

Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessin kuvaus ja kehittämistarpeiden tunnistaminen – tapaustutkimus

Title: Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessin kuvaus ja
kehittämistarpeiden tunnistaminen - tapaustutkimus

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.

© EFGS and Eurostat 2024

Sisältö

1. Johdanto.....	3
2. Tieto- ja tilastoprosessin loogiset tietovarannot ja prosessit.....	3
3. Organisaatietietovaranto	4
4. Sijaintitietojärjestelmään perustuva yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessi .	6
5. Sijaintitietojen päättelyprosessin peilaaminen GSGF-viitekehykseen	8
6. Haasteet ja kehittämissuhteet.....	12
7. Opet sijaintitiedon päättelyprosessin uudistuksesta.....	13
8. Lähteet.....	13

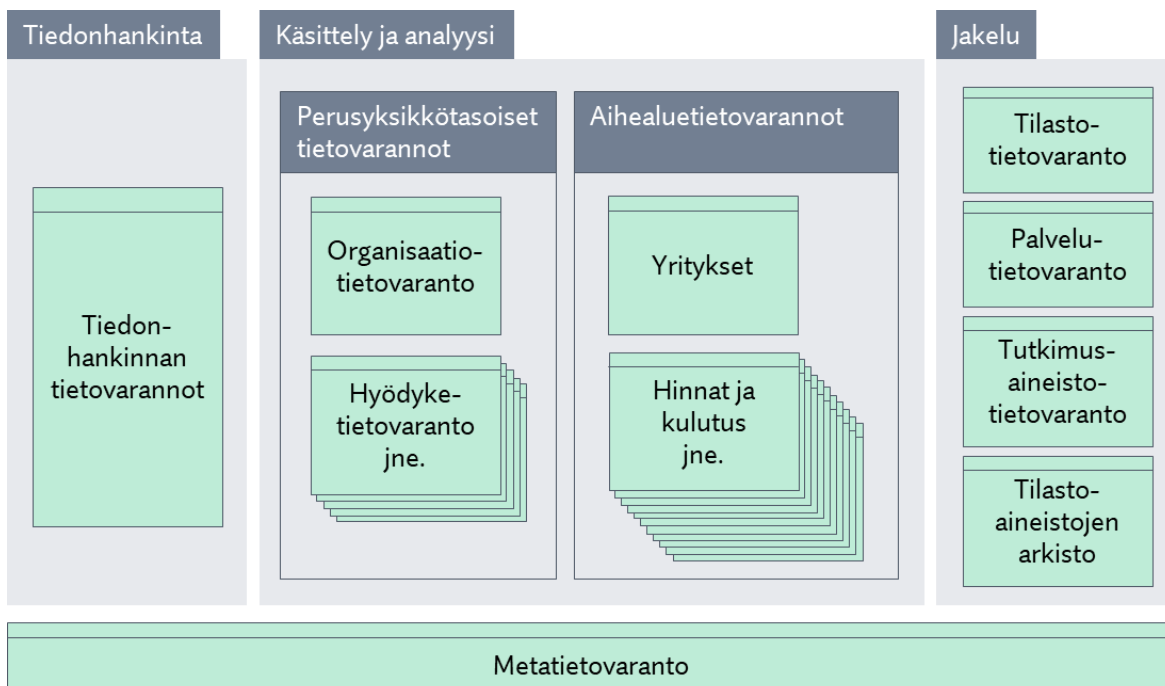
1. Johdanto

Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen johtamiseen käytetty päättelyprosessi uudistettiin Tilastokeskuksen yleisen tavoitteen ja kokonaisarkkitehtuurityössä tehtyjen määritysten mukaisesti vuosina 2020–2021.

Uusi sijaintitietojen päättelyprosessi rakennettiin yhteistyössä sijaintitietoasiantuntijoiden kanssa ja uusi prosessi käyttää uuden sijaintitietojärjestelmän teknologiariippumattomia palveluita sekä hyödyntää kansallista pysyvää rakennustunnusta sijaintitietojen johtamisessa. Uudistuksen myötä yritys- ja toimipaikkarekisterin sijaintitietojen laatu parani ja sijaintitietojen vertailukelpoisuus Tilastokeskuksen sisällä parani laajemmin.

2. Tieto- ja tilastoprosessin loogiset tietovarannot ja prosessit

Tilastokeskuksen yleinen tavoite on jäsentää tieto- ja tilastoprosessi tilastoinnin yleisen prosessimallin GSBPM 5.1:n mukaisesti tiedonhankinta-, käsittely-, analyysi- ja jakeluvaiheeseen sekä koota eri tuotantovaiheiden tiedot Tilastokeskuksen kokonaisarkkitehtuurissa määriteltyjen loogisten tietovarantojen mukaisesti (Kuva 1).

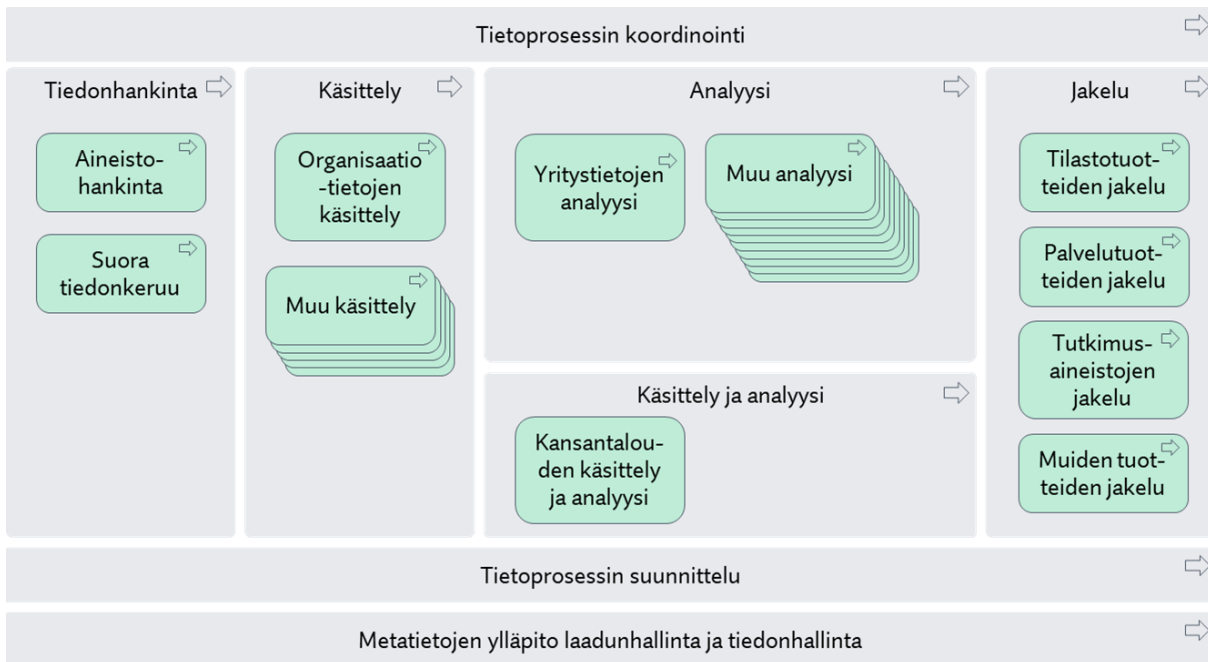


Kuva 1. Mallikuva Tilastokeskuksen tietoprosessin loogisista tietovarannoista.

Tavoitetilan tieto- ja tilastoprosessissa (Kuva 2):

- Kaikki tieto- ja tilastoprosessien lähdeaineistot ja suorat tiedonkeruutiedot hankitaan ja kerätään keskitetysti.
- Tieto tallennetaan ja käsitellään kertaalleen perusyksikkötasoisissa päätietyryhmien mukaan määritellyissä tietovarannoissa näiden varaan rakennettuja tietojärjestelmiä käyttäen.

- Valmiiksi käsiteltyjä aineistoja ja yhtenäisessä muodossa tallennettuja tietoja yhdistelemällä koostetaan aihealuekohtaiset yksikköaineistot tarkempaa aihealuekohtaista analysointia varten näin varmistuen, että eri aihealueiden tiedot ovat keskenään vertailukelpoisia.
- Jakeluvaiheessa valmiiksi analysoiduista tiedoista koostetaan ja julkaistaan jaeltavat tilasto- ja palvelutuotteet, muut tuotteet sekä tutkimusaineistot.



Kuva 2. Mallikuva Tilastokeskuksen tavoitetilan tieto- ja tilastoprosessista.

3. Organisaatietietovaranto

Organisaatietietovaranto on yksi kuudesta Tilastokeskuksen kokonaisarkkitehtuurissa määritellyistä perusyksikkötasoisesta tietovarannosta (Kuva 1). Tietovarannon keskiössä on Tilastokeskuksen ylläpitämä kansallinen Yritys- ja toimipaikkarekisteri, joka kattaa

- kaikki y-tunnukselliset yritykset, joiden tiedot ovat saatavana Verohallinnosta,
- yksityiset elinkeinonharjoittajat,
- julkisyhteisöt eli kaikki valtion virastot, kunnat, kuntayhtymät ja hyvinvointialueet,
- voittoa tavoittelemattomat yhteisöt sekä
- yritysten ja yhteisöjen toimipaikkatiedot.

Yrityksiä ja yhteisöjä koskevia tietoja ovat muun muassa y-tunnus, nimi, osoite, yhteystiedot, toimiala ja oikeudellinen muoto. Toimipaikkatietoja ovat esimerkiksi toimipaikkatunnus, sijaintiosoite, henkilöstön suuruusluokka ja henkilöstömäärä kunnassa. Yritysrekisterin koko tietosisältö on esitelty Tilastokeskuksen verkkosivuilla¹.

Tieto- ja tilastoprosessin tavoitetilan mukaisesti (Kuva 2) kaikki yrityksiä ja yhteisöjä koskevat perus- ja taloustiedot tallennetaan yhtenäisessä ja yhdisteltävässä muodossa organisaatietietovarantoon.

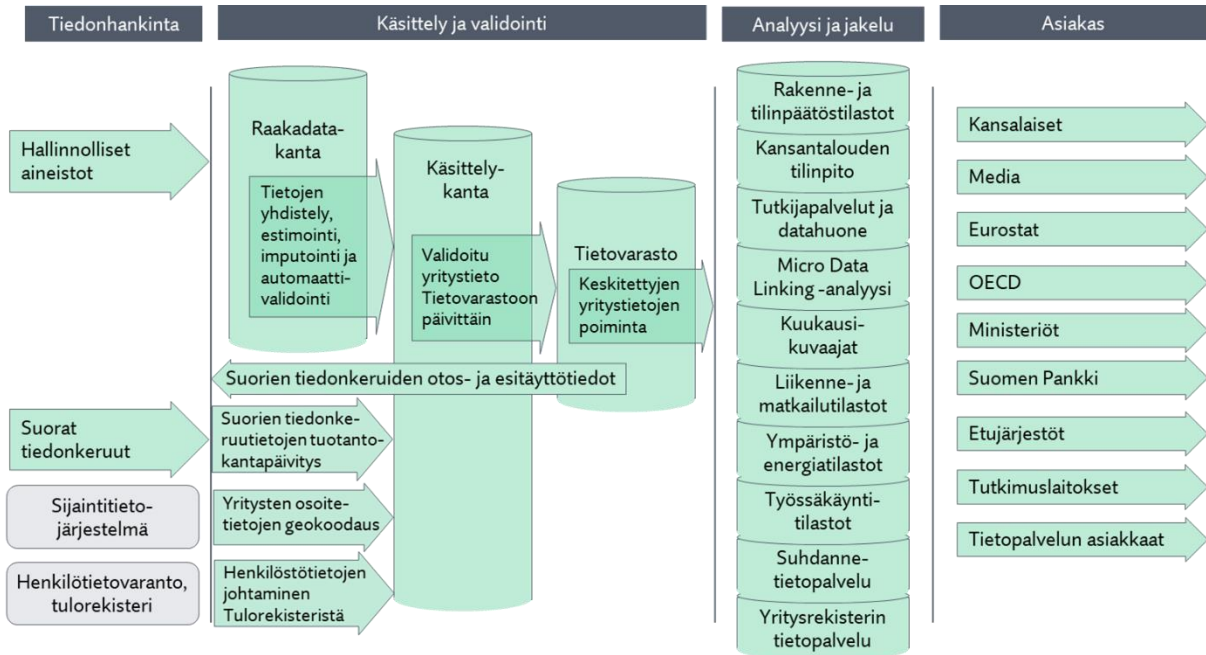
¹ Tilastokeskus, 2024. Yritysrekisterin tietosisältö. <https://stat.fi/yritysrekisteri/yritysrekisterin-tietosisalto>

Tietovarannon perustiedot koostuvat Yritys- ja toimipaikkarekisterin tiedoista ja valtaosa perustiedoista saadaan suoraan tai johdettua PRH:n ja Verohallinnon ylläpitämän Yritys- ja yhteisötietojärjestelmän (YTJ), Tulorekisterin ja Valtionkonttorin tiedoista. Monitoimipaikkaisten yritysten osalta joudutaan kuitenkin turvautumaan suorien tiedonkeruiden kautta saataviin tietoihin toimipaikkatasoisten hallinnollisten aineistojen puuttuessa.

Valtaosa tietovarannon taloustiedoista koostuu Verohallinnon elinkeinoveroaineistosta johdettavista vuositaso-tilinpäätöstiedoista, oma-aloitteisten verojen aineistoista johdettavista kuukausitaso- myyntitiedoista sekä tulorekisterin perusteella johdettavista henkilöstö- ja palkkasummatiedoista. Yritysten tilinpäätöstietoja ja kuukausitaso- myyntitietoja täydennetään suurimpien yritysten osalta suorien tiedonkeruiden avulla, koska Verohallinnon aineistoista ei ole johdettavissa kaikkia EU-asetusten mukaisia tietoja. Verohallinnon aineistot eivät myöskään kata kaikkia tietotarpeita mm. kansantalouden tilinpidon tarvitsemia liikevaihto- ja kuluerittelytietoja. Tietojärjestelmään tallennetaan edellä mainittujen tietojen lisäksi palveluiden ulkomaan kauppaan, palveluiden ja teollisuuden hyödykkeisiin ja teollisuuden ympäristösuojelumaihin liittyviä tietoja sekä rahoitus- ja vakuutussektorin tilinpäätöstiedot.

Organisaatietietovarannon tietojen kestitetty päivitys, käsittely, validointi ja varastointi tapahtuu Yritysrekisteri tilastoinnin ytimenä (YTY) -tietojärjestelmän avulla (Kuva 3). Tietojärjestelmä koostuu raakadata-, käsittely- ja tietovarasto -SQL-kannoista, tietojen manuaaliseen käsittelyyn tarkoitetuista .net sovelluksista sekä lukuisista SAS- ja SQL-proseduureista.

- Raakadatatietokanta toimii tietojärjestelmän lastauslaiturina (staging area), johon alkuperäisessä muun muassa csv-, dat-, json tai xml-formaatissa vastaanotettavat aineistot muunnetaan ja tallennetaan sql-formaatissa. Näin varmistetaan asiantuntijoille pääsy alkuperäiseen käsittelemättömään tietoon helposti yhdisteltävässä muodossa.
- Alkuperäisistä yhtenäisessä muodossa tallennetuista tiedoista johdetaan tilastoinnin tarvitsemat yksikkötason tiedot ja tallennetaan nämä yhtenäisessä muodossa käsittelykantaan. Käsittelykannassa ylläpidetään yritysten ja muiden yhteisöjen yksikkörekisteriä, eri yksikkötasojen välisiä suhteita ja fuusiotietoja sekä yksikköihin liittyviä perus- ja taloustietoja.
- Keskitetysti käsitellyt tiedot tallennetaan tietojärjestelmän tietovarastoon, mistä kaikki tiedot ovat koko Tilastokeskuksen mm. yritystilastojen, yritystiedonkeruiden, tietopalvelun, tutkijoiden ja kansantalouden tilinpidon käytettävissä. Koska tietojärjestelmän tietovarasto on laaja, on tietovarastoon rakennettu näkymiä, joihin poimitaan osajoukko tietovaraston tiedoista määriteltyä tarkoitusta varten. Tietojärjestelmän näkymiä hyödynnetään mm. BI-työkaluilla.



Kuva 3. Kuva 3. Yritysrekisteri tilastoinnin ytimenä (YTY) tietojärjestelmä.

Tietovarannon perus- ja taloustietojen alueellistamista varten tarvittavat kunta- ja koordinaattitiedot johdetaan yritysten ja muiden yhteisöjen osoitetietojen perusteella. Osoitetiedot saadaan kuukausittain YTJ:stä ja saatujen osoitetietojen voidaan katsoa olevan kattavat yksitoimipaikkaisten yritysten ja yhtiöiden osalta. Monitoimipaikkaisten yritysten osalta tarvitaan kuitenkin tarkempia osoitetietoja, koska monitoimipaikkaisten yritysten YTJ:n mukainen osoite sijaitsee monesti pääkaupunkiseudulla, vaikka yrityksellä on toimintaa ympäri Suomea. Monitoimipaikkaisten yritysten toimipaikkojen osoitetiedot joudutaan kartoittamaan vuosittain suorilla tiedonkeruilla muiden tietolähteiden puutteessa. Monitoimipaikkaisten yritysten tiedonkeruiden piiriin kuuluu kaikki Suomen monitoimipaikkaiset yritykset, joita oli vuonna 2021 tiedonkeruutietojen perusteella hieman yli 7000.

Parhaillaan on käynnissä PRH:n vetämä Yrityksen digitalous -hanke, jonka visiona on kansallinen, muiden pohjoismaiden kanssa yhteentoimiva talouden toimijoiden ekosysteemi vuonna 2030. Tilastokeskus tekee tiivistä yhteistyötä hankkeen kanssa ja tavoitteena on, että ainakin osa nykyään suorilla tiedonkeruilla kerättävistä tiedoista ml. toimipaikkatason osoitetiedot olisivat osana yrityksen digitaalista identiteettiä ja saatavissa koneluettavassa muodossa ilman suoria tiedonkeruita.

4. Sijaintitietojärjestelmään perustuva yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessi

Organisaatitietovarannon sijaintitietojen johtamiseen käytetty sijaintitietojen päättelyprosessi uudistettiin vuosina 2020–2021 hyödyntämään vuosina 2018–2019 kehitettyä sijaintitietojärjestelmää.

Sijaintitietojärjestelmä koostuu sijaintitietovarannosta ja sen päälle toteutetuista teknologiariippumattomista rajapintapalveluista. Teknologiariippumattomuuden myötä sijaintitietojärjestelmän palveluita voi käyttää eri aihealueita koskevien tilastojen paikkatietojen päättelyssä. Yhteisiä palveluita käyttämällä varmistetaan taas eri aihealueita koskevien tilastojen paikkatietojen vertailukelpoisuus. Sijaintitietojärjestelmän palveluita hyödynnetään mm. asumista,

rakentamista, väestöä sekä liikennettä ja matkailua koskevien tilastojen tuotannossa. Toistaiseksi sijaintitietojärjestelmän palvelut ovat tarkoitettu vain Tilastokeskuksen sisäiseen käyttöön.

Organisaatiotietovarannon paikkatietojen päättelyprosessin uudistamisen yhteydessä yritys- ja toimipaikkarekisterin toimipaikkojen osoitetietojen geokoodausprosessi toteutettiin hyödyntäen sijaintitietojärjestelmän palveluita. Näin varmistettiin yritys- ja toimipaikkarekisterin ja tätä hyödyntävien eri aihealueiden tilastojen vertailukelpoisuus muiden sijaintitietojärjestelmää hyödyntävien aihealueiden kanssa.

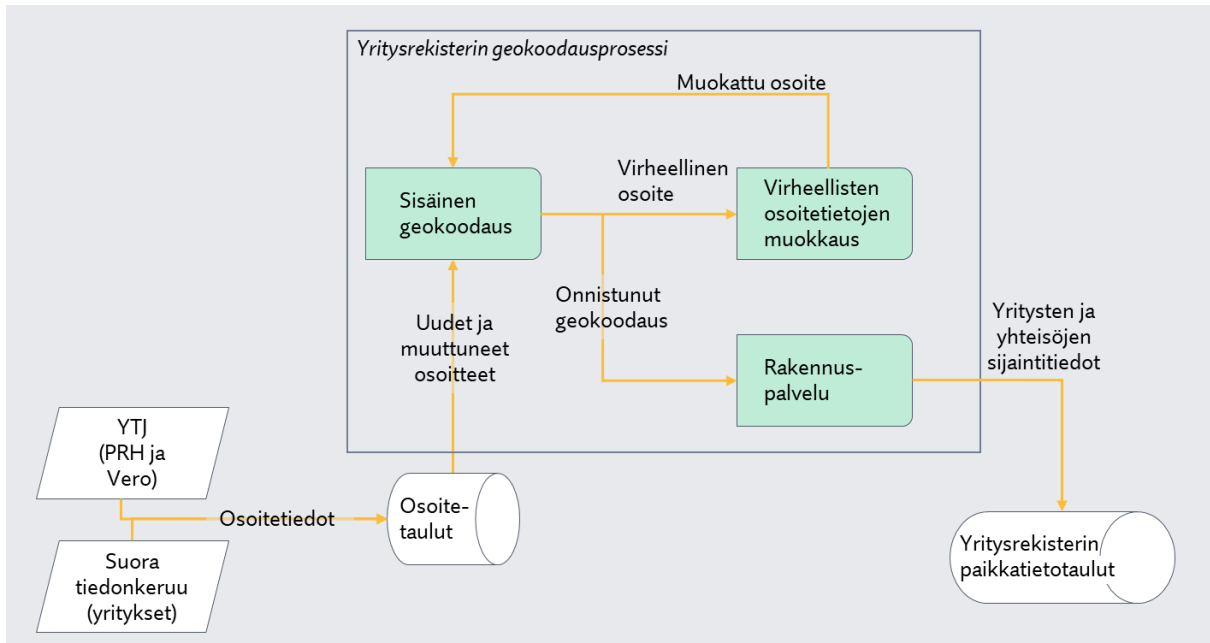
Ennen uudistusta organisaatiotietovarannon toimipaikkojen sijaintitiedot johdettiin käyttäen tähän tarkoitukseen tuotettua SAS-ohjelmaa, joka perustui Tilastokeskuksen vanhaan katutietokantaan ja alueluokituksiin. Vanha ohjelma tuotti kattavat sijaintitiedot toimipaikoille, mutta ohjelman tuottamat sijaintitiedot eivät olleet vertailukelpoisia muissa tilastoissa käytettävien sijaintitietojen kanssa ja etenkin vuonna 2017 katutietokannan ylläpidon lakkauttamisen jälkeen vanhan ohjelman tuottamien sijaintitietojen laatu ja kattavuus heikkeni.

Uusi sijaintitietojen päättelyprosessi suunniteltiin yhteistyössä IT-arkkitehdin, toiminta-arkkitehdin sekä sijainti- ja yritystietojen kohdealueiden arkkitehtuurivastaavien kanssa, ja toteutettiin yritystietojärjestelmän kehitystiimin toimesta. Näin varmistettiin, että uusi päättelyprosessi on Tilastokeskuksen arkkitehtuurilinjausten mukainen ja ratkaisuiltaan paras mahdollinen.

Uusi prosessi (Kuva 4) koostuu kuudesta päävaiheesta:

1. Osoitetietojen päivitys: Yritysten ja muiden yhteisöjen osoitetiedot päivitetään organisaatiotietovarantoon kuukausittain YTI:n perusteella ja tietovarannon osoitetietoja rikastetaan suorien tiedonkeruiden kautta saatavilla monitoimipaikkaisten yritysten toimipaikkojen osoitetiedoilla.
2. Osoitetietojen yhdistäminen pysyvään rakennustunnukseen: Sijaintitietojärjestelmän Geocoding -palvelua käyttäen etsitään toimipaikkojen osoitetietoja vastaavat pysyvät rakennustunnukset (prt).
3. Virheellisten osoitetietojen muokkaus: Osoite-prt-linkin puuttuessa osoitteet, joille osoitetta vastaava prt-arvoa ei löytynyt, yritetään korjata Sijaintitietojärjestelmän AdressSanitization-palvelua käyttäen. Palvelu palauttaa käyttäjällä korjausehdotusten lisäksi laatuindikaattorin 0-100. Korjausehdotukset katsotaan tarpeeksi laadukkaiksi, jos laatuindikaattorin arvo on yli 85.
4. Muokattujen osoitteiden yhdistäminen prt-tietoon: Onnistuneesti korjatuille osoitteille haetaan tämän jälkeen näitä vastaavat prt-arvot jälleen Geocoding -palvelua käyttäen.
5. Prt-tietoja vastaavien sijaintitietojen poiminta: Pysyvien rakennustunnusten (prt) ja toimipaikkojen osoitteiden yhdistelyn jälkeen haetaan prt- arvoja vastaavat sijaintitiedot Sijaintitietojärjestelmän BuildingService -palvelua käyttäen. Nykyisellään palvelun tarjoamista sijaintitiedoista poimitaan vain kunta- ja koordinaattitiedot, mutta palvelun läpi on poimittavissa kattavasti myös muita sijaintitietoja kuten esimerkiksi maakunta- ja seutukuntatiedot – kenties tulevaisuudessa myös pienaluejakomallin mukaiset sijaintitiedot.

6. Johdettujen sijaintitietojen tallennus organisaatietietovarantoon: Sijaintitietojärjestelmän BuildingServicen ulostulosta poimitut sijaintitiedot ja näitä vastaavat prt:t tallennetaan prosessin lopuksi organisaatietietovarannon käsittelykantaan.



Kuva 4. Sijaintitietojärjestelmän palveluita hyödyntävä yritysrekisterin toimipaikkojen osoitetietojen geokoodausprosessi.

Käyttämällä sijaintitietojärjestelmän palveluita toimipaikkojen sijaintitietojen johtamiseen ja tallentamalla prt:n mukaiset sijaintitiedot organisaatietietovarantoon, ovat organisaatietietovarannosta poimittavat sijaintitiedot vertailukelpoisia muiden sijaintitietojärjestelmää hyödyntävien tieto- ja tilastoprosessien kanssa. Organisaatietietovarannon sijaintitietoja hyödynnetään yritysten rakenne- ja tilinpäätöstilastojen ja Tilastokeskuksen yritysrekisterin lisäksi kansantalouden tilinpidossa aluetilinpidon koontiin, tutkijapalveluissa sekä laajasti Tilastokeskuksen tietopalveluissa mm. Suhdannetietopalvelussa ja Yritysrekisterin tietopalvelussa. Sijaintitietojärjestelmään käytetään organisaatietietovarannon toimipaikkojen sijaintitiedon johtamisen lisäksi asumisen ja rakentamisen sekä väestö ja yhteiskunta aihealueiden tilastoinnissa.

5. Sijaintitietojen päättelyprosessin peilaaminen GSGF-viitekehukseen

Alla olevissa taulukoissa 1–5 on peilattu uudistettua sijaintitietojen päättelyprosessi GSGF-viitekehukseen periaatteisiin. Analyysissä on hyödynnetty lisäksi GSGF Euroopassa kuvattuja vaatimuksia ja suosituksia (GEOSTAT4 and Eurostat, 2021).

Taulukko 1. Uudistetun geokoodausprosessin ja GSGF-viitekehyksen periaatteen 1 vertailu prosessin sisältä katsottuna. Periaatteessa 1 kansallinen paikkatietoinfrastruktuuri mahdollistaa sijainnin liittämisen tietoon yhdenmukaisella tavalla.

GSGF Europan esittämä vaatimus periaatteen 1 toteutumiseksi	Uudistetun geokoodausprosessin vertailu vaatimukseen
1.1 Käytä kansallisen paikkatietoinfrastruktuurin tietoja	Vaatimus toteutuu organisaatiotasolla. Uudistettu geokoodaus yhdistää Tilastokeskuksen ylläpitämät toimipaikkojen osoitetiedot Digi- ja väestötietoviraston toimittamiin rakennus- ja sijaintitietoihin. Suomesta puuttuu kuitenkin kansallisen osoiterekisteri ja toimipaikkojen osoitetiedot perustuvat yritysten ilmoittamiin tietoihin.
1.2 Käytä pistemäisiä paikkatietoja geokoodauksessa	Vaatimus toteutuu. Geokoodaus perustuu Tilastokeskuksen ylläpitämiin toimipaikkojen osoitetietoihin sekä rakennusten sijaintitietoihin.
1.3 Rakenna organisaatioiden välistä yhteistyötä virallisiin sopimuksiin perustaen	Vaatimus toteutuu, Tilastokeskus tekee tiivistä yhteistyötä eri viranomaistahojen kanssa tilastoyhteistyösopimusten puitteissa.
1.4 Edistä sijaintitietojen sisällyttämistä tilastoinnissa hyödynnettäviin kansallisiin rekistereihin	Vaatimus toteutuu osin. Uudistuksen myötä Tilastokeskuksen ylläpitämään kansalliseen Yritys- ja toimipaikkarekisteriin lisättiin DVV:n toimittamat tiedot pysyvistä rakennus tunnuksista ja sijaintitiedoista. Tilastokeskus edistää osana PRH:n vetämää Yrityksen digitalous -hanketta yritysten toimipaikkojen osoitetietojen saamista osaksi yrityksen digitaalista identiteettiä sekä kehitteillä olevia eKuitteja. Kansallinen osoiterekisteri pysyvän rakennuksen kanssa mahdollistaisi yhtenäiset sijaintitiedot eri viranomaistahojen kanssa.

Taulukko 2. Uudistetun geokoodausprosessin ja GSGF-viitekehyksen periaatteen 2 vertailu prosessin sisältä katsottuna. Periaate 2 tarkastelee yhdenmukaista sijaintia koskevan tiedon liittämistä tilasto- ja muihin tietoihin ja yhtenäisesti tuotetun paikkatiedon hallintaa.

GSGF Europan esittämä vaatimus periaatteen 2 toteutumiseksi	Uudistetun geokoodausprosessin vertailu vaatimukseen
2.1 Rakenna tehokas ja turvallinen paikkatiedon hallintaympäristö	Vaatimus toteutuu. Sijaintitietojen hallinta on keskitetty Tilastokeskuksessa Sijaintitieto-järjestelmään.

<p>2.2 Sisällytä sijainti- ja paikkatietonäkökulma organisaation kokonaisarkkitehtuuriin</p>	<p>Vaatus toteutuu. Sijainti- ja paikkatiedot on huomioitu Tilastokeskuksen kokonaisarkkitehtuurissa.</p>
<p>2.3 Varmista geokoodauksen tulosten laatu</p>	<p>Vaatus toteutuu osittain. Uuden prosessin hyödyntämät prt-tiedot ja näiden mukaiset sijaintitiedot tarkistetaan osana asumisen ja rakentamisen aiheen tilastointia ja tallennetaan keskitettyyn sijaintitietovarantoon kaikkien käytettäväksi. Sijaintitietojärjestelmän toimintalogiikka ja laadunvarmistustyö jäävät kuitenkin järjestelmän rajapintapalveluiden käyttäjillä vieraksi.</p>
<p>2.4 Varmista hallinnollisten ja muiden tietojen keruussa ja vastaanotossa sijaintitietojen oikeellisuus</p>	<p>Vaatus toteutuu osittain. Osalta yrityksiltä, vajaalta viideltä prosentilta, hallinnollisten aineistojen kautta saatavista sijaintiosoitteista puuttuu kaikki sijaintitietojen johtamiseen tarvittavat osoitetiedot. Hallinnollisten aineistojen kautta saatavissa osoitetiedoissa esiintyy myös merkittävien sijaintitietojen päättelyyn vaikuttavia virheitä.</p> <p>Monitoimipaikkaisten yritysten osoitetietoja ei löydy hallinnollisista aineistoista ja nämä joudutaan hankkimaan suorien tiedonkeruiden avulla.</p> <p>Kansallinen osoiterekisteri parantaisi hallinnollisten aineistojen kautta saatavien osoitetietojen oikeellisuutta.</p>
<p>2.5 Määrittele yhteiset paikkatiedon laadun määritelmät, jotka huomioivat alueellisen ja ajallisen johdonmukaisuuden</p>	<p>Vaatus toteutuu osittain. Uuden geokoodausprosessin myötä Tilastokeskuksen ylläpitämän yritys- ja toimipaikkarekisterin sijaintitiedot ovat ajallisesti ja alueellisesti johdonmukaisia saatavilla olevien osoitetietojen puitteissa, mutta yhteisiä sijaintitietojen laadun määritelmiä ei ole kirjattu ylös.</p>

Taulukko 3. Uudistetun geokoodausprosessin ja GSGF-viitekehyksen periaatteen 3 vertailu prosessin sisältä katsottuna. Periaatteessa 3 tavoitellaan yhteisiä aluejakoja tietojen ja tilastojen jakelussa.

<p>GSGF Europan esittämä vaatimus periaatteen 3 toteutumiseksi</p>	<p>Uudistetun geokoodausprosessin vertailu vaatimukseen</p>
<p>3.1 Luo ja ylläpidä johdonmukaista tilastollisten ja hallinnollisten alueiden kartta-aineistojen kokonaisuutta</p>	<p>Ei ole relevantti.</p>

<p>3.2 Vakiinnuta olemassa olevien tilastoruudukoiden käyttö ja tutki uusia kehittyviä maailmanlaajuisia tilastoruudukkoratkaisuja</p>	<p>Vaatus ei toteudu. Uudistuksen yhteydessä sijaintitietovarannosta poimitaan organisaatitietovarantoon tallennettavaksi ainoastaan toimipaikkojen osoitteita vastaavat prt-, kunta- ja koordinaattitiedot mutta ei tilastoruututietoja.</p> <p>Yritysten tilastoruutu- ja prt-tietoja tulisi käyttää jatkossa laajemmin osana Tilastokeskuksen yritystietopalvelua.</p>
--	---

Taulukko 4. Uudistetun geokoodausprosessin ja GSGF-viitekehyksen periaatteen 4 vertailu prosessin sisältä katsottuna. Periaate 4 on muut periaatteet poikkileikkaava ja nostaa keskiöön yhteentoimivat tilastot ja paikkatiedot.

GSGF Europen esittämä vaatimus periaatteen 4 toteutumiseksi	Uudistetun geokoodausprosessin vertailu vaatimukseen
<p>4.1 Paranna sijainti- ja paikkatietoihin liittyviä työnkuluja tietoprosesseissa</p>	<p>Vaatus toteutuu. Uudistuksen myötä parannettiin sijainti- ja paikkatietojen tietoprosessia ottamalla käyttöön pysyviä rakennustunnuksia vastaavat paikkatiedot organisaatitietovarannon yritys- ja toimipaikkarekisterissä samalla vähentäen sijaintitietojen tarkistamiseen käytettävää työmäärää.</p>
<p>4.2 Kehitä tietojen yhteentoimivuutta yhtenäistämällä eri aihealueiden semantiikkaa ja käsitteitä</p>	<p>Vaatus toteutuu. Uudistuksen myötä yritys- ja toimipaikkarekisterin sijaintitiedot perustuvat DVV:n toimittamiin tietoihin pysyvistä rakennus tunnuksista ja sijainneista.</p>
<p>4.3 Julkaise tiedot kerran ja mahdollista niiden uudelleenkäyttö alkuperäisestä lähteestä</p>	<p>Vaatus toteutuu osittain. Tilastokeskuksen yritys- ja toimipaikkarekisteriin tallennetut DVV:n toimittamiin tietoihin perustuvat sijaintitiedot julkaistaan Tilastokeskuksen sisäiseen käyttöön YTY-tietojärjestelmän tietovarastoon. Täältä tiedot ovat poimittavissa eri käyttötarkoituksiin, mm. aluetilinpidon laskentaan, maksulliseen tietopalveluun, Tilastokeskuksen tutkijapalveluun ja yritystilastointiin.</p>
<p>4.4 Mahdollista innovaatiot laajemmassa tietoekosysteemissä lisäämällä palveluiden ja semanttisen webteknologian käyttöä.</p>	<p>Vaatus toteutuu osittain, ks. periaate 4.3.</p>

Taulukko 5. Uudistetun geokoodausprosessin ja GSGF-viitekehyksen periaatteen 5 vertailu prosessin sisältä katsottuna. Periaatteessa 5 huomio on tilastojen ja paikkatietojen helppossa käyttöönotossa.

GSGF Europan esittämä vaatimus periaatteen 5 toteutumiseksi	Uudistetun geokoodausprosessin vertailu vaatimukseen
5.1 Ota käyttöön selkeät ja yksinkertaiset tiedon lisensointikäytännöt	Ei ole relevantti.
5.2 Hyödynnä palvelupohjaisia tietoportaaleja, jotka tukevat tiedon dynaamista integrointia	Vaatimus toteutuu. Uusi geokoodausprosessi on rakennettu Sijaintitietojärjestelmän yleiskäyttöisiä teknologiariippumattomia raja-pintapalveluita hyödyntäen ja prosessin tuottamat toimipaikkojen sijaintitiedot tallennetaan YTY-tietojärjestelmän tietovarastoon. Täältä tiedot ovat integroitavissa muihin tietojärjestelmiin ja siten Tilastokeskuksen sisäiseen käyttöön.
5.3 Määrittele selkeät säännöt yksityisyyden suojaamiseksi	Ei ole relevantti. Yritysten ja muiden yhteisöiden toiminnan sijaintiin ja toimipaikkoihin liittyvät tiedot ovat julkisia tilastolain 18 § mukaan.
5.4 Helpota tiedon löydettävyyttä paremman luetteloinnin ja ohjeistuksen avulla	Vaatimus toteutuu osittain. Käyttöön otetun sijaintitietojärjestelmän palveluiden käytölle löytyy ohjeet ja niiden palveluiden tarjoamat paikkatiedot on kuvattu. Sijaintitietojärjestelmän tai tämän palveluiden toimintaa ei kuitenkaan tunneta laajasti Tilastokeskuksessa. Ohjeiden merkitys kasvaa, kun palveluiden käyttäjäkunta kasvaa.

6. Haasteet ja kehittämissuhteet

Uudistuksen myötä organisaatitietovarannon yritys- ja toimipaikkarekisterin toimipaikkojen sijaintitiedot ovat laadukkaat ja vertailukelpoiset sijaintitietojärjestelmän teknologiariippumattomien rajapintapalveluiden ja kansallisen pysyvän rakennustunnuksen käyttöönoton myötä. Organisaatitietovarannon tietosisällössä ja sijaintitietojen päättelyprosessissa on kuitenkin vielä parannettavaa.

Monitoimipaikkaisten yritysten toimipaikkojen osoitetiedot kerätään suorilla tiedonkeruilla, koska tietoja ei löydy hallinnollisista aineistoista. Tilastokeskus edistää osana PRH:n vetämää Yrityksen digitalous -hanketta toimipaikkojen osoitetietojen saamista osaksi yrityksen digitaalista identiteettiä sekä kehitteillä olevia eKuitteja. Tavoitteena on saada kattavat osoitetiedot Suomen yritystoiminnan osalta. Tällä hetkellä noin viideltä prosentilta yrityksistä puuttuu osoitetiedot kokonaan.

Organisaatitietovarannon tietosisältö on osin puutteellinen esimerkiksi suurten kaupunkien toimipisteiden vrt. toimipaikkojen sekä toimipaikoiksi katsottavien miehittämättömien voimalaitosten

osalta. Tietovarannon kattavuutta on tarkoitus edistää uusien aineistojen mm. Opetushallituksen varhaiskasvatuksen tietovarannon ja Tilastokeskuksen päästö- ja energiatietovarannon käyttöönotolla.

Uudistuksen myötä organisaatietietovarantoon tallennettavaa pysyvää rakennustunnusta tai näitä vastaavia ruututietoja ei hyödynnetä yritys- ja toimipaikkarekisteripohjaisessa tietopalvelussa. Pysyvä rakennustunnus tulisi tuoda osaksi tietopalvelun käyttämiä näkymiä (ks. luku 3. Organisaatietietovaranto).

7. Opit sijaintitiedon päättelyprosessin uudistuksesta

Kehittämällä tilastotuotantoa kokonaisarkkitehtuurimenetelmiä hyödyntäen ja jäsentämällä tieto- ja tilastoprosessi tilastoinnin yleisen prosessimallin (GSBPM v.5.1) mukaisesti voidaan saada merkittäviä laatu- ja tehokkuushyötyjä.

Keskittämällä tilastotuotannossa käytettävät tiedot ja näihin liittyvä käsittely- ja validointityö tietovarantoihin, voidaan vähentää päällekkäistä käsittelytyötä, kun samaa aineistoa tai tietoa ei tallenneta ja käsitellä eri tilastoissa erikseen. Vastaavasti tietovarantojen hyödyntäminen tilasto- ja tietotuotannossa mahdollistaa vertailukelpoiset tiedot eri aihealueiden kesken.

Uudistuksen myötä organisaatietietovarannon yritys- ja toimipaikkarekisterin sijaintitietojen päättely toteutettiin yhteiskäyttöisiä teknologiariippumattomia sijaintitietojärjestelmän rajapintapalveluita käyttäen. Palveluiden käyttämällä sama asia on mahdollista tehdä samalla tavalla aihealueesta riippumatta. Samalla ylläpidon tarve vähenee, kun palveluiden ylläpito on keskitetty, ja tarve omille tilastoajoille vähenee.

Uudistuksessa organisaatietietovarannon yritys- ja toimipaikkarekisterin sijaintitiedot sidottiin pysyvään rakennustunnukseen. Käyttämällä rakennustunnuksen mukaisia paikkatietoja, parannetaan sekä tietojen vertailukelpoisuutta muiden tilastollisten paikkatietojen kanssa että tietojen toistettavuutta.

Pysyvän rakennustunnuksen ja keskitettyjen sijaintitietovarannon hyödyntäminen mahdollistaa uusien aluetietojen aiempaa ketterämmän käyttöönoton. Uusi aluetieto voidaan lisätä yhteiseen sijaintitietovarantoon ja huomioida yhteiskäyttöisissä palveluissa, minkä jälkeen uusi aluetieto on suoraan eri aihealueiden hyödynnettävissä.

8. Lähteet

GEOSTAT4 and Eurostat, 2021. GSGF Europe - Requirements and Recommendations. EU, 945503 - 2019-FI-GEOSTAT4. [Linkki GSGF Europe – Requirements and Recommendations -dokumenttiin: https://www.efgs.info/gsgf-europe-geostat-information-service/geostat-information-service/requirements-and-recommendations/](https://www.efgs.info/gsgf-europe-geostat-information-service/geostat-information-service/requirements-and-recommendations/)