja tutkimusaineistotoimitukset 2017 ja 2018	45
15.16	Kansallinen etäkäyttöpalvelu	46
	15.16.1 Tutkimusaineistojen etäkäytön mittarit 2010-luvulla	46

1 Tiivistelmä

Tämä on Tilastokeskuksen toinen tietotilinpäätös. Ensimmäisessä vuotta 2017 koskevassa tietotilinpäätöksessä kuvattiin mm. Tilastokeskuksen toimintaa ja Tilastokeskuksen asemaa Suomen tilastotoimessa, Tilastokeskusta koskevaa lainsäädäntöä, tilastotuotannon laatua ja ammattietoutta, tilastotuotantoa, tilastotarkoituksiin kerättyjen tietojen luovuttamista sekä tilastotiedon jakelua. Näihin kuvauksiin voi tutustua Tilastokeskuksen ensimmäisessä tietotilinpäätöksessä.

Tämä toinen tietotilinpäätös täydentää ensimmäistä raporttia ja siinä kerrotaan, mitä tietosuojaa ja tietoturvaa edistäviä toimenpiteitä Tilastokeskuksessa on tehty vuonna 2018. Tietotilinpäätöksen osana on mittaristo, johon on koottu keskeisiä kuvaajia henkilötietojen käytöstä tilastotuotannossa.

Avainsanat: tietotilinpäätös, tietosuojaa, tiedonhankinta, asianhallinta

2 Tilastokeskuksen strategia

Tilastokeskuksessa on tehty strategian uudistamistyötä yhteistyössä henkilöstön ja sidosryhmien kanssa syksyn 2018 aikana. Uusi visio ja strategiset tavoitteet ovat voimassa vuodesta 2019 eteenpäin.

Strategia pohjautuu viiteen Tilastokeskuksen toiminnan kannalta tärkeään toimintaympäristön muutostekijään, jotka ovat

- ilmiöpohjaisuus
- data ja mikroaineistot tilastojen rinnalle
- toimialan uusi järjestys
- toimintamallien murros
- yhteentoimivuus.

Tilastokeskuksen visio kirjoitettiin muotoon ”Tiedosta tulevaisuus”. Strategiset tavoitteet ovat ”kestävä tietoekosysteemi” ja ”tietoa ja ymmärrystä ilmiöitä ja tarpeita ennakoiden”.

Tilastokeskuksen strategian ytimessä on tieto ja sen käyttö. Strategian perustaksi valitut muutostekijät linkittyvät sisällöltään toisiinsa, ja useaan niistä liittyy suomalaisen yhteiskunnan tavoite käyttää nykyistä tehokkaammin hyödyksi eri lähteistä syntyvää tietoa. Eri lähteistä tulevaa tietoa halutaan yhdistellä toisiinsa nykyistä paremmin ja tehdä se eettisesti ja läpinäkyvästi niin, että tietojen käyttö on luotettavaa ja avointa. Toiminnan tulee olla sitä kautta yhteiskunnassa hyväksyttävää. Jokainen tietoja toiminnassaan käsittelevä organisaatio on vastuussa siitä, että luottamus säilyy eri toimijoiden välillä.

Tilastokeskus on ollut tiedon käyttäjien mielestä pitkään luotettava toimija, ja strategiassa pidetään tärkeänä säilyttää tämä luottamus toiminnan kehittämisen ja uudenlaistenkin tehtävien perustana. Tilastokeskus haluaa profiloitua erityisesti laadukkaalla ja eettisellä tietotuotannolla sekä tieto-osaamisella. Tietosuoja-asiat ja tiedonantajien näkökulma korostuvat entisestään kun tavoitellaan tiedon entistä parempaa yhteentoimivuutta ja aineistojen nykyistä laajempaa avaamista. Luottamus on jatkossakin Tilastokeskuksen toiminnan perusta ja Tilastokeskus näkee, että julkishallinnon organisaatioiden tulee toimia yhteistyössä niin, että kansalaisten luottamus säilytetään ja tiedon tehokas käyttö on mahdollista.

3 Kokonaisarkkitehtuuri

Tilastokeskuksessa ollaan parhaillaan siirtymässä arkkitehtuurityössä uudelle tekemisen tasolle. Tavoitteena on järjestelmällinen kokonaisarkkitehtuurityö koko organisaation kattavana kokonaisuuden hallintana strategiasta ja kyvykkyyksistä alkaen. Muutos käynnistettiin esiselvitystyönä vuonna 2017. Tätä on

seurannut kokonaisarkkitehtuurityötä Tilastokeskuksessa suunnitteleva ja käynnistävä projekti (2018–2019). Työtä nivotaan tiiviisti yhteen Tilastokeskuksen strategian uudistamisprosessin kanssa.

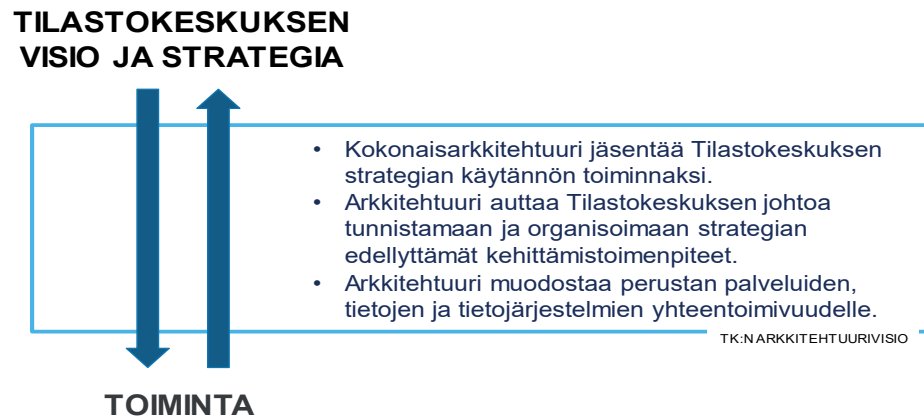
Kokonaisarkkitehtuuri (KA) ei ole uusia asia Tilastokeskuksessa, mutta aiemmin se on nähty ensisijaisesti yleiskäyttöisten ja yhtenäisten prosessien ja tietojärjestelmien kehittämiseen tähtäävänä määrittelytyönä. Työ on keskittynyt ns. ruohonjuuritasolle, kun yksittäiset projektit ja hankekokonaisuudet ovat hyödyntäneet työssään erilaisia arkkitehtuurikuvauksia. Tavoiteltu yleiskäyttöisyys ja yhtenäisyys ovat toteutuneet esimerkiksi luokitusten osalta, kun on rakennettu kaikkia tilastoprosesseja hyödyntäviä palveluja erillisten komponenttien sijaan.

Tilastokeskuksen arkkitehtuurivisio asettaa nyt käynnissä olevalle kokonaisarkkitehtuurityölle kolme selkeää tavoitetta:

- Kokonaisarkkitehtuuri jäsentää Tilastokeskuksen strategian käytännön toiminnaksi.
- Arkkitehtuuri auttaa Tilastokeskuksen johtoa tunnistamaan ja organisoimaan strategian edellyttämät kehittämistoimenpiteet.
- Arkkitehtuuri muodostaa perustan palveluiden, tietojen ja tietojärjestelmien yhteentoimivuudelle.

Uudenlaista lähestymistapaa on hyödyntänyt ensimmäisten joukossa sijaintitietokokonaisuus, jolle on tuotettu Tilastokeskuksen strategiaan ja visioon pohjautuva tavoitetila.

Kuvio 1. Kokonaisarkkitehtuuri Tilastokeskuksen vision ja strategian välineenä



4 Tietoarkkitehtuuri

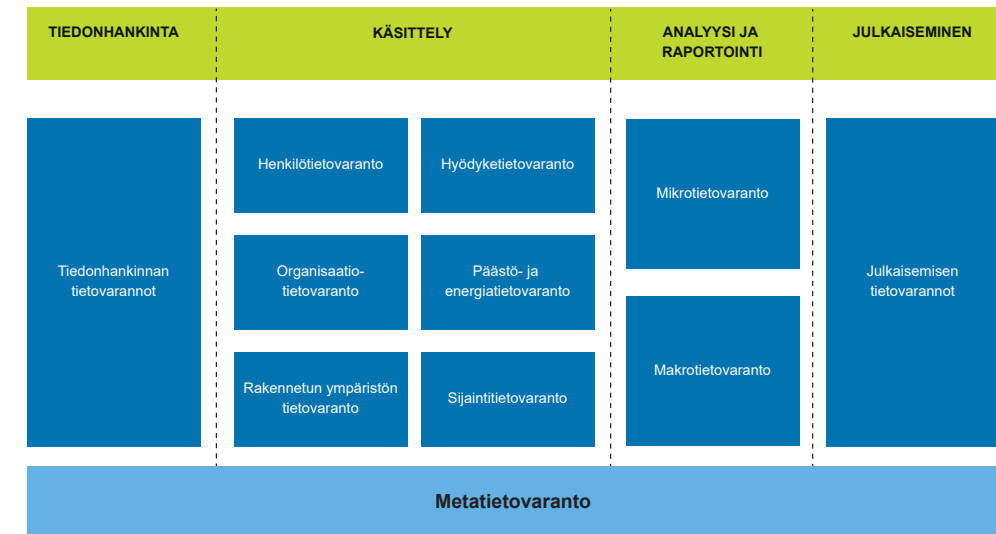
Tietoarkkitehtuurihanke edisti vuosina 2017–2018 tilastotuotannon tietoarkkitehtuurin kehittymistä ja tuotti ylätasoa tietoarkkitehtuurimäärittämiä. Tilastotuotannon tietoarkkitehtuuri -hankkeessa tuotettujen tietoarkkitehtuurikuvausten keskiössä on tieto. Suhteessa nykytilaan tietolähtöisyyden voidaan katsoa olevan ajattelutavan muutos perinteiseen tilastolähtöiseen kehittämistoimintaan verrattuna. Keskittymällä tietoihin voidaan tavoitella määritellä siten, että tiedot ovat yhteiskäyttöisiä sekä laadukkaita, ja toiminnassa voidaan välttää tarpeetonta kopiointia ja tietojen päällekkäisyyttä tai epäyhtenäisyyttä.

Tietoarkkitehtuurihanke tuotti tilastoprosessille loogisten tietovarantojen ylätasoa tavoitetilan ja vei eteenpäin Tilastokeskuksen tilastotuotannon Generic Statistical Information Model (GSIM) -pohjaista käsitelmää. Hanke myös suunnitteli Tilastokeskuksen tieto-objektien tunnistamisen hallinnan tulevaisuuden ja koordinoi Tunnistepalvelun määrittelyä ja toteuttamista. Tämän lisäksi hanke hahmotteli käsitteiden hallinnan tulevaisuutta ja yhteyksiä kansalliseen yhteentoimivuuteen ja sen välineisiin.

Tilastotuotannon tavoitetilan ylätasoa loogiset tietovarannot ilmenevät kuviossa 2. Tilastoja tuotetaan käsittelemällä ja yhdistelemällä tietoja yhdestä tai use-

ammasta tietovarannosta. Tietovaranto voi sisältää useiden tietojärjestelmien tietokantoja tai rekistereitä. Vastaavasti sama looginen tietovaranto voi sisältää useiden eri tahojen hallinnoimia tietoja, vaikka tiedot sijaitsisivatkin samassa fyysisessä tietokannassa.

Kuvio 2. Tilastotuotannon loogiset tietovarannot



Yksittäisten tietovarantojen täsmällinen sisältö ja rakenne tulee määritellä erikseen. Henkilöpohjaisten tilastojen STIINA-kehittämishankkeessa toteutettiin tarkemman tason loogisten tietovarantokokonaisuuksien kuvaus henkilötietokokonaisuuden osalta. Seuraava vaihe kohti tavoitetilaa etenemisessä on tarvittavien käsitteellisen ja loogisen tason kuvausten tuottaminen kaikille tarkemille osakokonaisuuksille.

5 Muuttuvan tietosuojalainsäädännön vaikutus Tilastokeskuksen toimintaan

Vuonna 2018 Tilastokeskuksessa jatkettiin toteutusprojektia, jonka tarkoituksen oli osaltaan varmistaa henkilötietojen tietosuojalainsäädännön mukainen käsittely Tilastokeskuksessa.

Tilastokeskuksessa kartoitettiin viraston omaa henkilöstöä, asiakas- ja sidosryhmiä sekä tilastotuotantoa koskevat aineistot. Kaikkien näiden osalta tehtiin joitain toimenpide-ehdotuksia, joiden toteutumista on vuoden aikana seurattu. Toiminnanohjausjärjestelmään tehtiin laajennus ja sinne kuvattiin Tilastokeskuksen saapuvat aineistot ja suorat tiedonkeruut henkilötietojen osalta.

Tiedonantajille suunnattua informointia parannettiin lisäämällä Tilastokeskuksen verkkosivuille tietosuojasivu ja uudistamalla tiedonkeruiden sivuja. Henkilötietolain edellyttämät rekisteriselosteet päivitettiin tietosuojaselosteiksi osana tietosuojasivustoa vastaamaan EU:n yleisen tietosuoja-asetuksen vaatimuksia ja ne julkaistiin suomeksi ja ruotsiksi. Tilastokeskuksen oman henkilöstön henkilötietoja koskevat selosteet pidetään saatavilla viraston sisäisessä verkossa.

Tilastokeskuksen omissa ja asiakasrahoitteisissa suorissa tiedonkeruissa tiedonantajille toimitettavia informointikirjeitä uudistettiin. Tilastokeskus tarkensi suoran tiedonkeruun kohteena olevan henkilön informointia sekä verkkosivullaan että paperilla tapahtuvassa ensimmäisessä yhteydenotossa eli saatekirjeessä, jonka Tilastokeskus lähettää vastaajalle.

Tilastokeskuksen tietovarantojen ylläpito on ulkoistettu. Vuosien 2017–2018 aikana palvelujen hankintaa koskevia sopimuksia on täydennetty tietojenkäsittely- ja/tai turvallisuussopimuksilla. Näin on varmistettu, että henkilötietojen käsittelijät noudattavat tietosuojalainsäädännön vaatimuksia. Sopimuksia on tehty yhteensä 15 sopimuskumppanin kanssa. Näistä merkittävimmät ovat Valtion tieto- ja viestintätekniikkakeskus Valtorin, Valtion talous- ja henkilöstö-

hallinnon palvelukeskus Palkeiden ja Tieteen tietotekniikan keskuksen CSC:n kanssa tehdyt sopimukset.

Tietosuoja koskevien vaikutustenarviointien tekeminen aloitettiin vuoden 2018 alkupuolella, jolloin tilattiin ulkopuolelta konsultilta tietosuojan vaikutustenarviointi kahden tilaston tuotantoprosessiin. Toinen arviointi tehtiin tilastoon, joka laaditaan eri viranomaisilta saaduista hallinnollisista aineistoista ja toinen tilastoon, jossa tietoja kerätään suoraan rekisteröidyiltä. Tämän jälkeen oma henkilöstö on jatkanut vaikutustenarviointien tekemistä omien tilastojensa osalta.

Henkilötietojen käsittelyyn liittyvää ohjeistusta on päivitetty vuoden 2018 aikana. Uusi, pienempi päivityskierros on tulossa vuoden 2019 alussa, kun tietosuojalaki on tullut voimaan.

Tilastokeskuksen ensimmäinen tietotilinpäätös valmistui kesäkuussa 2018

6 Henkilöstön osaamisen varmistaminen

Tilastokeskus järjesti tietosuojakoulutusta henkilöstölleen vuoden mittaan lukuisia kertoja. Keväällä 2018 jatkettiin yksikkökohtaisia koulutuksia. Lisäksi koko henkilöstölle järjestettiin keväällä kaksi tietosuojainfoa.

Koulutussuunnitelman mukainen uusille työntekijöille suunnattu, kolme kertaa vuodessa pidettävä tietosuojainfo järjestettiin ensimmäistä kertaa lokakuussa 2018. Tietosuoja oli esillä myös esimiesten koulutustilaisuuksissa. Lisäksi tilastotehtävissä työskenteleville järjestettiin kurssi tilastollisesta tietosuojasta kaksi kertaa vuonna 2018.

Tiedonhankintayksikössä tietosuojasioiden koulutusta on suunniteltu käsiteltävän säännöllisesti henkilöstön kanssa muutamia kertoja vuodessa. Suunnitelman mukaisesti tietosuojakoulutusta järjestettiin tiedonhankintayksikössä kesäkuussa 2018. Tiedonhankintayksikössä järjestettiin koulutusta myös noin 150:lle Tilastokeskuksen palveluksessa olevalle haastattelijalle. Koulutuksen sisältö vaihteli kohderyhmän mukaan.

Lisäksi Tilastokeskuksessa on valmisteltu ohjeistusta henkilötietojen käsittelystä tiedonkeruussa. Ohjeluonnosta on käsitelty laajasti yksikön henkilöstön kanssa osana ohjeen valmistelua. Tällä tavoin on pyritty laajentamaan ja syventämään henkilöstön osaamista henkilötietojen käsittelystä ja tietosuojasta sekä tietoturvasta.

Tietosuojasiointia käsiteltiin myös pohjoismaisten kollegavirastojen haastattelutiedonkeruuseen liittyvässä tapaamisessa kesällä 2018.

Julkisen hallinnon tietohallinnon neuvottelukunta (JUHTA) ja julkisen hallinnon digitaalisen turvallisuuden johtoryhmä (VAHTI) järjestivät 2017–2018 kaksi tietosuojan yhteishanketta, joiden työpajatilaisuuksiin Tilastokeskus osallistui.

7 Tietoturvan hallinta

Tilastokeskuksen tietoturvaperiaatteet toteuttavat viraston toimintaan liittyvää lainsäädäntöä. Tietoturvaperiaatteet ovat johdon hyväksymät ja henkilökunta soveltaa niitä työssään. Tilastokeskus hyödyntää toiminnassaan myös julkisen hallinnon digitaalisen turvallisuuden johtoryhmän (VAHTI) laatimaa ohjeistusta.

Käyttövaltuuksien hallinta

Tilastokeskuksen käyttövaltuushallinnan kehittämiseksi käynnistettiin vuonna 2018 kaksi projektia.

Käyttövaltuushallinta – toteutus -projektissa tavoitteena oli määritellä uusi roolipohjainen käyttövaltuuksien hallintamenetelmä. Työroolipohjaisella käyttövaltuushallinnalla saavutetaan tilanne, jossa käyttöoikeudet samanlaista työtehtävää hoitavien keskuudessa ovat täysin yhteneväiset, eikä käyttöoikeuksissa ole enää henkilöittäin tapahtuvaa vaihtelevuutta. Tämä helpottaa myös käyttöoikeuksien hallintaa kokonaisuutena sekä turhien käyttöoikeuksien pois karsimista.

Käyttövaltuushallinnan järjestelmän (IdM) hankinta -projekti puolestaan keskittyi hankkimaan Tilastokeskuksen käyttöön soveltuvan käyttövaltuushallinnan järjestelmän. Ilman tarkoitukseen soveltuvan järjestelmän tarjoamaa tukea käyttövaltuushallinnan kokonaisvaltainen kehittäminen olisi mahdotonta ja uuden prosessin jalkauttaminen vaikeaa. Järjestelmä asettaa myös puitteet niille käytännöille, joiden mukaisesti käyttövaltuushallintaa tullaan toteuttamaan: järjestelmän ohi tapahtuvia käyttövaltuushallinnan menettelyjä ei siis jatkossa tulla sallimaan kuin niiden järjestelmien tai vastaavien osalta, joita ei eritellystä syystä saada järjestelmätuen piiriin.

Projekti päättyi valitsemaan käyttövaltuushallinnan järjestelmäksi Valtorin tarjoaman Avain-palvelun, joka on keskitetty, valtionhallinnon käyttöön tarkoitettu käyttövaltuushallinnan ratkaisu ja joka on käytössä myös muissa valtionhallinnon virastoissa. Käyttöönoton esiselvitys toteutettiin kesällä 2018 ja järjestelmän käyttöönotto alkoi syksyllä 2018. Vuoden 2018 loppuun mennessä järjestelmän testiympäristö oli valmis, mutta järjestelmän hyväksyntätästä ja käyttöönotto jää vuoden 2019 alkupuolelle.

Tekninen tietoturva

Tilastokeskuksen järjestelmien teknisestä ylläpidosta vastaa Valtion tieto- ja viestintäteknikkakeskus Valtori yhdessä Tilastokeskuksen tietohallinnon kanssa. Tilastokeskuksen tietohallinto ylläpitää ja kehittää viraston omia järjestelmiä ja sovelluksia tilastointitarkoituksiin. Näiden tietoturvasta huolehditaan mm. tekemällä säännöllisiä auditointeja. Auditoinnit suorittaa aina ulkopuolinen toimija.

Tilastokeskuksen alihankkijat tai palveluntuottajat, joiden henkilöstö liikkuu Tilastokeskuksen tiloissa tai joiden henkilökunta käsittelee salassa pidettäviä tietoja Tilastokeskuksen toimeksiannosta, allekirjoittavat salassapitositoumuksen. Henkilötietojen käsittelyyn liittyvät velvoitteet kuten salassapitositoumuksen allekirjoittaminen, on kirjattu myös tietojen käsittelijöiden kanssa tehtyihin sopimuksiin.

Lokitiedot

Tilastokeskuksen tuotantojärjestelmien lokitiedot tallennetaan keskitettyyn lokijärjestelmään, jota Valtori ylläpitää. Erityistä huomiota on kiinnitetty siihen, että lokitietoja ei voi muuttaa järjestelmään kirjaamisen jälkeen. Uusi järjestelmä otettiin käyttöön keväällä 2018.

Tietoturvapoikkeamien raportointi

Tilastokeskuksen järjestelmiin ja toimintaympäristöön kohdistuneet tietoturvapoikkeamat raportoidaan ja kirjataan keskitettyyn poikkeamienhallintajärjestelmään. Poikkeamien arvioinnista vastaa viraston valmiusryhmä. Tietoturvapoikkeamia käsitellään viraston johtoryhmässä vähintään kaksi kertaa vuodessa.

Jatkuvuus ja riskienhallinta

Toiminnan jatkuvuudesta ja riskienhallinnasta vastaa viraston valmiusryhmä. Ryhmä kartoittaa riskejä osana riskienhallintaprosessia yhdessä viraston johdon

kanssa. Riskit arvioidaan ja niille nimetään vastuutahot. Tietojärjestelmiin kohdistuvia riskejä seurataan viraston tietohallinnossa.

Poikkeamatilanteiden harjoittelu osana jatkuvuuden varmistamista

Vuoden 2018 aikana Tilastokeskuksessa harjoiteltiin erilaisten tietoturva- ja suojaloukkausten varalta. Harjoituksissa testattiin henkilöstön osaamista paineen alla sekä kriisiviestintää niin omalle henkilökunnalle kuin mediallekin.

Tilastokeskus osallistui myös valtakunnalliseen Väestörekisterikeskuksen järjestämään TAISTO18-harjoitukseen.

Sisäinen tarkastus

Vuoden 2018 aikana valmistui kaksi tarkastusta ja kaksi selvitystä sisäisestä tarkastuksesta. Tarkastettavina alueina olivat käyttöoikeudet ja niiden hallinnointi sekä hallinnollisten aineistojen vastaanoton toimivuus ja varmuus. Selvityksen kohteina olivat henkilöstön etätyö ja toiminta monitoimitiloissa. Tarkastuksissa ja selvityksissä annettujen suositusten toimeenpano on käynnistynyt.

8 Tiedonhankinta

8.1 Henkilötietojen elinkaari

Tilastointitarkoituksiin kerättävän tiedon elinkaari Tilastokeskuksessa määritellään tiedonkeruukohtaisissa tiedonkeruupäätöksissä tai tiedontuottajien kanssa tehdyissä sopimuksissa. Digitaalisessa muodossa olevien tilastoaineistojen pysyvä säilyttäminen perustuu Kansallisarkiston tekemiin päätöksiin.

Henkilötiedon elinkaari alkaa, kun Tilastokeskus saa toisen viranomaisen tai muun tiedontuottajan tietoaineiston hallintaansa (Kuvio 3.). Kansalaisilta tai

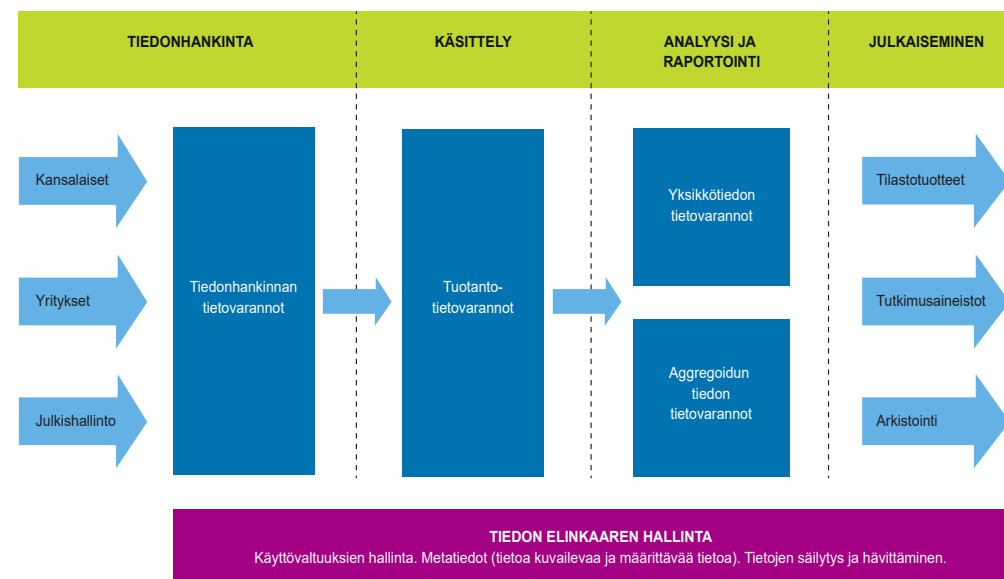
yrittäjiltä suoraan kerättävien tietojen elinkaari alkaa sillä hetkellä, kun tiedonantaja vastaa Tilastokeskuksen tiedusteluun. Suoraan tiedonantajilta kerätyt tiedot sekä tiedontuottajilta saadut tiedot talletetaan tiedonkeruukohtaisiin tietovarantoihin, joissa tieto on sellaisena kuin se on Tilastokeskukseen saatu.

Tietojen käsittely tapahtuu erillisellä tietojen käsittelyalueella, jossa eri lähteistä kerätyt yksittäistä henkilöä tai organisaatiota koskevia tietoja yhdistetään ja käsitellään tilastotiedon tuottamisen vaatimusten mukaisesti. Käsittely tallentuu ns. tuotantotietovarantoihin. Tuotantotietovarantoja ja aiemmin mainittuja tiedonkeruukohtaisia tietovarantoja säilytetään tilastoprosessien vaatimuksista riippuen eripituisia aikoja.

Tuotantotietovarannon perusteella muodostetaan kunkin ilmiön kuvaamisessa tarvittavat yksikötason (esimerkiksi henkilö tai yritys) tietovarannot, jotka säilytetään pysyvästi. Pysyvä säilyttäminen tarkoittaa sitä, että tiedot ja tietoa koskevat kuvaukset (aineiston ja sen muuttujien kuvaukset) arkistoidaan siten, että niitä on mahdollista käsitellä uudelleen ilmoitetuissa käyttötarkoituksissa vuosikymmentenkin jälkeen.

Yksikötietovarannosta muodostetaan myös aggregoituja tietoja sisältävät tilastotietovarannot, joista tuotetaan erilaisia tilastotuotteita kuten julkistuksia ja taulukoita. Tässä vaiheessa käsiteltävät tiedot ovat tilastotietoa, jota voidaan jakaa erilaisina tietotuotteina yhteiskunnan käyttöön. Näihin tilastoinnin tuloksiin voi jokainen tutustua mm. Tilastokeskuksen internetsivuilla.

Kuvio 3. Henkilötietojen elinkaari



8.2 Tiedonkeruurekisteri

Tilastot laaditaan joko juuri niitä varten kerätyn tiedon pohjalta tai esimerkiksi käyttämällä muiden viranomaisten hallinnolliseen käyttöön keräämiä tietoja. Tilastolaki edellyttää, että tietoja kysytään yrityksiltä ja kansalaisilta vain siinä tapauksessa, että tarvittavaa tietoa ei ole jo olemassa.

Tilastokeskuksen toiminnanohjausjärjestelmän osana on vuodesta 2006 alkaen toiminut tiedonkeruurekisteri. Siihen kuvataan tilastotuotannossa hyödynnettävät muilta toimijoilta joko tilastolakiin tai (vakiintuneisiin) kahdenvälisiin sopimuksiin perustuen saatavat hallinnolliset rekisterit ja yksityisen sektorin valmiit aineistot sekä tiedonantajilta suoraan kerättävien tietojen pohjana olevat **suorat tiedonkeruut**.

8.2.1 Hallinnolliset ja yksityisen sektorin aineistot

Virallisen tilastotuotannon kannalta hallinnollisella aineistolla tarkoitetaan mitä tahansa aineistoa, joka on ensisijaisesti kerätty muuta tarkoitusta kuin virallista tilastotuotantoa varten. Tällaisia aineistoja ovat esimerkiksi Väestötietokeskuksen väestötietojärjestelmä sekä Verohallinnon ja Kelan rekisterit. Yksityisen sektorin toimijoiden kanssa on lisääntyvässä määrin sovittu aineistojen hankinnoista tilastotuotannon käyttöön.

Tiedonkeruurekisteriin kuvattu aineisto voi koostua useammasta tiedostosta ja tulla Tilastokeskukseen useassa toimituserässä. Eri tiedontoimittajilta tulevat tiedostot kuvataan omiksi aineistoikseen, vaikka aihe/ilmiö olisi sama.

8.2.2 Suorat tiedonkeruut

Tilastokeskuksella on yksityishenkilöille, kotitalouksille, yrityksille, yhteisöille ja julkishallinnolle kohdistettuja suoria tiedusteluita. Suorat tiedonkeruut on kuvattu Tilastokeskuksen verkkosivuilla osoitteessa <http://tilastokeskus.fi/keruu>.

Henkilöpohjaiset tiedonkeruut perustuvat yleensä otoksiin. Tutkimuksiin valitaan tilastollisin perustein satunnaisesti ihmisiä, jotka edustavat laajaa joukkoa suomalaisista esimerkiksi iän tai asuinpaikan perusteella. Vastaavasti yritys vastaaja edustaa muita yrityksiä kokonsa, sijaintinsa, toimialansa tai henkilöstömääränsä pohjalta. Tämä on myös EU:n yleisen tietosuoja-asetuksen minimointiperiaatteen mukaista.

8.3 Pseudonymisointi tietoaineistojen vastaanotossa

Pseudonymisointi eli tunnusten korvaaminen satunnaisilla tunneilla tietoaineistojen vastaanotossa otetaan vaiheittain käyttöön kaikissa henkilötietoa sisältävissä tilastoprosesseissa. Henkilötietojen pseudonymisointi on keino estää henkilöiden suora tunnistaminen. Tilastokeskuksessa on toteutettu palvelu, jonka mahdollistaa tilastotuotannossa käsiteltävien henkilötietojen sisältävien tietojen pseudonymisoinnin yhtenäisellä tavalla. Tietojen pseudonymisointi toteutetaan korvaamalla henkilötunnus satunnaisella tunnisteella ja poistamalla henkilöiden nimitiedot. Poistetut ja korvatut tiedot säilytetään hyvin suojattuna erillään muusta aineistosta.

Pseudonymisointipalvelu saatiin valmiiksi vuonna 2018 ja sitä on käytetty yksittäisten aineistojen pseudonymisointiin loppuvuodesta 2018 alkaen. Laajemmin palvelu otetaan käyttöön vastaanotto-prosessin uudistuksen käyttöönoton yhteydessä, joka tapahtuu vaiheittain vuoden 2019 alkupuolelta lähtien.

Uudistetussa vastaanotto-prosessissa kaikki Tilastokeskukseen saapuva henkilötieto pseudonymisoidaan heti vastaanoton ensimmäisessä vaiheessa riippumatta siitä, onko kyseessä viranomaisten tai yksityisten tahojen toimittamat tiedot vai Tilastokeskuksen suoraan henkilöiltä tai yrityksiltä keräämä tieto.

Käyttövaltuudet vastaanotto-prosessin ensimmäiseen vaiheeseen tietoihin, joissa on vielä henkilöiden suorat tunnistetiedot mukana, ovat hyvin rajatut ja perustuvat prosessissa tapahtuvien mahdollisten virhetilanteiden selvittämiseen. Käyttövaltuudet suorien tunnistetietojen palauttamiseen voidaan myöntää vain henkilöille, jotka työtehtäviensä suorittamiseksi tarvitsevat käyttöönsä henkilöiden suoria tunnistetietoja. Tällaisia työtehtäviä ovat esimerkiksi yhteydenotot tiedonantajiin. Kaikki pseudonymisointipalvelussa tapahtuvat tietojen käsittelyt kirjataan keskitettyyn lokijärjestelmään, josta tapahtumien kulku ja syy käsittelylle voidaan selvittää.

8.4 Verkkoharavointi uutena tiedonkeruumenetelmänä

Tilastokeskuksessa on aloitettu hanke uusien tietoaineistojen käyttöönotosta, joihin verkkoharavointi lukeutuu. Verkkoharavoinnilla tarkoitetaan automatisoitua tiedonkeruuta internetsivuilta. Verkkoharavoinnin käyttöä on Tilastokeskuksessa selvitetty ostopalveluna vuosina 2016–2017 ja sen jälkeen osana omaa kehitystyötä vuonna 2018.

Verkkoharavointi on tiedonkeruutapana myös lainsäädännöllisesti uusi menetelmä. Euroopan unionin tilastoviraston Eurostatin asettama Big Data -työryhmä kokosi helmikuussa 2017 [dokumentin](#), jossa käsitellään verkkoharavointiin liittyviä laki- ja etiikka-asioita työryhmään osallistuvissa eri maiden tilastovirastoissa. Käytänteet ja arvot eroavat maiden välillä, mutta haravointia koskevat lait ovat pääosin samat. Erityisesti verkkoharavointiin liittyviksi ylikansalliseksi lainsäädännöksi työryhmä identifioi tekijän- ja siihen liittyvien oikeuksien suojaan liittyvät direktiivit [93/98/EEC](#), [2001/29/EC](#) ja [2006/116/EC](#), tietokantojen oikeussuojaan liittyvän direktiivin [96/9/EC](#) sekä tietokoneohjelmien oikeussuojaan liittyvän direktiivin [2009/24/EC](#). Verkkoharavointia koskettaa myös kaksi kansainvälistä sopimusta: [Bernin konventio](#), joka liittyy kaunokirjallisten ja taiteellisten töiden suojaan ja [TRIPSin sopimus](#), joka liittyy immateriaalioikeuksiin. Lisäksi huomioon täytyy ottaa verkkoharavoinnin kohteiden yksityisyydensuoja.

Arviota Suomen tilanteesta ei ole toistaiseksi tehty. Tällä hetkellä Tilastokeskuksessa on periaatteena tehdä tiedonkeruusta sopimus haravoitavien sivustojen kanssa samalla tavalla kuin muiden sopimusperusteisesti tietoja antavien tahojen kanssa. Periaatteen mukaan testimateriaalia voidaan haravoida ilman sopimusta kohtuullisissa määrin ja sivustojen kuormitus huomioon ottaen. Kaikessa haravoinnissa hakurobotti ilmoittaa haravoinnin kohteena olevalle palvelimelle aineiston haravoinnista vastaavien henkilöiden yhteystiedot.

Ennen verkkoharavoinnin käyttöönottoa tilastotuotannossa asiasta tehdään selvitys, jossa arvioidaan lainsäädännöstä tulevat reunaehdot verkkoharavoinnille.

8.5 Henkilötiedonkeruiden kehittäminen vuosina 2015–2019

Tilastokeskus määritteli strategiassaan vuosille 2016–2019 tavoitteeksi verkko-vastaamisen käyttöönoton kaikissa henkilö pohjaisissa tiedonkeruissa. Käytännössä tämä tarkoittaa yhdistelmä tiedonkeruuta (mixed mode), jossa verkkovastaaminen on käytössä muiden tiedonkeruumenetelmien rinnalla. Muutoksen toteuttamiseksi perustettiin hanke, jonka alla on toteutettu lähes 20 kehittämisprojektia vuosina 2015–2019.

Yhdistelmä tiedonkeruun käyttöönotto tarkoittaa yksittäisen tiedonkeruun näkökulmasta kokonaisvaltaista tiedonkeruuprosessin uudistamista. Sillä on vaikutuksia mm. otanta-asetelmaan, tiedonkeruuaikatauluihin, saatekirjeisiin ja muuhun vastaajille suuntautuvaan viestintään, tiedonkeruulomakkeisiin, haastattelutyön organisointiin ja lopullisen tilastoaineiston muodostamiseen.

Muutokset prosessinäkökulmasta

Henkilö pohjaiset tiedonkeruut on aiemmin toteutettu kukin omana itsenäisenä ja pitkälle räätälöitynä tiedonkeruuprosessinaan. Prosessin vaiheet ja välineet ovat vaihdelleet tiedonkeruukohtaisesti ja sisältäneet runsaasti manuaalisia työvaiheita. Yksittäiseen tiedonkeruuseen liittyvä osaaminen on ollut pitkälle henkilöitynyttä.

Yhdistelmä tiedonkeruun käyttöönoton myötä tavoitellaan tiedonkeruuprosessin sujuvuuden parantumista, prosessien yhtenäistämistä ja automatisointia sekä henkilösidonnaisuuden vähentämistä. Keruiden tehokasta toteutusta varten on rakennettu uusi tiedonkeruujärjestelmä, joka sai äänestyksen jälkeen nimen Ruuti. Se valmistui tuotantokäyttöön vuoden 2018 lopussa. Kaikki henkilötiedonkeruut siirretään tiedonkeruuhjelman mukaisesti Ruuti-ympäristöön vuoden 2019 alusta lähtien.

Ruuti-järjestelmän kautta voidaan ohjata keruun toteutusta kokonaisuutena sekä verkkovastaamisen että haastattelujen osalta sekä seurata tiedonkeruiden

etenemistä ja haastattelijoiden työtä. Se mahdollistaa myös töiden entistä joustavamman jakamisen haastattelijoiden kesken, koska kaikki haastattelijat toimivat samassa ympäristössä.

Muutokset henkilötiedonantajan näkökulmasta

Tiedonantajalle muutos näkyy lisääntyneenä valinnanvarana, kun verkkovastaaminen tulee mahdolliseksi. Verkkovastaamisen suosioon eri keruissa vaikuttavat monet tekijät kuten keruun aihe, kysymysten määrä sekä vastaamisajan pituus.

Kaikille tiedonantajille verkossa vastaaminen ei ole mahdollista, joten vastauskadon hallitsemattoman kasvun välttämiseksi on tarjottava myös vaihtoehtoinen tiedonkeruutapa. Vaihtoehtoinen tapa on useimmiten puhelinhaastattelu, joissain tapauksissa myös paperilomake tai käyntihaastattelu.

Muutokset tilastotuotannon näkökulmasta

Muutoksen yhteydessä on huolehdittava siitä, että henkilöitä ja kotitalouksia koskevien tilastoaineistojen laatu säilyy ja tilastoihin kohdistuvat EU- ja muut sisältövaatimukset tulevat täytettyä. Riskinä on, että muutokset lomakkeisiin ja tiedonkeruuprosessiin aiheuttavat kadon kasvua tai tietojen vertailukelpoisuuden heikkenemistä. Riskiä voidaan pienentää mm. huolellisen lomakesuunnittelun, keruuprosessin suunnittelun ja testaamisen, keruujärjestelmän häiriöttömyyden varmistamisen sekä ammattimaisen haastattelu- ja katotyön avulla.

8.6 Koski-palvelun käyttöönottoon valmistautuminen

Koski eli Kansallisten opiskeluoikeuksien ja suoritusten keskitetty integraatiopalvelu on opetushallinnon valtakunnallinen palvelu, jonne kootaan tietoja kunkin oppijan koulutuksesta yksittäisistä opintosuorituksista suoritettuihin

tutkintoihin asti. Palveluun viedään tietoja jo olemassa olevista koulutusrekistereistä, joita Tilastokeskus käyttää jo tällä hetkellä hallinnollisina tietolähteinä, sekä uusista esi- ja perusopetuksen, lukiokoulutuksen ja ammatillisen koulutuksen opiskelija- ja tutkintorekistereistä, jotka sisältävät pitkälti vastaavaa tietoa, kuin mitä Tilastokeskus on kerännyt suorilla tiedonkeruillaan koulutuksen järjestäjiltä.

Koulutuksen järjestäjät ovat siirtäneet tietojaan esi- ja perusopetuksesta sekä toisen asteen koulutuksesta Koski-palveluun helmikuusta 2018 lähtien. Palvelu on käytössä täydessä laajuudessaan vuonna 2019.

Koski-palvelun tietojen käyttöönoton esiselvitys -projekti toteutettiin keväällä 2018 Tilastokeskuksen koulutusalueen tiedonkeruiden ja opetushallinnon valtakunnallisen palvelun tietojen vertailemiseksi, jotta tilastovuodesta 2019 voitaisiin siirtyä käyttämään Koski-palvelun tietoja ja luopua Tilastokeskuksen omista suorista tiedonkeruista. Esiselvityksessä vertailtiin kuvausten tasolla Koski-palvelun ja tiedonkeruiden tietosisällöt, jotka vastasivat suurimmalta osalta hyvin toisiaan. Ensimmäinen karkean vertailun mahdollistava testiaineisto saatiin käyttöön kuitenkin vasta vuoden lopulla. Opetushallituksen kanssa käynnistettiin syksyllä Koski-tietojen luovutussopimuksen laadinta, mutta sopimus ei valmistunut vuoden 2018 aikana eikä yksityiskohtaisen vertailun mahdollistavaa henkilötason aineistoa saatu siten käyttöön.

Vuoden aikana tehtiin suunnitelma Koski-palvelun käyttöön siirtymisen ja omista tiedonkeruista luopumisen aikataulusta vuosina 2019–2020. Lisäksi määriteltiin Tilastokeskukselle jäljelle jäävät koulutustiedonkeruut ja Koski-rekisteriä täydentämään tarvittavat tiedonkeruut.

8.7 Tulorekisterin käyttöönottoon valmistautuminen

Tulorekisteri on kansallinen sähköinen tietokanta, joka sisältää henkilöiden reaaliaikaiset palkka-, työsuhde-, eläke- ja etuustiedot. Tulorekisteri otetaan Suo-

messa käyttöön palkkatietojen osalta 1.1.2019 ja etuustietojen osalta 1.1.2020.

Tilastokeskus tulee hyödyntämään tilastoinnissaan tulorekisteriä 1.1.2020 alkaen. Tilastokeskuksen tulorekisterin käyttöönottoa tukee useampi kehittämisprojekti. Tulorekisterin käyttöönottoa koordinoidaan tiedonhankinta-yksikössä, jossa on myös tulorekisterin testausta suorittavat projektit. Tulorekisteritiedot tullaan Tilastokeskuksessa viemään uudistuvaan henkilötietovarantoon, jota rakennetaan väestö- ja elinolot -yksikön kehittämishankkeessa 2017–2020. Tulorekisteri mahdollistaa myös yritys- ja suhdannetilastoinnin uudistuksia.

Tulorekisterin toivotaan tulevaisuudessa keventävän sekä mahdollisuuksien mukaan osin lopettavan Tilastokeskuksen nykyisiä omia ammatti-, työsuhte- ja palkkatiedonkeruita. Koska tulorekisteri on reaaliaikainen, se mahdollistaa myös tilastoinnissa nykyistä nopeamman tilastotiedon tuottamisen ja jakamisen.

Tulorekisterin aineistoa ja prosessia testattiin Tilastokeskuksessa vuonna 2018. Testausta jatketaan vielä vuonna 2019.

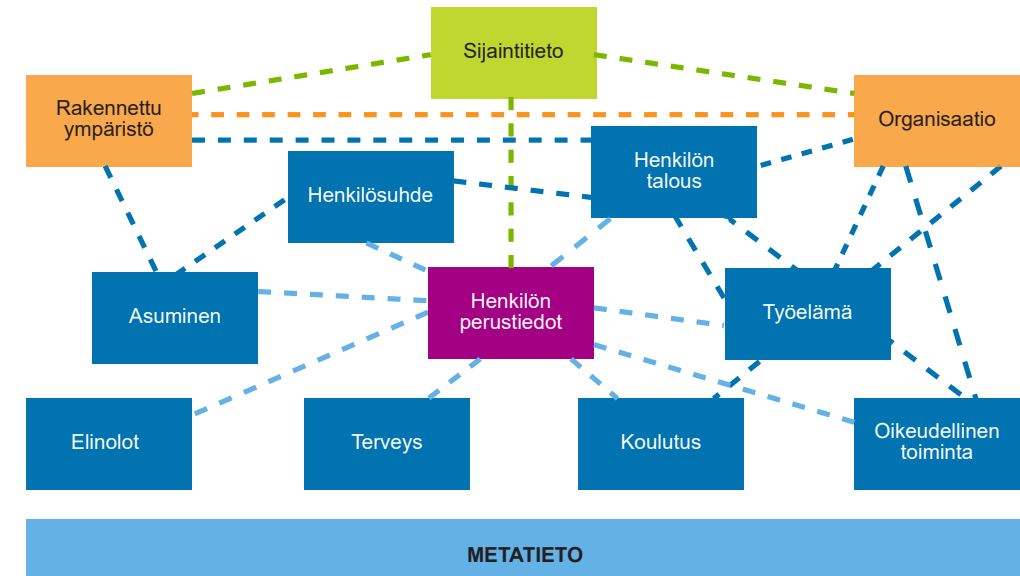
9 Henkilötietovarannon uudistamishanke (STIINA)

Tilastokeskus on uudistamassa henkilötietovarantoa henkilötieto-kokonaisuuden uudistushankkeessa. Hankkeessa toteutetaan uusi henkilötietovaranto ja siihen liittyvät tilastotuotantoprosessit osaprojekteina vuosien 2016–2020 aikana. Hankkeen tavoitteena on rakentaa palvelupohjaisen tietoarkkitehtuuriin mukaisesti yhteiskäyttöiset, jatkuvapäivitteiset, metatieto-ohjautuvat tuotantojärjestelmät.

Hankkeen taustalla on henkilökokonaisuuteen kuuluvien tilastojen tietojärjestelmien uudistustarpeen lisäksi tiedossa olevat suuret muutokset toimintaympäristössä. Näistä merkittävimpiä ovat tulorekisterin ja opetushallinnon Koski-palvelun käyttöönotto sekä mahdollinen sote-uudistus ja sen tuomat

muutokset kuntien toimintaan. Uudistuksessa huomioidaan tärkeänä osana henkilötietojen käytön tietosuojaa. Hankkeen myötä otetaan käyttöön henkilön suorien tunnistetietojen pseudonymisointi kaikissa henkilö pohjaisissa tuotantoprosessissa. Kun uusi tietojärjestelmä saadaan käyttöön, ei tilastoaineistoissa käytetä enää henkilön suoria tunnistetietoja keruuvaiheen jälkeen.

Kuvio 4. Henkilötietokokonaisuuden loogiset tietovarannot



10 Yritys- ja toimipaikkarekisteri

Tilastokeskus ylläpitää tilastolain perusteella tietopalvelutehtäviä varten yritys- ja toimipaikkarekisteriä. Se sisältää kaikki yritykset, yhteisöt ja yksityiset elinkeinonharjoittajat, jotka ovat liiketoiminnasta arvonlisäverovelvollisia tai toimivat työnantajina. Yritysrekisterin tietoja ovat mm. yritysten osoitteet, toimialat, henkilöstön ja liikevaihdon suuruusluokat.

Yritys- ja toimipaikkarekisterissä oli vuoden 2018 lopussa noin **761 000** (2017: 735 000) luonnollista henkilöä, joista edelleen toimintaa harjoittaa noin **279 000** (2017: 287 000) luonnollista henkilöä.

Yritysrekisteristä tehdyistä yksikkötason tietotoimituksista suurin osa tehdään aktiivisten toimivien yritysten joukosta. Näistä yrityksistä noin 39 prosenttia on oikeudelliselta muodoltaan luonnollisia henkilöitä. Kun yritysrekisteristä poimitaan tietopalvelun tilaajalle tietoja, luonnolliset henkilöt kuuluvat osana yritysten poimintakehikkoon.

11 Avoin data

11.1 Tilastokeskuksen tuottama avoin data

Tilastokeskuksen avoimen datan avaukset ovat ajoittuneet pääasiassa vuosille 2012–2015. Toimenpiteet toteutettiin tällöin valtion talousarvioehdotuksessa Tilastokeskukselle julkisten tietovarantojen avaamiseksi myönnettyjen määrärahojen puitteissa. Vuonna 2018 Tilastokeskus panosti lähinnä datan löydettävyyteen ja käytön ohjeistukseen. Kaikki avatut aineistot (www.stat.fi/avoindata) ovat nykyisin käytettävissä CC BY 4.0 –ehdoin.

Tietovarantojen avaaminen on keskeinen julkishallinnon kansallinen tavoite. Tavoite mainitaan Valtioneuvoston periaatepäätöksessä datan hyödyntämistä liiketoiminnassa (19.5.2016), hallitusohjelmien kirjauksissa, avoimen hallinnon hankkeissa (Open Government Partnership) sekä ministerityöryhmän vuonna 2016 hyväksymissä [digitalisoinnin periaatteissa](#) (2.2.2016). Tiedon avointa saatavuutta käsitellään VNK:n raportissa [Avoimen datan hyödyntäminen ja vaikuttavuus](#) (16.3.2017), VM:n muistiossa [Suomi tarvitsee tietopolitiikkaa](#) (8.8.2017) sekä VM:n tietopoliittisessa selonteossa [Eettistä tietopolitiikkaa tekoälyn aikakaudella](#) (20.9.2018). Valtioneuvoston alkuvuonna 2018 ilmestynyt [Tiedolla johtaminen aluekehittämisessä](#) –raportti painottaa lisäksi, että alueellisen tiedon ja tiedon selkeän esittämisen tarve nousee sekä kansallisen että alueellisen tason päätöksenteon ja politiikkavalmistelun tarpeista.

Tilastokeskuksessa aloitettiin vuonna 2018 asiakkaiden alueellisen tiedon tarpeiden kartoitus maakunta- ja SOTE-uudistuksen näkökulmasta. Kartoitus oli osa laajempaa hankekokonaisuutta, jossa tavoitteena on tulevaisuudessa lisätä alueellisen tiedon tarjontaa avoimena datana. Esiselvityksen pohjalta voidaan todeta, että vuosina 2018–2020 Tilastokeskuksessa on käynnissä useita hankkeita, joiden tavoitteiden toteutuminen edistää alueellisen tiedon tietohuollon tarpeisiin vastaamista niin sisällöllisesti (tiedot ja metatiedot) kuin teknisestikin. Aluetiedon laadukas jakelu edellyttää kuitenkin Tilastokeskuksen tietosuojalinjausten tarkentamista sekä tiedon suojauksessa käytettävien teknisten ratkaisujen edelleen kehittämistä.

11.2 Tilastokeskuksen hyödyntämä avoin data

Tilastokeskus hyödyntää avointa dataa tilastotuotannossaan. Vuonna 2018 Tilastokeskus hyödynsi Liikenneviraston tuottamaa avointa dataa (Digiroad) tieliikenneonnettomuustilaston onnettomuuspaikkatietojen korjaamiseen ja täydentämiseen. Lisäksi Ilmatieteen laitoksen avoimesta rajapinnasta haetaan onnettomuuspaikkaa ja onnettomuusaikaa lähimmän sääaseman tallentama tieto lämpötilasta. Datasta imputoidaan virheellisiä ja puuttuvia tieliikenneonnettomuustilaston lämpötilatietoja.

Muulla avointa dataa hyödynnetään niin toimeksiantoissa ja kokeiluissa kuin tuotannossakin.

12 Toimeksiantoina tehtävä henkilötietojen käsittely

Toimeksiantoina asiakkaille tehtävien tiedonkeruiden aiheet ja kokoluokka vaihtelevat paljon. Toimeksiantoina tehdään ainoastaan yhteiskunnallisesti tai tieteellisesti merkittäviä tutkimuksia. Aiheesta riippuen toteutustavat vaihtelevat.

vat esimerkiksi kymmenille tuhansille vastaajille lähetettävistä verkkokyselyistä 3 000 henkilön käyntihaastattelututkimuksiin ja 5 000 henkilön verkko-/puhelinkeruisiin. Osa keruista on säännöllisesti toistuvia, osa kertaluonteisia.

Tiedonkeruita tilaavat pääasiassa julkishallinto ja tutkimuslaitokset –tyypillisimmin ministeriöt, yliopistot sekä tutkimusorganisaatiot kuten esimerkiksi THL ja Väestöliitto. Tällöin Tilastokeskus toimii pääsääntöisesti henkilötietojen käsittelijänä ja toimeksiantaja rekisterinpitäjänä. Aineiston tietosisällöstä, tiedonkeruusopimuksesta ja otoshenkilöiden informoinnista riippuen anonymisoitu aineisto voidaan toimittaa toimeksiantajan lisäksi Yhteiskuntatieteelliseen tietoarkistoon (FSD).

Toimeksiantoina tehtäviä tutkimuksia koskevat samat salassapitovelvoitteet kuin Tilastokeskuksen omiakin tiedonkeruita. Henkilötietoja käsitellään aina luottamuksellisesti, eikä Tilastokeskus missään olosuhteissa luovuta eteenpäin henkilön tunnistamisen mahdollistavia tietoja.

Tutkimusaineiston käyttöön ja henkilötietojen säilyttämiseen Tilastokeskuksessa vaikuttaa EU:n yleisen tietosuoja-asetuksen lisäksi asiakkaan kanssa tehtävä tiedonkeruusopimus. Aineiston keruusta, säilyttämisestä ja jatkokäytöstä sekä henkilötietojen käsittelystä ja hävittämisestä kerrotaan saateviestissä ja muissa materiaaleissa, jotka toimitetaan tiedonkeruuseen poimituille otoshenkilöille ennen aineistonkeruun aloittamista.

Tiedonkeruun ja aineiston koostamisen jälkeen tutkimusaineisto voidaan luovuttaa toimeksiantajalle joko tiedostona tai avata käytettäväksi etäyhteydellä. Kummassakaan tapauksessa toimeksiantaja ei pääse näkemään sellaisia tietoja, joiden perusteella vastaajien tunnistaminen olisi mahdollista. Ennen aineiston toimittamista/avaamista käyttäjille Tilastokeskuksen tutkijat poistavat kaikki yksilöivät henkilötiedot (kuten nimet) ja karkeistavat sellaiset tiedot, joita yhdistelemällä yksittäisen vastaajan tunnistaminen olisi mahdollista. Tällaisia tietoja ovat esimerkiksi asuinpaikkakunta, tutkinto ja työpaikka, jotka karkeistetaan esimerkiksi kunnasta maakuntaan, tutkinnosta tutkintoalaan ja työpaikasta työpaikan toimialaan tai ammattiryhmään.

Valmiit tiedonkeruut arkistoidaan niin, että kaikesta keruun aikana kertyneestä materiaalista säilytetään vain lopullinen saatemateriaali, lomake ja anonymisoitu aineisto selosteineen. Joissain tapauksissa säilytetään lisäksi henkilöt ja vastaukset yhdistävä avaintiedosto, jolla aineiston anonymisointi voidaan tarvittaessa purkaa.

Pääsääntöisesti henkilötiedot hävitetään lopullisesti parin kuukauden sisällä aineiston valmistumisesta. Poikkeustapauksissa henkilötietoja voidaan säilyttää myös kauemmin, esimerkiksi mikäli odotetaan aineistoon myöhemmin liitettävien rekisteritietojen valmistumista. Toisen poikkeuksen henkilötietojen pitkän hävittämisen käytäntöön muodostavat toistuvat tutkimukset (mm. seuranta- ja paneelitutkimukset), joiden kohdalla voidaan toimeksiantajan kanssa sopia henkilötietojen, aineiston tai materiaalin pitkäaikaisemmasta säilyttämisestä esimerkiksi saman tiedonkeruun seuraavaa kierrosta varten. Myös tästä kerrotaan vastaajille saatemateriaaleissa ennen aineistonkeruun alkamista.

13 Asiakas- ja sidosryhmätiedot

Vuonna 2018 varmistettiin myös järjestelmien ja käytäntöjen osalta, että Tilastokeskuksen asiakas- ja sidosryhmätietojen käsittely sekä viestintä ja markkinointi toteutetaan EU:n tietosuoja-asetuksen vaatimusten mukaisesti. Asiakas- ja sidosryhmärekisteriseloste sekä evästeinfo päivitettiin. Asiakas- ja sidosryhmätiedon hallintajärjestelmää kehitettiin vastaamaan tietosuoja-asetusta ja tietosisällön säilytysajat määriteltiin ja dokumentoitiin. Yksiköiden omia asiakaslistoja ja -luetteloita siirrettiin CRM-järjestelmään.

14 Asianhallinta ja arkistointi

14.1 Asianhallinta ja analoginen asiakirja-aineisto

Tilastokeskuksen hallinnollisten asioiden arkistonmuodostussuunnitelma on päivitetty ja vahvistettu yleis-, talous- ja henkilöstöhallinnon osalta vuonna 2012. Kansallisarkistolle hyväksyttäväksi lähetettyyn muut kuin em. toiminnot sisältävään säilytysaikaesitykseen saatiin päätösluonnos joulukuussa 2018.

Tilastokeskuksessa lokakuussa 2017 käynnistynyt asian- ja dokumentinhallinnan kehittämishanke (ADOHA) jatkui. Hankkeen tavoitteena on luoda asian- ja dokumentinhallinnalle selkeät prosessit ja toimintatavat sekä ottaa käyttöön helppokäyttöiset asian- ja dokumentinhallinnan sähköiset järjestelmät. Asianhallinnan osalta hankkeen alaprojekti, jossa luonnosteltiin tiedonhallintasuunnitelma, päättyi vuoden 2018 lopussa. Uuden sähköisen asianhallintajärjestelmän hankinta- ja käyttöönottoprojektit alkavat vuosina 2019–2020.

Kirjaamo rekisteröi ja jakeli vuoden 2018 aikana 1 883 uutta asiaa sekä arkistoi manuaalisesti ja skannasi Dynasty-diaarijärjestelmään noin 7 000 asiakirjaa.

Valtioneuvoston periaatepäätöksen 21.6.2017 mukaisesti viranomaisten hallussa oleva analoginen asiakirja-aineisto digitoidaan vuoteen 2030 mennessä. Suurin osa analogisesta asiakirja-aineistosta hävitetään digitoinnin jälkeen. Vuonna 2018 Tilastokeskus osallistui Kansallisarkiston koordinoimaan massadigitoinnin Proof of Concept (PoC)-hankkeeseen. Lisäksi jatkettiin aineistojen inventointia ja hävitettiin niitä määräajan säilytettäviä asiakirja-aineistoja, joiden säilytysaika on päättynyt. Manuaaliaineistoa oli vuoden 2018 lopussa noin 2 500 hyllymetriä, josta pysyvästi säilytettävä noin 1 300 hyllymetriä. Suuri osa siitä on käyttörajoitettua tai henkilötietoja sisältävää tietoaineistoa.

Tilastokeskuksen toimitilojen muutoshanke toteutettiin vuosina 2016–2018. Virasto siirtyi perinteisestä huonetoimistosta valtion toimitilastrategian mukaiseen monitilaympäristöön kevään 2018 aikana. Uusien ratkaisujen pohjana

olivat rakentamisen säädökset ja laatuvaatimukset, joiden kautta täyttyvät esimerkiksi materiaalien laadun, akustiikan ja talotekniikan turvallisuuden ja toimivuuden kriteerit. Turvallisuus ja sen myötä myös tietosuoja paranivat, koska sähköinen kulunvalvonta asennettiin kulkuovien lisäksi kellarin arkistotiloihin, hisseihin ja kiinteistön yhteisiin neuvottelu- ja kokoustiloihin. Pääsy käyttörajoitettua tietoaineistoa sisältäviin tiloihin on nyt rajattu.

14.2 Tilastoaineiston arkistointi

Arkistoinnin toteuma on vuonna 2018 ensimmäistä kertaa mittarina Tilastokeskuksessa tulosityksiköiden tuloskorteissa. Arkistointisuunnitelmat ja niiden toteumat ovat seurattavissa ajantasaisesti raportointijärjestelmässä Portissa. Vuoden 2018 aikana tulosityksiköt arkistoivat yhteensä 183 tilastoaineistoa.

Tilastokeskus sai kesällä 2018 Kansallisarkistolta säilytysaikapäätökset pysyvästi säilytettävistä digitaalisessa muodossa arkistoitavista tilastoaineistoista. Tämän jälkeen pyritään edistämään määräajan säilytettävien eri tuotantovaiheiden tilastoaineistojen säilytysaikojen yhtenäistä määrittystä.

15 Mittarit

15.1 Tilastot

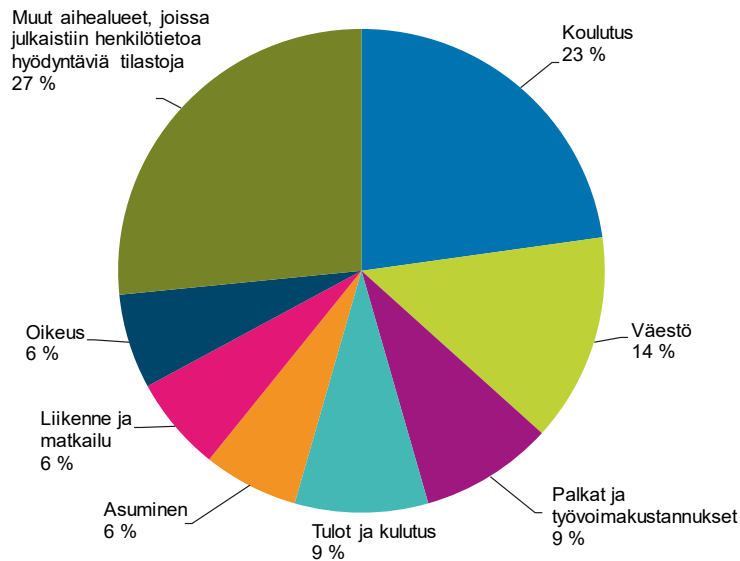
Vuonna 2018 Tilastokeskuksella oli tuotannossa 156 tilastoa. Niistä kaksi, elin-olotilasto ja maatalouden taloustilit (EAA), perustettiin tilastovuonna 2018. Omistusasumisen hintaindeksit -tilasto lakkautettiin, tilasto ilmestyi viimeisen kerran 4.4.2018. Jatkossa tiedot ovat saatavilla Eurostatin sivuston kautta.

Vuonna 2018 Tilastokeskuksen kaikista tilastoista 145 julkaisi tietoja. Kaikkia tilastoja ei julkaista vuosittain. Esimerkiksi innovaatiotoiminta-tilasto tuotetaan

joka toinen vuosi ja ajankäyttötilaston pohjana oleva ajankäyttötutkimus tehdään vain noin kymmenen vuoden välein. Vaalitulastoja tuotetaan vaalivuosina.

Vuonna 2018 julkaistuista tilastoista 79:ssä käytettiin henkilötietoja. Suurin osa henkilöitä kuvaavista tilastoista tuotetaan mikrotietovarannosta eli prosessissa käsitellään henkilötietoja. Jotkut tilastot, esimerkiksi kotitaloustuotannon arvo-tilasto, tuotetaan makrotietovarannosta eli aggregoidusta tilastoaineistosta

Kuvio 5. Henkilötietoja vuonna 2018 hyödyntävien tilastojen sijoittuminen aihealueille. (N=79)



Suurin osa vuonna 2018 henkilötietoon pohjautuvista tilastoista sijoittui koulutuksen ja väestön aihealueille.

Lisätietoa aihealuejakoista sivulta: <http://www.stat.fi/til/aihealuejako.html>

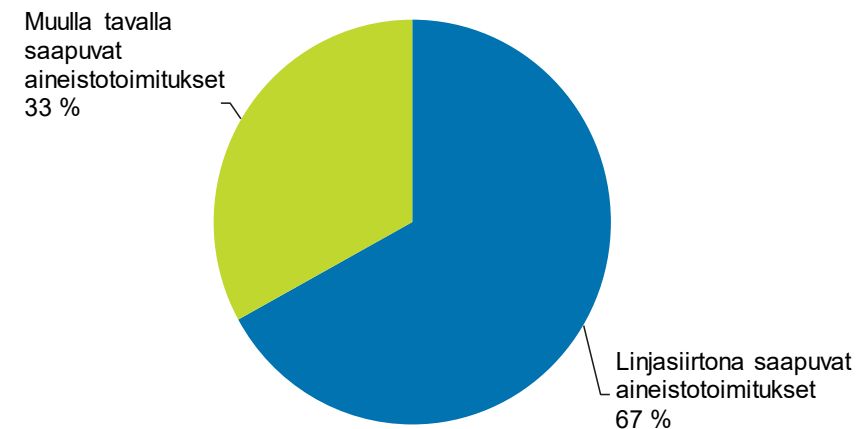
15.2 Tilastokeskuksen tilastotuotannossaan käyttämien aineistojen vastaanotto

Tilastokeskuksessa toimii keskitetty aineistojen vastaanottopalvelu (AVa), jonka kautta kulkevien tiedostojen vuosivolyymi oli 54 000 tiedostoa vuonna 2018.

Edellä mainittujen aineistojen lisäksi Tilastokeskukseen saapuu keskitetyn vastaanoton ohi jonkin verran lähinnä taulukkomuotoisia aineistoja.

15.3 Aineistojen vastaanoton tavat vuonna 2018

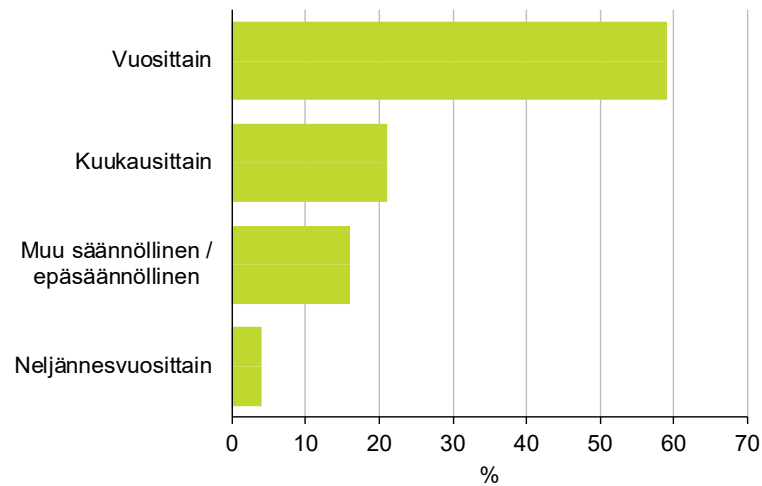
Kuvio 6. (N=220)



Keskitetyn aineistojen vastaanoton kautta tietoja toimittaa 42 eri tiedontuottajaa. Tiedon siirtää Tilastokeskukseen joko sen tuottaja tai jokin heidän valtuuttamansa taho (esim. Tieto Oy, CGI).

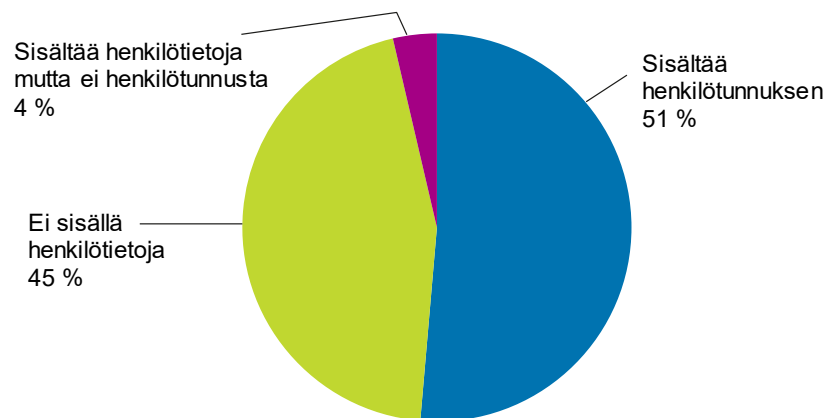
15.4 Saapuvien aineistojen frekvenssit 2018

Kuvio 7. (N=220)



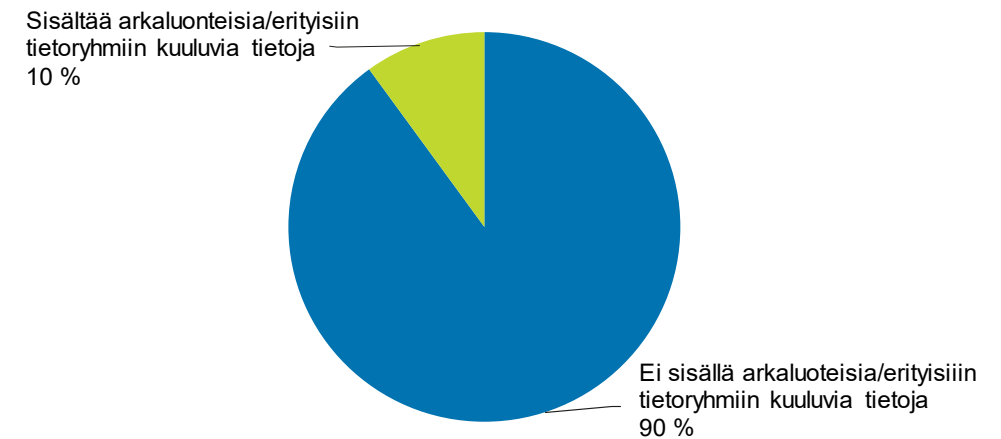
15.5 Aineistotoimitukset Tilastokeskukseen tilastotuotannon käyttöön vuonna 2018

Kuvio 8. (N=220)



15.6 Arkaluonteisia*/erityisiä tietoryhmiä** kuvaavia tietoja

Kuvio 9. Arkaluonteisia*/erityisiä tietoryhmiä** kuvaavia tietoja sisältävien aineistojen osuus tilastotuotantoon saaduista aineistoista vuonna 2018. Kuvio perustuu tiedonkeruurekisteriin kuvattujen aineistokuvausten tietoihin. (N=220)

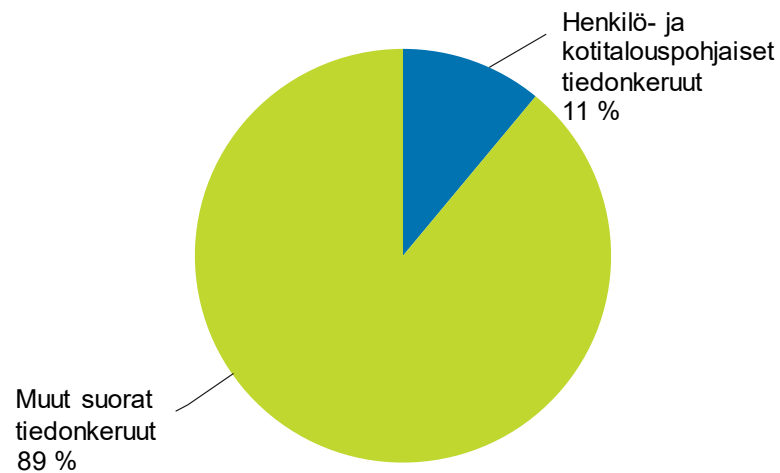


* Arkaluonteisia tietoja ovat henkilötietolain 11 § määritellyt tiedot (rotu tai etninen alkuperä; henkilön yhteiskunnallinen, poliittinen tai uskonnollinen vakaumus tai ammattiliittoon kuuluminen; rikollinen teko, rangaistus tai muu rikoksen seuraamus; henkilön terveydentila, sairaus tai vammaisuus taikka hänen kohdistetut hoitotoimenpiteet tai niihin verrattavat toimet; henkilön seksuaalinen suuntautuminen tai käyttäytyminen; taikka henkilön sosiaalihuollon tarve tai hänen saamansa sosiaalihuollon palvelut, tukitoimet ja muut sosiaalihuollon etuudet)

**Erityisiin tietoryhmiin kuuluvat tiedot ovat EU:n yleisen tietosuoja-asetuksen 9 artiklassa määritellyt (rotu tai etninen alkuperä, poliittinen mielipide, uskonnollinen tai filosofinen vakaumus tai ammattiliiton jäsenyys sekä geneettiset tai biometriset tiedot, terveyttä koskevat tiedot taikka luonnollisen henkilön seksuaalista käyttäytymistä ja suuntautumista koskevat tiedot) ja EU:n yleisen tietosuoja-asetuksen 10 artiklassa määritellyt rikostuomioihin ja rikkomuksiin liittyvät tiedot.

15.7 Suorat tiedonkeruut

Kuvio 10. Henkilö- ja kotitalouspohjaiset suorat tiedonkeruut kaikista vuoden 2018 suorista tiedonkeruista. (N=80)



Vuoden 2018 keruiden tarkastelusta puuttuvat aiemmin tehdyt kertaluonteiset henkilö- ja kotitalouspohjaiset suorat keruut: Asuntojen vuokrat -verkko-/puhelintiedonkeruu, ICT-tutkimuksen verkkopilotti, Itsensä työllistäjät 2013, Palkansaajien kokema työsyryntä Suomessa sekä Ulkomaista syntyperää olevien työ ja hyvinvointi -tutkimus (UTH).

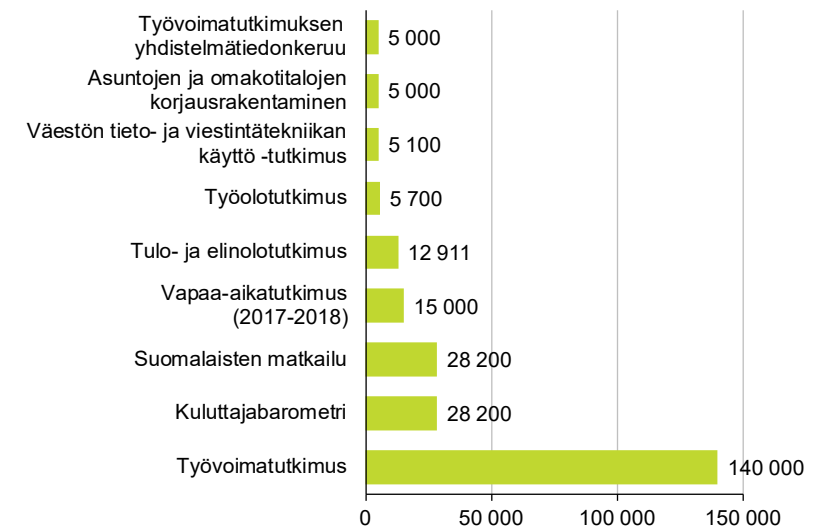
Vuonna 2018 ei toteutettu aikuiskoulutustutkimusta, ajankäyttötutkimusta eikä kulutustutkimusta.

15.9 Vuonna 2018 toteutetut suorat henkilöpohjaiset tiedonkeruut

- Asuntojen ja omakotitalojen korjausrakentaminen
- Kuluttajabarometri
- Suomalaisten matkailu
- Tulo- ja elinolotutkimus
- Työolotutkimus
- Työvoimatutkimuksen yhdistelmätiedonkeruu
- Työvoimatutkimus
- Väestön tieto- ja viestintätekniikan käyttö -tutkimus
- Vapaa-aikatutkimus (päätyi tammikuussa 2018)

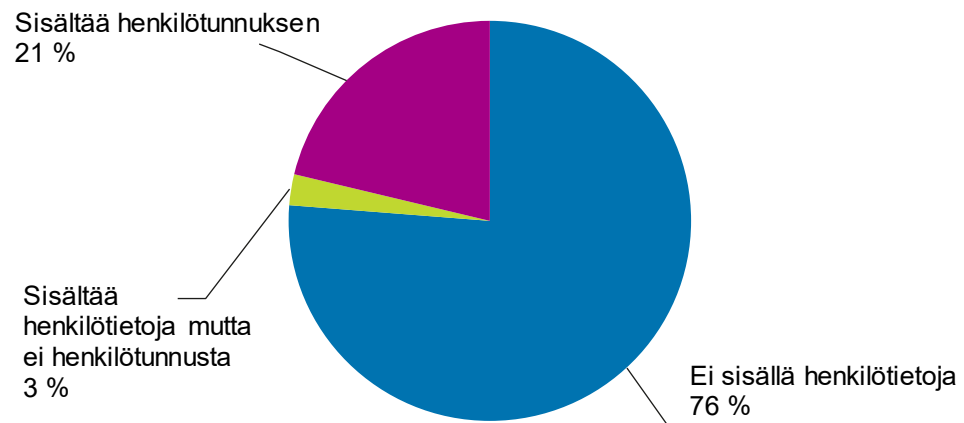
15.10 Henkilö- ja kotitalouspohjaiset suorat tiedonkeruut vuonna 2018 otoskoon mukaan

Kuvio 11. Vuoden 2018 aikana tehtyjen henkilö- ja kotitalouspohjaisten suorien tiedonkeruiden **otoksissa oli yhteensä 216 911 kohdetta.** (N=9)



15.11 Henkilötieto ja suoralla tiedonkeruulla kerättävät tiedot vuonna 2018

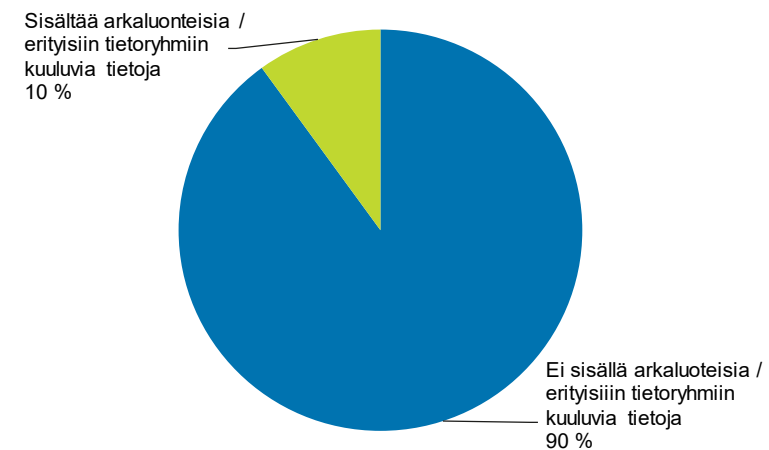
Kuvio 12. (N=80)



Henkilötiedoksi luettavaa tietoa kerätään myös yrityskeruissa, kuten esimerkiksi palkkatiedusteluissa. Tilastokeskus tarvitsee tiedonkeruussa henkilötunnuksen, jotta työnantajan toimittamia tietoja voidaan yhdistää muihin aineistoihin – esimerkiksi tutkintorekisterin tutkintotietoihin.

15.12 Arkaluonteinen/erityisiin tietoryhmiin kuuluva tieto suorassa tiedonkeruussa vuonna 2018

Kuvio 13. (N=80)



15.12.1 Haastattelujen määrät vuosina 2017–2018

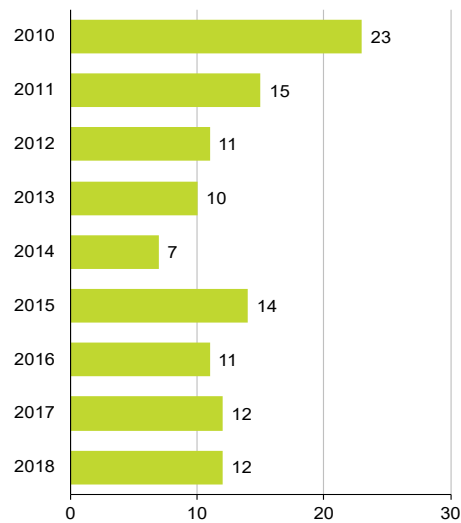
Taulukko 1. Tilastokeskus on suoran tiedonkeruun yhteydessä tehtävien haastattelujen osalta rekisterinpitäjänä. (N=80)

	2017	2017 %	2018	2018 %
Käyntihaastattelu	6 412	4	3 905	3
Puhelinhaastattelu	144 919	96	122 408	97
Yhteensä	151 331	100	126 313	100

15.13 Tilaustutkimukset

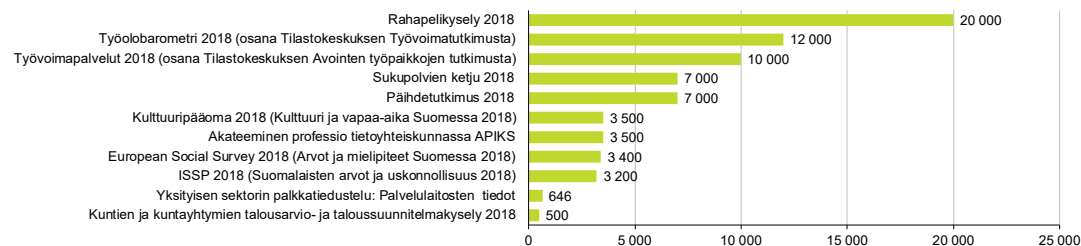
15.13.1 Toimeksiantojen lukumäärä vuosittain: asiakasrahoitteiset tilaustutkimukset vuosina 2010–2018

Kuvio 14. Toimeksiantojen lukumäärä 2010 - 2018



15.13.2 Tilaustutkimukset 2018 otoskoon mukaan

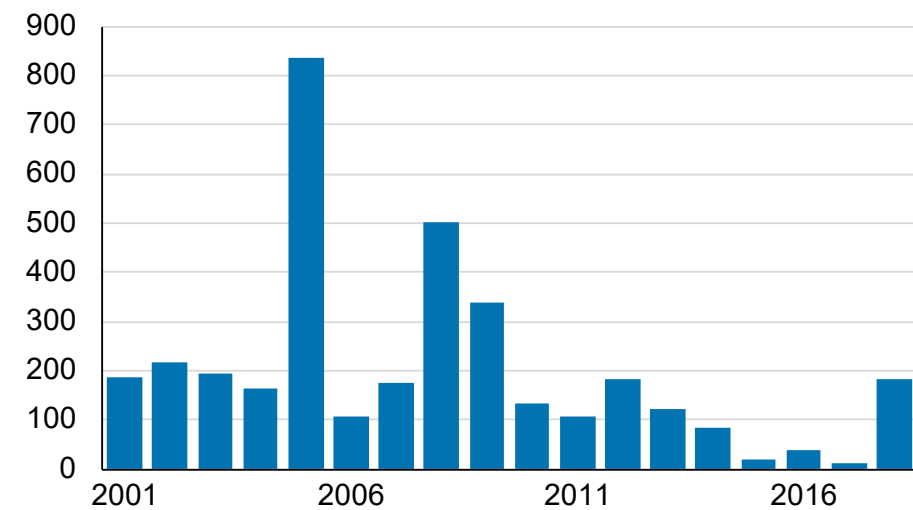
Kuvio 15. Tilastokeskuksen tekemien asiakasrahoitteisten tilaustutkimusten otokset olivat vuonna 2018 yhteensä 70 746 yksikköä.



15.14 Arkistointi

15.14.1 Arkistoitujen tilastoaineistojen lukumäärät vuosina 2001–2018

Kuvio 16.



Tilastoaineistojen arkistointi on osa tilastotuotantoprosessia. Arkistoinnin tarkoituksena on varmistaa tutkimus- ja tilastointitarkoituksen kannalta arvokkaan tiedon säilyminen käyttökelpoisessa muodossa. Aineistojen hyödyntämisen kannalta on tärkeää, että arkistoitavat tilastoaineistot kuvataan laadukkaasti.

15.15 Tutkimusaineistot

15.15.1 Tutkimusaineistojen käyttöluvat ja luovuttaminen

Tilastolain mukaan Tilastokeskuksella on velvollisuus edistää tilastotarkoituksia varten kerättyjen tietojen käyttöä tieteellisissä tutkimuksissa ja yhteiskuntaoloja koskevissa tilastollisissa selvityksissä.

Tilastokeskuksessa tutkimusaineistojen pohjana ovat Tilastokeskuksen tilastointia varten saadut aineistot. Aineistoja sijaitsee tilastojen tuotantojärjestelmissä ja tilastointia varten perustetuissa tietokannoissa. Aineistojen yhdistelyn avaimena voivat toimia henkilönumero, henkilötunnus, Y-tunnus, rakennustunnus, postinumero tai vastaava.

Taulukko 2. Tutkimusaineistojen käyttöluvat ja luovuttaminen.

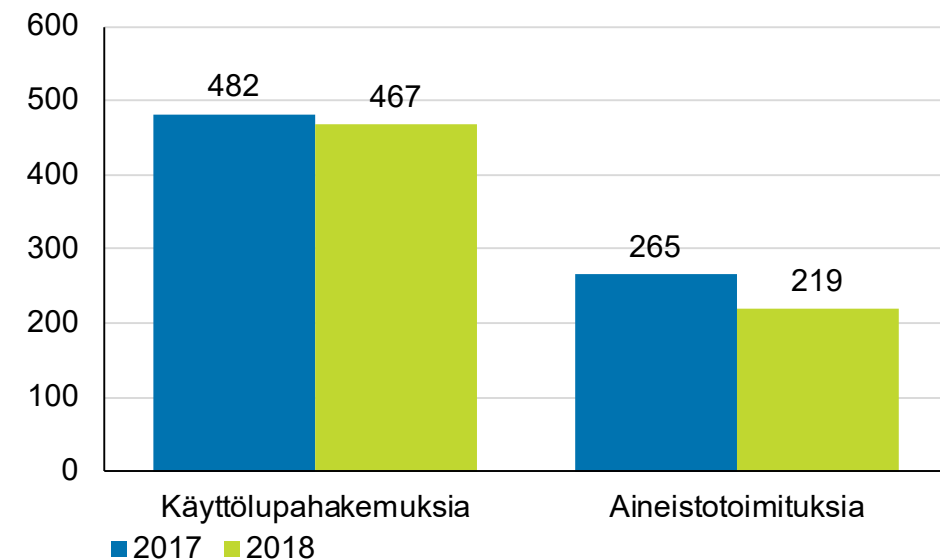
Tuoteluokka myönnetyt	käyttöluvat		aineistotoimitus	
	2017	2018	2017	2018
Kuolemansyyaineistot - laki kuolemansyy selvittämisestä 15 §	80	70	85	64
Voimassaolevien käyttöluoppien jatkot ja laajennukset	144	135		
Mikrosimulointi	21	12	14	7
Valmisaineistojen etäkäyttö	49	62	41	49
Räätälöityjen tutkimusaineistojen etäkäyttö	90	74	80	56
Räätälöityjen tutkimusaineistojen luovutus*	29	26	18	20
Tunnisteellisenä luovutettavat tutkimusaineistot** - tilastolaki 19 § (ammatti, sosioekonominen asema, koulutus, ikä, sukupuoli)	26	23	21	22
Muut (ainoastaan toisten viranomaisten aineistot)	5	4	6	1
Yhteensä	444	406	265	220

*Anonymisoituja aineistoja, joissa tilastoyksiköitä ei voi tunnistaa välillisesti.

**Tilastolain 19 § mukaan käyttöluvalla voidaan luovuttaa tunnisteteollisessa muodossa ammattia-, ikää, sukupuolta ja sosioekonomista asemaa koskevia sekä koulutustietoja sisältäviä tutkimusaineistoja.

15.15.2 Tutkimusaineistojen käyttöluvahakemukset ja tutkimusaineistotoimitukset 2017 ja 2018

Kuvio 17. Tutkimusaineistojen käyttöluvahakemukset (sis. jatkoluvat) sekä tutkimusaineistotoimitukset 2017 ja 2018.



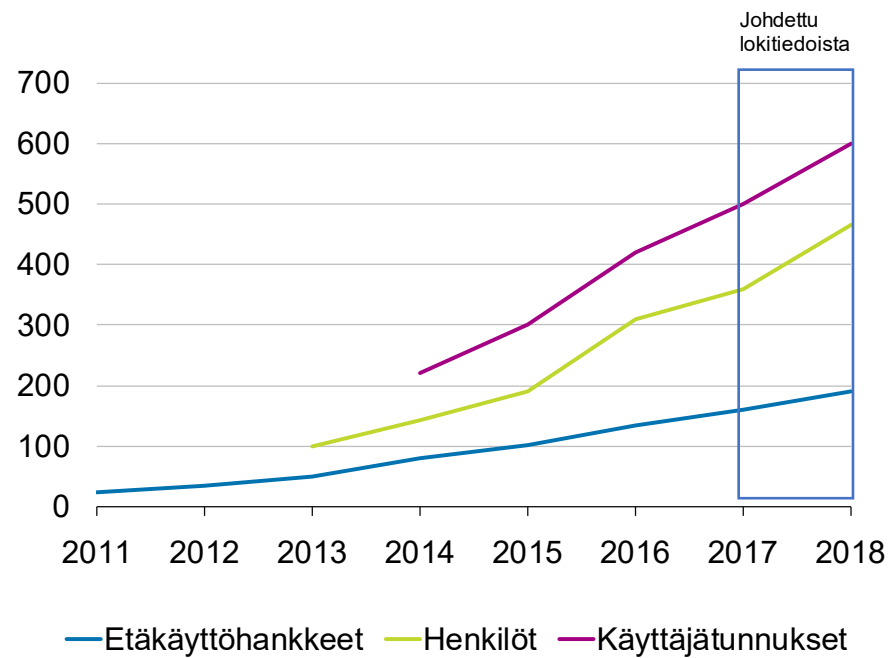
15.16 Kansallinen etäkäyttöpalvelu

Taulukko 3. Kansallinen etäkäyttöpalvelu 2017 ja 2018.

	2017	2018
Etäkäytön organisaatiosopimukset	71	88
Etäkäyttöhankkeet	163	196
Henkilöitä (mahdollisesti mukana useassa hankkeessa)	363	466
Käyttäjätunnuksia (hanke ja henkilökohtaisia tunnuksia)	501	601

15.16.1 Tutkimusaineistojen etäkäytön mittarit 2010-luvulla

Kuvio 18. Etäkäytön mittarit.



Tilastokeskus 

www.tilastokeskus.fi
www.facebook.com/Tilastotohtori
<https://twitter.com/tilastokeskus>