

Tietosuoja tilasto- ja paikkatietojen integroinnissa

Title: Tietosuoja tilasto- ja paikkatietojen integroinnissa

Title in English: Data confidentiality in the integration of statistical and geospatial information

Project: The GSGF in Finland – Integration of geospatial and statistical information in Finland (GSFI)

Grant agreement number: 101112903 - 2022-FI-GEOS-GSFI

It is permitted to copy and reproduce the content in this report. When quoting, please state the source.

© GSFI and Eurostat 2024

Sisältö

1. Yhteenveto.....	3
2. Summary.....	3
3. Tietosuojan merkitys paikkatietoja julkaistaessa.....	3
4. Paikkatiedon vaikutus paljastumisriskiin.....	4
4.1. Paikkatiedon tarkkuus.....	4
4.2. Havaintojen muodostamat rypäät ja autokorrelaatio	5
4.3. Leikkaavat tilastoalueet.....	5
5. Tilastolliset tietosuojamenetelmät	5
6. Kansainväliset suositukset tietosuojasta tilasto- ja paikkatietoja yhdistettäessä	6
7. Tilasto- ja paikkatietojen yhdistäminen Suomessa ja Tilastokeskuksessa	7
8. Lähteet.....	9

1. Yhteenveto

Tämä dokumentti käsittelee tietosuojaa tilasto- ja paikkatietojen integroinnissa keskittyen GEOSTAT 4 -projektin GSGF Europe kokonaisuuden tuotoksiin ja kansainvälisiin julkaisuihin sekä tuoden esiin kansallisia näkökulmia.

Tietosuojan merkitys korostuu, kun tilastotietoja yhdistetään tarkkoihin sijaintitietoihin, sillä yksittäisten tilastoyksiköiden tunnistaminen voi olla mahdollista. Tilastoyksikön maantieteellinen sijainti voi toimia epäsuorana tunnisteena, mikä lisää paljastumisriskiä. Tämä riski liittyy erityisesti paikkatietojen tarkkuuteen, havaintojen muodostamiin rypäisiin ja autokorrelaatioon sekä leikkaaviin tilastoalueisiin.

Paljastumisriskin pienentämiseksi käytetään tilastollisia tietosuojamenetelmiä, joita ovat mm. record swapping, kohinan lisääminen, peittäminen ja aluetietojen aggregointi. Kansainväliset suositukset (esim. UN EG-ISGI) korostavat tietoisuuden lisäämistä paikkatietojen erityispiirteistä ja niiden huomioimista julkaisu- ja suojauskäytännöissä.

Suomessa Tilastokeskus ja muut viranomaiset yhdistävät tilastotietoja paikkatietoihin tarjotakseen erilaisia palveluita, kuten Ruututietokanta ja YKR-aineisto.

2. Summary

This document addresses data confidentiality in the integration of statistical and geospatial information, focusing on the outputs of the GEOSTAT 4 project deliverable “GSGF Europe” and international publications, while also highlighting national perspectives.

The importance of data confidentiality is emphasized when statistical data is integrated with location data, as it may be possible to identify individual statistical units. The geographical location of a statistical unit can act as an indirect identifier, increasing the unit’s disclosure risk. This risk is particularly related to the accuracy of geospatial data, the clustering of observations and spatial autocorrelation, and intersecting statistical areas.

To lower the disclosure risk, statistical disclosure control methods are used, such as record swapping, noise addition, suppression, and the aggregation of geographical information. International recommendations (e.g., UN EG-ISGI) emphasize raising awareness of the specific characteristics of geospatial data and considering these characteristics in publication policies and protection practices.

In Finland, Statistics Finland and other authorities integrate statistical data with geospatial information to offer various services, such as Grid Database and YKR data.

3. Tietosuojan merkitys paikkatietoja julkaistaessa

Tilastotietoja julkaistaessa tulee aina kiinnittää huomiota tilastolliseen tietosuojaan, jotta voidaan estää yksittäisten tilastoyksiköiden, kuten henkilöiden tai yritysten, tunnistaminen ja näitä koskevien tietojen paljastuminen julkaistuista tiedoista. Tämä vaatimus tulee sekä kansallisen tilastolain (280/2004) 11 §:stä että Euroopan tilastoasetuksen (N:o 223/2009) 20 artiklassa.

Tilastoyksikön maantieteellinen sijainti, mahdollisesti yhdistettynä vielä muihin yksikköä kuvaaviin tietoihin, voi toimia epäsuorana tunnisteena, eli sijaintitiedon avulla yksikön identiteetin päättely voi olla mahdollista. Aivan tarkkaa osoitetietoa voidaan pitää suorana tunnisteena, jonka perusteella tilastoyksikkö voidaan identifioida (HE 154/2012). Tämän takia tietosuojan merkitys korostuu, kun tilastotietoja yhdistetään tarkkoihin sijaintitietoihin.

Samalla kun mahdollisuus tuottaa tilastotietoja tarkemmilla alueluokituksilla kasvaa, myös kysyntä tällaisille tiedoille kasvaa. Jotta tietosuojasta voitaisiin asiaankuuluvasti huolehtia pienaluetilastojen yhteydessä, on tärkeää tunnistaa tietosuojaan liittyvät erityispiirteet ja haasteet, joita tarkat sijaintitiedot aiheuttavat.

4. Paikkatiedon vaikutus paljastumisriskiin

Tilastollisen tietosuoja yhteydessä puhutaan tilastollisesta paljastumisesta ja paljastumisriskistä. Tilastollinen paljastuminen (engl. *statistical disclosure*) tapahtuu, kun jokin taho pystyy tunnistamaan yksittäisen tilastoyksikön (esim. henkilön tai yrityksen) julkaistuista tilastotiedoista tai oppii jotain uutta kyseisestä tilastoyksiköstä (Hundepool ym. 2012, 2). Tilastoyksiköihin kohdistuvaa paljastumisriskiä voidaan mitata tai määritellä useilla eri tavoilla. Esimerkiksi aggregoidussa, taulukkomuotoisessa aineistossa paljastumisriski voidaan määritellä koskemaan ns. pieniin soluihin kuuluvia tilastoyksiköitä. Tilastollisen tietosuojan avulla paljastumisriski pyritään laskemaan hyväksyttävälle tasolle.

Paikkatietojen osalta paljastumisriskiin vaikuttavia tekijöitä ovat muun muassa paikkatietojen tarkkuus, havaintojen muodostamat rypäät ja autokorrelaatio sekä erilaisten alueluokitusten käytön seurauksena muodostuvat leikkaavat tilastoalueet. (GSGF Europe 2022, 2.)

4.1. Paikkatiedon tarkkuus

Mitä tarkempi paikkatieto on kyseessä, sitä vähemmän kyseeseen paikkaan yleensä sijoittuu tilastoyksiköitä (vrt. paikkatietona Suomi, yksittäinen kunta tai yksittäinen 1 km x 1 km karttaruutu). Toisin sanoen mitä pienempää tunnistettavaa maantieteellistä aluetta tilastotieto koskee, sitä suurempi on yleensä tiedon taustalla olevien tilastoyksiköiden paljastumisriski. Tilastotietoja ei pitäisi julkaista, jos tieto koskee vain yksittäistä havaintoa, ts. tilastoyksikköä, koska tällöin kyseisen tilastoyksikön paljastumisriski on liian suuri.

Kun tilastotietoja yhdistetään (tarkkoihin) paikkatietoihin, tulee siis aina ennen tietojen julkaisua varmistua siitä, että maantieteellinen sijainti ei rajaudu niin pieneksi, että tietoja tuotetaan esimerkiksi vain yhteen tai kahteen tilastoyksikköön perustuen. Tässä on huomioitava myös muut tilastoyksiköitä kuvaavat ja niiden tunnistamista helpottavat tiedot (epäsuorat tunnisteet), kuten esimerkiksi henkilön ikä, sukupuoli ja pääasiallinen toiminta.

Esimerkki: Ruututietokanta on Tilastokeskuksen julkaisema tuote, joka sisältää koordinaattipohjaisia tilastoaineistoja karttaruutuihin laskettuna. Keskeiset tiedot esimerkiksi asukasrakenteesta, tuloista ja työpaikoista on koottu koko Suomen alueelta joko 250 m x 250 m-, 1 km x 1 km tai 5 km x 5 km-karttaruuduittain. Tietosuojan vuoksi ruudun tiedot (pl. perusjoukkojen lukumäärätiedot, esim. henkilöt yhteensä, tulonsaajat yhteensä) peitetään, mikäli ruudussa on alle 3 tai 10 tapausta (tietoryhmästä riippuen). Ruudun asukkaiden tulotiedot suojataan, jos ruudussa on alle 10 täysi-ikäistä asukasta. Näin myös siinä tapauksessa, että asukkaita olisi alaikäiset ja täysi-ikäiset mukaan lukien yhteensä yli 10. Paljastumisriskin ja suojauksen tarpeen määrittelyssä huomioidaan siis paitsi

henkilöiden kokonaismäärä ruudussa, niin myös henkilöiden lukumäärä iän perusteella tehdyissä luokissa. (Ruututietokanta Tilastokeskus; Tilastokeskus 2023.)

4.2. Havaintojen muodostamat rypäät ja autokorrelaatio

Toblerin Maantieteen ensimmäisen lain mukaan maantieteessä kaikki on riippuvaista kaikesta, mutta lähellä olevat asiat ovat riippuvaisempia toisistaan kuin kauempana olevat (Dempsey 2014). Tobler kuvaa ilmiötä, jota kutsutaan spatiaaliseksi autokorrelaatioksi.

Lyhyesti ilmaistuna aineisto on spatiaalisesti autokorreloitunutta, kun (maantieteellisesti) lähellä toisiaan olevat havainnot saavat samankaltaisia arvoja (positiivinen autokorrelaatio) tai päinvastoin lähellä toisiaan olevilta havainnoilta mitataan hyvin erilaisia arvoja (negatiivinen autokorrelaatio). Vahva positiivinen autokorrelaatio nostaa havaintojen paljastumisriskiä, sillä havaintojen perusteella muodostetut aggregaatit voivat olla hyvin lähellä yksittäisten havaintojen arvoja. (GSGF Europe 2022, 2–3.)

Paljastumisriskin arvioinnissa tulee siis kiinnittää huomiota havaintojen mahdollisesti muodostamiin rypäisiin (engl. *cluster*): onko rypäitä havaittavissa ja esiintyykö vahvaa positiivista autokorrelaatiota. Eräs yleisesti käytetty testi spatiaaliseen autokorrelaatiolle on Moranin I-testi (de Bellefon & Bouayad-Agha 2018, 56).

4.3. Leikkaavat tilastoalueet

Pistegeometriaan perustuvaa tilastointia hyödyntäen tilastotietoja on mahdollista tuottaa erilaisilla alueluokituksilla. Erilaisia alueluokituksia ovat mm. hallinnolliset alueet (kunnat, maakunnat, suuralueet, hyvinvointialueet), maantieteelliset alueet (esim. erikokoiset ruututiedot tai eri koordinaattijärjestelmien mukaiset ruututiedot) sekä muut eri käyttötarkoituksia varten olevat alueluokitukset, kuten esim. postinumeroalueet ja vaalipiirit. Tilastotietoja käyttävät asiakkaat saattavat toivoa tietoja omia tarpeitaan parhaiten palvelevalla alueluokituksella, joka ei välttämättä vastaa muussa tilastotuotannossa käytettyjä luokituksia. Erilaisissa paikkatietopalveluissa tilastotietojen tuottaminen pieniltäkin alueilta on mahdollista ja käyttäjä saattaa pystyä räätälöimään alueluokitusta tietotarpeisiinsa sopivaksi.

Kun samasta aihepiiristä tuotetaan tietoja eri alueluokituksilla, jotka eivät ole sisäkkäisiä (engl. *non-nested classifications*), saattaa luokitukset päällekkäin asettamalla syntyä ns. leikkaavia tilastoalueita. Mikäli tällainen leikkaava tilastoalue on hyvin pieni, toisin sanoen siihen kuuluu vain yksi tai muutama tilastoyksikkö, voi tietojen julkaiseminen johtaa näiden tilastoyksiköiden identiteetin ja/tai ominaisuustietojen paljastumiseen. Tällaista leikkaavien tilastoalueiden muodostumista kutsutaan englanninkielisellä termillä *geographic differencing* (GSGF Europe 2022, 3).

5. Tilastolliset tietosuojamenetelmät

Koska paikkatiedot monesti nostavat tilastoyksiköiden paljastumisriskiä, on paikkatietoja sisältäviin tilastotietoihin syytä soveltaa tilastollisia tietosuojamenetelmiä riskin pienentämiseksi. Paikkatietoja sisältävien tilastotietojen erityispiirteet on huomioitava sopivan suojausmenetelmän valinnassa. Tällaisia erityispiirteitä ovat mm. edellisessä kappaleessa kuvailtu leikkaavien tilastoalueiden mahdollinen syntyminen sekä alueluokitusten laajuudesta johtuvat datan suuri määrä (vrt. Suomessa on vain reilu 300 kuntaa, mutta yli 300 000 neliökilometrin ruutua). Datan suuri määrä voi vaikuttaa

olennaisesti laskentakapasiteettiin, joka vaaditaan aineiston käsittelyyn ja suojausmenetelmien toteuttamiseen (GSGF Europe 2022, 5).

Tilastolliset tietosuojamenetelmät voidaan karkeasti jakaa kahteen luokkaan: ennen aggregointia (engl. *pre-tabular*) ja aggregoinnin jälkeen (*post-tabular*) sovellettaviin menetelmiin.

Ennen aggregointia sovellettavat menetelmät ovat luonteeltaan sotkevia menetelmiä. Niillä pyritään muuttamaan yksikkötason aineiston arvoja siten, että sotketusta aineistosta muodostettujen aggregaattien perusteella ei voida enää varmuudella päätellä mitään alkuperäisistä yksikkötason arvoista. Esimerkkejä tällaisista menetelmistä ovat ns. *record swapping* sekä kohinan lisääminen. (GSGF Europe 2022, 4.)

- Record swapping -menetelmässä nimensä mukaisesti vaihdetaan tilastoyksiköiden arvoja keskenään. Yleensä vaihtaminen kohdistuu yksiköiden paikkatietoon, eli yksiköitä ”siirrellään” alueilta toisille. *Targeted record swapping* eli TRS on menetelmä, jossa aluetietojen vaihtaminen kohdistetaan yksiköihin, joihin tietyillä mittareilla kohdistuu suurin paljastumisriski. Lisää TRS-menetelmästä voi lukea esim. Shlomo ym. (2010).
- Kohinan lisääminen (engl. *blurring*, *noise masking*) tarkoittaa satunnaiskohinan lisäämistä sellaisiin tilastoyksikön arvoihin, joihin kohdistuu suuri paljastumisriski. Aineiston epätarkkuus luonnollisesti kasvaa, mitä enemmän kohinaa lisätään. Kohinan lisäämisestä voi lukea lisää esim. Hundepool ym. (2012, 54–60).

Aggregoinnin jälkeen sovellettavat tilastolliset tietosuojamenetelmät ovat usein rajoittavia menetelmiä. Toisin sanoen menetelmät perustuvat siihen, että aggregoinnin tuloksena saatujen tietojen julkaisemista rajoitetaan, mikäli niihin arvioidaan kohdistuvan liian suuri paljastumisriski. Hyvä esimerkki tällaisesta menetelmästä on peittäminen (engl. *suppression*), jossa tilastoluvut jätetään liian suuren paljastumisriskin tapauksessa julkaisematta. Peittämisen sijaan tietoja voidaan myös aggregoida lisää esim. yhdistelemällä sellaisia alueita, joihin kohdistuu paljastumisriski, kunnes uusiin aggregaatteihin ei enää katsota kohdistuvan paljastumisriskiä. (GSGF Europe 2022, 5.)

Rajoittavien menetelmien lisäksi on myös olemassa sotkevia menetelmiä, joita voidaan soveltaa aggregoinnin jälkeen. Esimerkki tällaisesta menetelmästä on Cell Key -menetelmä, eli CKM, joka perustuu tilastoyksiköille satunnaisesti määriteltyihin avainlukuihin, joiden perusteella tilastoyksiköistä laskettuihin summatietoihin lisätään kohinaa. Cell Key -menetelmästä tarkemmin kertoo esim. Fraser & Wooton (2006).

6. Kansainväliset suositukset tietosuojasta tilasto- ja paikkatietoja yhdistettäessä

Tilasto- ja paikkatietojen yhdistäminen on kansainvälisesti kiinnostava aihe, kuten myös tietosuojasta huolehtiminen tässä yhteydessä. Muun muassa YK:n alainen asiantuntijaryhmä Expert Group on the Integration of Statistical and Geospatial Information (UN EG-ISGI) on laatinut suosituksia tietosuojan ja siihen liittyvien haasteiden huomioimiseksi tilasto- ja paikkatietoja yhdistettäessä. Nämä suositukset voidaan tiivistää seuraaviin kohtiin.

- Tietoisuutta paikkatietojen erityispiirteistä tietosuojasta huolehdittaessa tulee lisätä maailmanlaajuisella, alueellisella ja kansallisella tasolla.

- Edellä mainitut erityispiirteet tulee huomioida kansallisissa tilasto- tai tietosuojalaeissa, tietojen julkaisukäytännöissä, kansallisissa ohjeistuksissa sekä laadunvarmistukseen käytettävissä viitekehyksissä.
- Yhteistyötä akateemisen yhteisön ja muiden virallisten tahojen kanssa tulee edistää kehittämisen tukemiseksi.
- On laadittava käsikirjoja paikkatietoja sisältäviin tilastotietoihin soveltuvista tilastollisista tietosuojamenetelmistä ja -käytännöistä, joita voidaan hyödyntää kansallisella tasolla tai yleisesti.
- Paikkatietojen teknisen käsittelyn ja julkaisun tietosuojaa tulee parantaa mm. integroimalla olemassa oleviin tietosuojaohjelmistoihin aluetietoja koskeva ulottuvuus.
- Kansallisella tasolla tulee kiinnittää huomiota paikkatietojen erilaisiin käyttöoikeuksiin eri käyttäjäryhmien välillä.
- Kansallisten tilastovirastojen osaamista ja resursseja paikkatietoja sisältävien tilastojen tietosuojamenetelmissä tulee kehittää.

(UN EG-ISGI 2022, 29–30.)

7. Tilasto- ja paikkatietojen yhdistäminen Suomessa ja Tilastokeskuksessa

Suomessa joillain viranomaisilla (ml. Tilastokeskus) ja tutkimuslaitoksilla on pääsy hallinnollisiin rekisteriaineistoihin, joiden avulla voidaan tuottaa erilaisia palveluita ja tuotteita, joissa tilastotietoja yhdistetään erilaisiin paikkatietoihin. Tällaisia tuotteita ovat mm. Tilastokeskuksen Ruututietokanta¹ ja postinumeroalueittaista tietoa tarjoava Paavo-palvelu² sekä yhdyskuntarakenteen seurantaan tarkoitettu ruututasoinen YKR-aineisto³, joka on käytettävissä Suomen ympäristökeskuksen Liiteri-palvelussa. Valmiiden tuotteiden lisäksi Tilastokeskuksen Asiakas- ja toimeksiantopalvelut tuottavat asiakkaiden tarpeisiin räätälöityjä maksullisia toimeksianto- ja tietokantapohjaisia palveluita, joissa asiakas voi pyytää tilastotietoja erilaisilla pienaluejaoilla, esim. kaupunkikortteleittain. Tilastokeskuksen Tutkijapalvelut tarjoavat tutkijoille mahdollisuuden analysoida etäkäyttöjärjestelmässä käyttöluovallisia henkilö- tai yritystason aineistoja, joihin voidaan yhdistää tarkimmillaan ruututasoinen sijaintitietoja INFRA Sijaintitietomodulin⁴ kautta.

Kansainvälisten suositusten (ks. edellinen kappale) mukaan kansallisessa lainsäädännössä tulisi huomioida erityispiirteet, jotka liittyvät paikkatietoja sisältävien tilastotietojen tietosuojakysymyksiin. Tällä hetkellä Suomessa Tilastokeskuksen toimintaa ohjaa vahvasti tilastolaki (280/2004), josta seuraa myös velvoite tilastointia varten kerättyjen salassa pidettävien tietojen suojaamiseen. Paikkatietoja ei ole tässä tilastolaissa erikseen mainittu. Sen sijaan hallituksen esityksessä tilastolain muuttamiseksi (HE 154/2012) todetaan "Tarkkaa koordinaattitietoa voidaan kuitenkin perustellusti pitää osoitteeseen verrattavana suorana tunnisteena ainakin silloin, kun tilastoyksikkö voidaan yksilöidä yksinomaan sen

¹ <https://stat.fi/tup/ruututietokanta/index.html>

² <https://stat.fi/tup/paavo/index.html>

³ <https://stat.fi/tup/ykraineistot/index.html>

⁴ https://taika.stat.fi/fi/aineistokuvaus.html#!?dataid=INFRA_19872022_jua_sijainti_001.xml

avulla. Sen sijaan tarkkakin koordinaattitieto, joka koskee useampaa tilastoyksikköä samanaikaisesti esimerkiksi karttapohjalla esitettynä, olisi lähtökohtaisesti tulkittavissa vain välillisen tunnistamisen mahdollistavaksi tiedoksi.”

Paikkatietojen ja pienaluetilastojen tietosuojaan on Tilastokeskuksessa kuitenkin viime vuosina kiinnitetty yhä enemmän huomiota. Yksi syy tälle on nyky-yhteiskunnan tietotekniset mahdollisuudet, jotka tuovat mukaan myös haasteita turvalliselle tiedonkäsittelylle. Avoimien, laadukkaiden, kattavien ja yhdisteltävien aineistojen määrä on kasvussa, samalla kun paikkatietosovellukset eivät enää ole pelkästään GIS (engl. *geographic information system*) -asiantuntijoiden käytettävissä. Tietosuojan varmistamisen ohella yhteiskunnallisena tavoitteena on jo olemassa olevan tiedon laaja hyödyntäminen ja avoimen tiedon lisääminen.

Yksi esimerkki paikkatietojen tietosuojan tarkemmalle pohtimiselle liittyy ruututietoihin ja erityisesti tapauksiin, joissa ruudussa on vain yksi tilastoyksikkö, kuten henkilö tai rakennus. Tilastokeskuksen soveltamiskäytännön mukaan ruututieto (ts. ruudun koordinaatit) on tällaisissa tapauksissa ollut verrattavissa henkilön tai rakennuksen osoitteeseen paljastumisriskiä arvioitaessa. Tilastoruudukko on ajassa muuttumaton, tarkasti kartalle sijoitettava aluejako, joka on avoimesti saatavissa. Mikäli ruudussa on vain yksi tilastoyksikkö, se on sijaintinsa perusteella tunnistettavissa siitä huolimatta, että ruudun vasemman alakulman koordinaatit eivät suoraan kohdistu tilastoyksikön osoitteeseen. Ruudun koolla ei tunnistamisessa ole merkitystä. Mikäli ruudussa on vain yksi tilastoyksikkö (esim. henkilö, asunto), sen ominaisuustiedot (esim. ikä tai tulot, hallintaperuste ja pinta-ala) paljastuvat vaikka ruutu olisi kooltaan viisi neliökilometriä. Ainoa tapa olla paljastamatta yksittäisten tilastoyksiköiden tietoja, on suojata niiden ruutujen tiedot, joissa tapaukset sijaitsevat.

Tällä hetkellä Tilastokeskuksessa käytetään paikkatietoihin yhdistettyjen tilastotietojen suojaamiseen rajoittavia tietosuojamenetelmiä, kuten peittämistä ja aluetietojen aggregointia. Paikkatietoja sisältävässä tilastotuotannossa suositaan myös havaintojen suoria jakaumia ristiintaulukoinnin sijaan. Esimerkiksi Ruututietokannassa ilmoitetaan ruudussa olevien henkilöiden ikä- ja sukupuolijakaumat erikseen. Aineistosta siis selviää tietyssä ruudussa olevien alle naisten lukumäärä ja alle 18-vuotiaiden henkilöiden lukumäärä, mutta ei alle 18-vuotiaiden naisten lukumäärää (elleivät kaikki ruudun henkilöt ole naisia).

Paikkatietoihin yhdistettyjen tilastotietojen suojaaminen on paikoin haasteellista. Suojauspäätöksiä joudutaan esim. Tilastokeskuksen Asiakas- ja toimeksiantopalveluissa tekemään tapauskohtaisesti, toimeksianto kerrallaan. Kun tietoja suojataan peittämällä, tulisi peittämisen olla yhteneväistä: jos jokin tieto on peitetty yhden toimeksiannon yhteydessä, niin sama tieto tulisi olla peitetty kaikissa muissakin toimeksiannoissa, joissa se esiintyy. Peittämistä on kuitenkin haastavaa, ellei jopa mahdotonta koordinoita kaikkien toimeksiantojen välillä. Vaikka rajoittavien menetelmien käyttöön liittyy haasteita, niin niiden käyttö on Tilastokeskuksessa hyvin vakiintunutta ja siirtyminen sotkeviin suojausmenetelmiin olisi suuri muutos, jota pitäisi huolellisesti valmistella huomioiden paitsi tilastotiedon tuottajat niin myös tilastotiedon käyttäjät.

Toinen esimerkki paikkatietojen tietosuojan tarkemmalle pohtimiselle liittyy aineistojen avoimuuteen muuttuvissa geopoliittisissa tilanteissa. Aiemmin avoimeksi määriteltyjen paikkatietojen voidaan jatkossa katsoa kuuluvan tunnistautumisen tai muulla tavoin säännellyn pääsyn edellyttämiksi aineistoiksi. Mikäli aiemmin avoimiksi määriteltyihin aineistoihin pääsyä päätetään rajoittaa, tulee päätöksen olla kansallinen ja koskea kaikkia vastaavia aineistoja tuottavia toimijoita.

Puhtaasti tilasto- ja paikkatietojen integraation lisäksi muutkin avoimet paikkatietoaineistot ja niihin liittyvät riskit ovat saaneet huomiota kansallisella tasolla. Valtiovarainministeriö asetti vuonna 2023 työryhmän laatimaan paikkatietoja koskevan kansallisen riskiarvion ja toimintamallin paikkatietojen käsittelyyn (hankenumero VM111:00/2023). Työryhmän laatimasta muistiosta on laadittu myös julkinen versio, jossa on mainittu myös Tilastokeskuksen paikkatietoja sisältävät aineistot osana keskeisiä viranomaisen paikkatietovarantoja (Valtiovarainministeriö 2024, 13).

8. Lähteet

de Bellefon, Marie-Pierre & Bouayad-Agha, Salima (2018). Spatial autocorrelation indices. Julkaisussa Vincent Loonis (ohj.), *Handbook of spatial analysis, Theory and practical application with R*. Insee Méthodes n°131, sivut 51–70. <https://ec.europa.eu/eurostat/documents/3859598/9462709/INSEE-ESTAT-SPATIAL-ANA-18-EN.pdf/c4f87d5b-b508-4aff-ad7d-264da463077e?t=1545319662000> . Viitattu 3.9.2024.

Dempsey, Caitlin (2014). Tobler's First Law of Geography. Kirjoitus Geography Realm -sivustolla 29.4.2014. <https://www.geographyrealm.com/toblers-first-law-geography/> . Viitattu 3.9.2024.

Euroopan parlamentin ja neuvoston asetus (EY) N:o 223/2009.

Fraser, Bruce & Wooton, Janice (2006). A proposed method for confidentialising tabular output to protect against differencing. Julkaisussa *Monographs of Official Statistics. Work session on Statistical Data Confidentiality*, sivut 299–302. Office for Official Publications of the European Communities, Luxemburg.

GSGF Europe (2022). *GSGF Europe: Managing Confidentiality in Geospatial Statistics*. Eurostat ESSnet grant project GEOSTAT4, 2020–2022. https://www.efgs.info/wp-content/uploads/2022/02/ManagingConfidentiality_in_GeospatialStatistics.pdf . Viitattu 3.9.2024.

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

Hundepool, Anco; Domingo-Ferrer, Josep; Franconi, Luisa; Giessing, Sarah; Schulte Nordholt, Eric; Spicer, Keith & de Wolf, Peter-Paul (2012). *Statistical Disclosure Control*. Wiley.

Ruututietokanta (Tilastokeskus). <https://stat.fi/tup/ruututietokanta/index.html> . Viitattu 3.9.2024.

Shlomo, Natalie; Tudor, Caroline & Groom, Paul (2010). Data swapping for protecting census tables. Julkaisussa Josep Domingo-Ferrer & Emmanouil Magkos (toim.), *Privacy in Statistical Databases*, osa 6344 sarjasta *Lecture Notes in Computer Science*, sivut 41–51. Springer.

Tilastokeskus (2023). Ruututietokanta 2023 – Käyttäjän opas. https://stat.fi/media/uploads/tup/ruututietokanta/rttk2023_opas_fi.pdf . Viitattu 3.9.2024.

Tilastolaki 280/2004.

UN EG-ISGI (2022). *The Global Statistical Geospatial Framework: Implementation Guide*. United Nations, New York. <https://ggim.un.org/documents/EG-ISGI-GSGF-Implementation-Guide-E.pdf> . Viitattu 4.9.2024.

29.11.2024 10 (10)

Valtiovarainministeriö (2024). Paikkatiedon kansallisen riskiarvion työryhmän muistio (julkinen).
VN/15876/2023. https://api.hankeikkuna.fi/asiakirjat/4e3fe1bo-3564-4c8d-a3f2-bo8909b85d22/832b2cb2-a5fb-4c74-9e56-795a4f8dfcof/MUISTIO_20240613060728.PDF . Viitattu
6.9.2024.