

Paikkatieto Tilastokeskuksen yritystietojärjestelmässä – tapaustutkimus

Title: Paikkatieto Tilastokeskuksen yritystietojärjestelmässä – tapaustutkimus

Title in English: Geospatial Information in the Business Data Production Process Finland – A Case Study

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. Yhteenveto.....	3
2. Summary in English	3
3. Lähtötilanne ja kehittämistarpeet	3
3.1. Nykytilan tarkastelu arvoketjujen avulla.....	4
4. Tavoitelaehdotuksen kuvaus.....	5
5. Tavoitelaehdotuksen eheyden ja kattavuuden tarkastelu	6
6. Lähteet.....	7
Liite 1 - Sijaintitietojen päättelyprosessi sekä prosessiin liittyvät roolit ja toimijat, nykytila	8
Liite 2 - Sijaintitietojen päättelyprosessi sekä prosessiin liittyvät roolit ja toimijat, tavoitetila	9
Liite 3 – Toimenpide-ehdotukset	10

1. Yhteenveto

Tässä dokumentissa on hahmoteltu tavoitetilaa sijaintitietojen tietoekosysteemille Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessin näkökulmasta. Työ aloitettiin kuvaamalla päättelyprosessin nykytila ja jatkettiin tunnistamalla kehittämistarpeet sekä kuvaamalla tavoitetilaehdotus. Tavoitetilan eheyttä ja kattavuutta testattiin vielä asiantuntijahaastatteluin. Lisäksi dokumentissa on avattu työn aikana saatu oivallus siitä, kuinka arvoketjuja soveltamalla voidaan tunnistaa tehokkaasti prosessiin liittyviä kehittämistarpeita sekä kuinka arvoketjuja ja arvovirtoja soveltamalla voidaan mitata tiedon laatua läpi sijaintitietoihin liittyvien tuotantoprosessien.

Työn taustalla on ollut tarve testata globaali tilastojen ja paikkatiedon yhteentoimivuuden viitekehysten, GSGF:n, mallia kansallisissa yhteistuotantoratkaisuissa tilastoprosessien tehostamiseksi. Toteutettu kehittämistarpeiden tunnistaminen ja tavoitetilan kuvaaminen on nojannut GSGF Europe -tietokokonaisuuteen ja siinä erityisesti GSGF Europe – vaatimukset ja suositukset -dokumenttiin. Tarkastelu toteutettiin EU-rahoitteisessa Paikkatiedot Suomen tilastotuotannossa (GSFI) -projektissa (2/2023–1/2025).

2. Summary in English

This document, "Geospatial Information in the Business Data Production Process Finland – A Case Study", examines the vision and plan for the integration of geospatial information in different statistical processes, utilizing a centralized business data source and other data sources related to geospatial information. The work started by producing a description of the current state of the business data production process and continued by identifying development needs and describing a proposal for the target state. The integrity and comprehensiveness of the target state were further tested with expert interviews. In addition, it is shown how value chains can be applied in defining process-related development needs and how quality aspects can be integrated into geospatial production processes by applying value chains and value streams.

The background of the work has been the need to test the solutions of the global framework for the interoperability of statistics and geospatial information, GSGF, in national co-production solutions in order to enhance statistical processes. The implemented identification of development needs and description of the target state has relied on the GSGF Europe data set and in particular the GSGF Europe – requirements and recommendations document. The review was carried out in the EU-funded Integration of geospatial and statistical information in Finland (GSFI) project (February 2023–January 2025).

3. Lähtötilanne ja kehittämistarpeet

Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen johtamiseen käytetty päättelyprosessi uudistettiin Tilastokeskuksen yleisten tavoitteiden ja arkkitehtuurilinjausten mukaisesti vuosina 2020–2021. Liitteessä 1 on esitetty sijaintitietojen päättelyprosessin sekä prosessiin liittyvien roolien ja toimijoiden nykytilakuvaus. Liitteen kuva on kansallinen vastine GEOSTAT 4 -projektin yhteydessä tuotetulle arkkitehtuurikuvaukselle (EFGS and Eurostat, 2022) tarkentaen sijaintitietojen tuotantoon liittyviä rooleja ja toimijoita esimerkkiproessin osalta.

Jo nykytilassa päättelyprosessi toimii hyvänä käytännön esimerkkinä siitä, miten sijaintitietojen ja muiden tilastointitarkoitukseen käytettävien tietojen yhdistely voidaan toteuttaa tehokkaasti

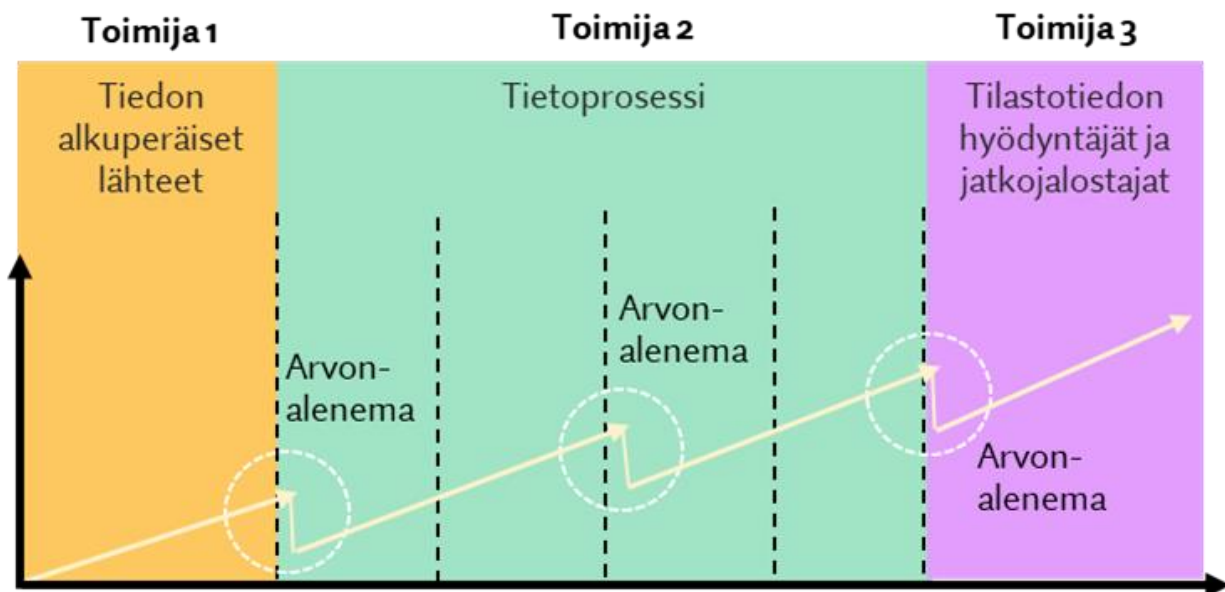
keskitettyjä sijaintitietoja hyödyntäen. Päättelyprosessiin voi tutustua tarkemmin projektissa koostetun nykytilakuvauksen avulla (EFGS and Eurostat, 2023).

Nykytilan kuvauksen yhteydessä kirjattiin ylös projektin tunnistamat nykyisen sijaintitietojen päättelyprosessin nykytila ja puutteet verrattuna GSGF Europe: vaatimukset ja suositukset - dokumenttiin. Suunnitteluvaiheessa näin tunnistetut kehittämistarpeet analysoitiin ja kuvauksia tarkennettiin. Tunnistetut kehittämistarpeista koostettuun listaukseen voi tutustua liitteessä 3.

Suurimmat kehittämistarpeet tunnistettiin liittyvän kansalliseen tietoekosysteemiin sijaintitietojen osalta. Etenkin kansallisen osoiterekisterin ja yritys-toimipaikka-palkansaaja linkin puuttuminen vie pohjaa sijaintitietojen tehokkaalta hyödyntämiseltä. Kansallisen osoiterekisterin puuttumisen myötä ainakin Tilastokeskuksessa osoitetietojen selvittelyyn ja korjaamisen käytetään paljon aikaa ja resursseja eikä osoitteet ole koneluettavia. Yritys/organisaatio-toimipaikka-palkansaaja linkin uupumisen myötä puuttuu yritysten ja muiden organisaatioiden taloustietoja. Siksi tietoja ei ole mahdollista alueellistaa muuten kuin turvautumalla kalliisiin suoriin tiedonkeruisiin, joilla saadaan kiinni vain suurimpien yritysten ja muiden organisaatioiden tarkempi sijaintitieto.

3.1. Nykytilan tarkastelu arvoketjujen avulla

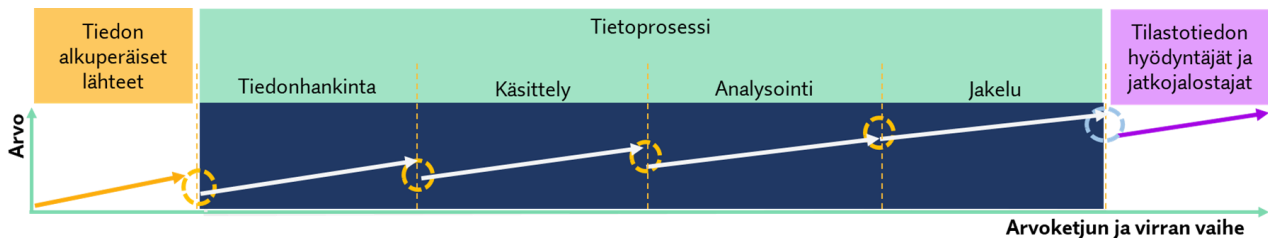
Yksi tapa kuvantaa ongelmaa on kuvata sijaintitietojen arvoketju (Kuva 1). Kuvasta näkyy, miten tiedon arvo laskee tiedon siirtyessä toimijalta toiselle. Kansallisen osoiterekisterin ja yritys-toimipaikka-palkansaaja linkin puuttumisesta seuraa arvonalentuma ja osoitetietojen joudutaan editoimaan käsin ja yritysten tarkempien sijaintitietojen osalta joudutaan turvautumaan suoriin tiedonkeruiseen kasvattaen tiedonantajarasitetta. Lisäksi puutteet osoitetiedoissa säteilevät joissain tapauksissa tilastollisiin tuotteisiin asti, mistä seuraa arvonalentuma tilastollisten tuotteiden hyödyntäjien ja jatkojalostajien näkökulmasta.



Kuva 1. Arvoketju, jossa tiedon arvo laskee tiedon siirtyessä toimijalta toiselle.

Sijaintitietojen arvoketju- ja tarkemman tason arvovirtakuvia voidaan soveltaa myös sijaintitietojen laadun mittaamiseen läpi tuotantoprosessien. Kuvantamalla eri toimijoiden välinen prosessi tiedon arvoketjuna ja määräämällä tiedon laadun mittauspisteet eri toimijoiden välisiin saumakohtiin (Kuva 2), voidaan tiedon laatua mitata eri toimijoiden näkökulmasta esimerkiksi tiedonlaatukriteereitä hyödyntäen (Kuva 3). Vastaavasti tiedon arvoketjun eri toimijoiden sisäisiä prosesseja voidaan kuvantaa tiedon arvovirtakuvan avulla ja määrittellä tiedon laadun mittauspisteitä tuotantoprosessien vaiheiden saumakohtiin (Kuva 3).

Tilastokeskuksen tietoprosessi on jäsenetty GSBPM-viiteprosessin (UNECE, 2019) mukaisesti. Viiteprosessin päävaiheiden eli tiedonhankinnan, käsittelyn, analyysin ja jakelun saumakohtat toimivat luontevina tiedon laadun mittauspisteinä. Tiedon arvoketjujen soveltamista on käsitelty tarkemmin erillisessä paperissa (Tammisto et al., 2024).



Kuva 2. Arvoketjun ja arvovirran vaiheet laadun mittauspisteinä.

4. Tavoitelaehdotuksen kuvaus

Työssä laaditussa ehdotuksessa Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessin tavoitetilasta, Tilastokeskuksen Yritys- ja toimipaikkarekisterin henkilöstö, osoite- ja sijaintitiedot perustuvat kansallisen tietoekosysteemin tietoihin ilman tarvetta erillisille sijaintitietojen suorakeruille tai manuaaliselle osoitetietojen korjaamiselle. Ehdotettu malli olisi sekä kustannustehokas että nopeuttaisi nykyistä tiedontuotantoa.

Tavoitelaehdotus kuulostaa kaukaiselta nykytilassa, missä kansallisen osoiterekisterin puuttuessa eri lähteiden kautta saatavat osoitetiedot voivat poiketa muodoltaan toisistaan, olla puutteellisia tai sisältää jopa suoranaisia virheitä. Myöskään organisaatioiden toiminnasta tai henkilöstöstä ei ole saatavissa sijaintitiedot sisältävää alueellisesta tietoa muutoin kuin suorien tiedonkeruiden avulla.

Tavoitelaehdotuksessa sijaintitiedot perustuvat Suomen ympäristökeskuksen vetämän Ryhti-hankkeen myötä rakennettavan RYHTI-tietojärjestelmän rakennusten sijaintitietoihin. Kattavat rakennusten osoite- ja sijaintitiedot mahdollistaisivat muiden tilastollisten yksikkötietojen alueellistamisen, kun osoitetietojen perusteella olisi johdettavissa osoitetietoa vastaava rakennus ja rakennusta vastaavat sijaintitiedot. Laadukkaat ja kattavat osoitetiedot ovatkin kriittisiä laadukkaiden sijaintitietojen johtamisen ja siten tavoitetilan toteutumisen kannalta.

Tavoitetilan toteutumiselle kansallinen osoiterekisteri onkin tunnistettu välttämättömäksi. Osoiterekisterin avulla on mahdollista standardoida Suomen osoitteisto ja varmistaa eri ilmoituskanavien kautta ilmoitettavien osoitetietojen oikeellisuus. Kun osoitetietojen oikeellisuus varmistetaan jo ilmoitusvaiheessa kansallisen osoiterekisterin avulla, ei osoitetietoja tarvitse validoida tai korjata ilmoittamisen jälkeen.

Laadukkaiden osoitteiden lisäksi tavoitetilan kannalta on tärkeä saada organisaatioiden alueelliset, sijaintitiedot sisältävät toimintatiedot kansallisen tietokesysteemin piiriin. Patentti- ja rekisterihallituksen vetämää parhaillaan Yritysten digitalous -hanketta. Siinä rakennetaan tietokesysteemi, jossa on tarkoitus jakaa myös Tilastokeskuksen yritys- ja toimipaikkarekisterin yritysten ja toimipaikkojen perustiedot ml. osoitetiedot. Ekosysteemissä tiedot ovat myös muiden valtionhallinnon toimijoiden mm. Verohallinnon Tulorekisteriyksikön käytettävissä. Tavoitetilassa Tulorekisteristä löytyy linkki tuloilmoitusten, palkansaajien, palkanmaksajien ja palkanmaksajien toimipaikkojen välille. Näiden linkitystietojen perusteella pystytään johtamaan esimerkiksi tieto siitä, missä organisaatioiden toiminta sijaitsee sekä organisaatioiden alueellinen työpanos. Tarkempien alueellisten työpanostietojen perusteella voidaan laskea mm. nykyistä tarkemmin kunnille jaettavan yhteisöveron osuudet.

Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessin näkökulmasta ehdotettu tavoitetila tarkoittaisi, että tarvittavat tiedot saadaan kansallisen tietokesysteemin eri tietojärjestelmistä, eikä suoria tiedonkeruita tarvittaisi:

- Kansallisesta osoiterekisteristä saadaan laadukkaat osoitetiedot
- Ryhti-hankkeessa rakennettavasta Ryhti-tietojärjestelmästä saadaan rakennusten osoite- ja sijaintitiedot.
- Yritysten ja toimipaikkojen perustiedot ml. osoitetiedot saadaan Yritysten digitalous -hankkeessa rakennettavasta talouden toimijoiden ekosysteemiä.
- Yritys- ja toimipaikkakohtaiset tulonsaaja- ja tulotiedot saadaan Verohallinnon Tulorekisteristä.

Liitteessä 2 kuvataan tavoitetilan mukainen Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessi, jossa ei ole tarvetta kalliille suorille tiedonkeruulle tai käsin tehtävälle osoitetietojen validointi- ja korjaustyölle. Nykytilaan verrattuna tavoitetilassa poisjäävät prosessivaiheet on liitteen kuvassa korostettu valkoisella.

5. Tavoitetilaehdotuksen eheyden ja kattavuuden tarkastelu

Työn lopuksi edellä kuvattua tapaustutkimusta ja tavoitetilaehdotusta tarkasteltiin eheyden ja kattavuuden näkökulmista peilaamalla näitä GSGF Europe viitearkkitehtuuriin, kansalliseen paikkatietoinfrastruktuuriin sekä Tilastokeskuksen viiteprosessiin ja tavoitetilan tietoprosessiin.

- Tapaustutkimuksessa tarkastellun Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessin nyky- ja tavoitetilasta tuotettiin GSGF Europe viitearkkitehtuuriin perustuvat toiminta-arkkitehtuurikuvat Archimate kuvauskielellä (ks. liitteet 1 ja 2). Tavoitetilan toimintaehdotuksen katsottiin olevan eheä GSGF Europe viitearkkitehtuurin kanssa.
- Kansallinen paikkatietoinfrastruktuuri on pyritty huomioimaan tavoitetilaehdotusta suunniteltaessa osana GSFI-projektia.
- Tapaustutkimuksena käytetty Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessi on GSBPM viiteprosessin mukainen ja kehittämisehdotusten myötä etenkin

yhteistyö kansallisten yhteistyötahojen kanssa tiivistyisi. Hyödyksi arvioidaan kansallisesti 2–5 htv:n vuosisäästö.

- Tapaustutkimuksen prosessi on nykytilassa Tilastokeskuksen tavoitetilan tietoprosessin mukainen tavoitetilan tieto. Kehittämisehdotusten myötä päättelyprosessi tehostuu etenkin manuaalisten käsittelytarpeiden sekä suorien tiedonkeruutarpeiden osalta. Hyödyksi arvioidaan Tilastokeskuksessa 1–3 htv:n vuosisäästö.

6. Lähteet

EFGS and Eurostat, 2022. GSGF Europe: Reference Architecture. EU, 945503 - 2019-FI-GEOSTAT4.
Linkki: [GSGF Europe: reference architecture, EU, https://www.efgs.info/gsgf-europe-geostat-information-service/geostat-information-service/methodology/](https://www.efgs.info/gsgf-europe-geostat-information-service/geostat-information-service/methodology/)

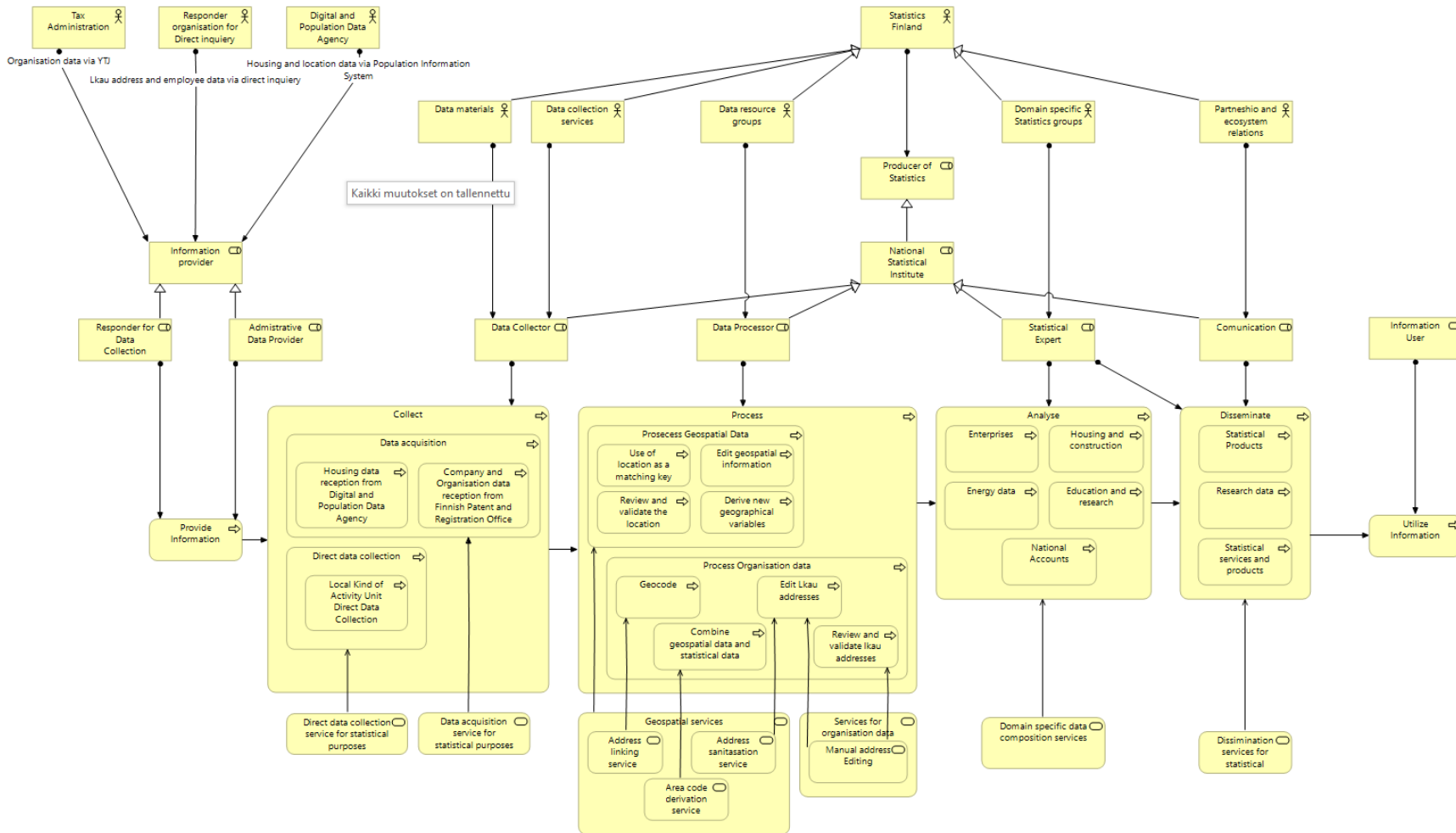
EFGS and Eurostat, 2023. Tilastokeskuksen yritys- ja toimipaikkarekisterin sijaintitietojen päättelyprosessin kuvaus ja kehittämistarpeiden tunnistaminen – tapaustutkimus. EU, 101112903 - 2022-FI-GEOS-GSFI.

Tammisto, R., Lowndes, G., Haakana, M., 2024. Practical approach in developing the integration of statistics and geospatial information through value chains and data quality. Estoril, Portugal, 4. – 7. June 2024.

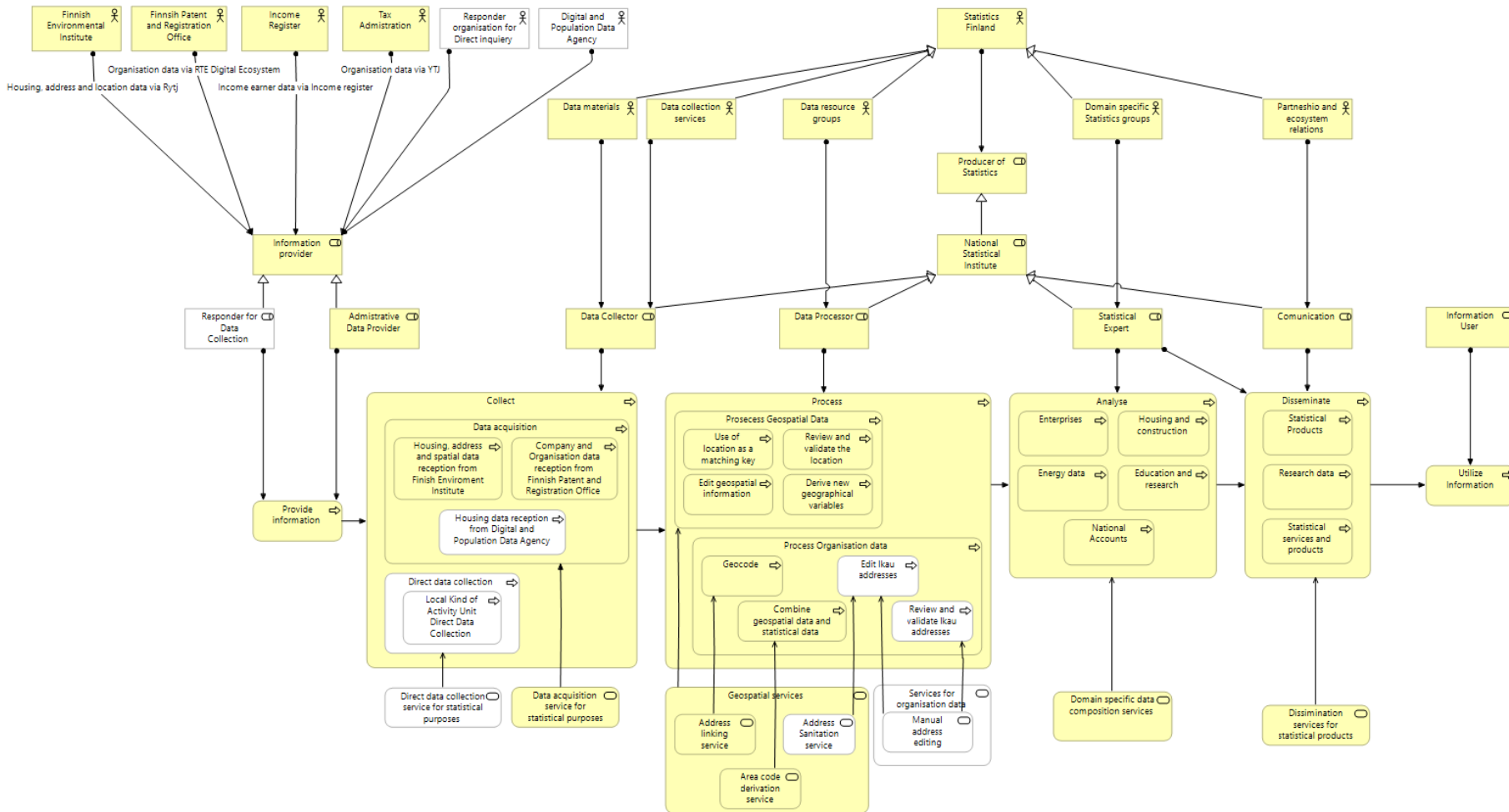
UNECE, 2019. Generic Statistical Business Process Model GSBPM. Version 5.1.

Linkki: [GSBPM prosessimalli, UNECE, https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1](https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1)

Liite 1 - Sijaintitietojen päättelyprosessi sekä prosessiin liittyvät roolit ja toimijat, nykytila



Liite 2 - Sijaintitietojen päättelyprosessi sekä prosessiin liittyvät roolit ja toimijat, tavoitetila



Liite 3 – Toimenpide-ehdotukset

Tiedon käytön vaihe	Nykytila	Kehittämistarve	GSGF-periaate ja vaatimus, jota toteuttaa	Ehdotettu toimenpide Suomessa	Ehdotetut Tilasto-keskuksen ulkoiset sidosryhmät	Hyödyn mittaaminen kansallisen laatukehikon ¹ laatukriteerein
Tiedonhankinta ja käsittely	Ei kansallista osoiterekisteriä. Kukin organisaatio rakentaa omat ratkaisunsa ja tarvittavat osoiterekisterinsä. Aineistojen ja tiedonkeruiden kautta tulevilla tiedoilla virheellisesti syötettyjä ja yhteentoimimattomia osoitteita.	Kansallinen osoiterekisteri.	1.1 Käytä kansallisen paikkatieto-infrastruktuurin tietoja 1.4 Edistä sijaintitietojen sisällymistä tilastoinnissa hyödynnettäviin kansallisiin rekistereihin	Tuota suunnitelma kansallisen osoiterekisterin toteuttamiselle.	MML, Syke	Virheettömyys Johdonmukaisuus Suositusten mukaisuus

¹ Tiedon laatukehikko, 2022. Tiedon laatukehikko -osahanke (7/2020–12/2022), Tilastokeskus, Tiedon hyödyntämisen ja avaamisen hanke (VN/5386/2020), VM.
Linkki: stat.fi/tiedonlaatu

Tiedonhankinta ja käsittely	Yritys voi syöttää PRH:n YTJ-palveluun osoitteen vapaaseen tekstikenttään ja YTJ sisältää virheellisiä osoitteita.	PRH:n YTJ-palvelun kehittäminen osoitetietojen syöttämisen osalta siten, että palveluun voi syöttää vain kansallisen osoiterekisterin mukaisia osoitetietoja.	1.4 Edistä sijaintitietojen sisältymistä tilastoinnissa hyödynnettäviin kansallisiin rekistereihin 2.4 Varmista hallinnollisten ja muiden tietojen keruussa ja vastaanotossa sijaintitietojen oikeellisuus	Asian esiin tuominen PRH:n kanssa. Toteutus riippuvainen kansallisesta osoiterekisteristä.	PRH	Johdonmukaisuus Alkuperäisyys Ajantasaisuus
Tiedonhankinta ja käsittely	Toimipaikkatason tiedot suorien tiedonkeruiden varassa, koska toimipaikkatason tietoa ei ole saatavilla hallinnollisissa aineistoissa.	Toimipaikkojen osoitetietojen lisääminen osaksi yrityksen digitaalista identiteettiä Yrityksen digitalous -hankkeessa.	1.4 Edistä sijaintitietojen sisältymistä tilastoinnissa hyödynnettäviin kansallisiin rekistereihin	Seuraa ja vaikuta aktiivisesti asian etenemiseen Yritysten digitaloushankkeessa	PRH\YD	Johdonmukaisuus Alkuperäisyys Ajantasaisuus Kattavuus
Tiedonhankinta ja käsittely	Tiedonantaja voi syöttää Tilastokeskuksen suorien	Yritystiedonkeruiden kehittäminen osoitetietojen syöttämisen osalta	2.4 Varmista hallinnollisten ja muiden tietojen keruussa ja	Vie eteenpäin kehittämissuorien suorien tiedonkeruiden	-	Suosittelun mukaisuus

	tiedonkeruiden lomakkeille osoitteen vapaaseen tekstikenttään ja tiedonkeruiden kautta tulee virheellisiä osoitteita.	siten, että tiedonkeruulomakkeelle on mahdollista syöttää vain valideja osoitteita esimerkiksi dynaamisen tekstikentän kautta.	vastaanotossa sijaintitietojen oikeellisuus	lomakkeiden osoitekenttien kehittämiseksi siten, että jatkossa osoitekentät eivät ole vapaita tekstikenttiä vaan dynaamisia alavetovalikoita tekstikentän kera, jolloin tekstikenttää voisi syöttää vain osoiterekisterin mukaisia tietoja.		
Käsittely	Tulorekisterissä vain vähän käyttökelpoista toimipaikkatason tietoa.	Toimipaikkatiedon standardointi tulorekisterissä ja toimipaikkatiedon tuominen tulorekisteriin yrityksen digitaalisen identiteetin kautta.	1.4 Edistä sijaintitietojen sisältymistä tilastoinnissa hyödynnettäviin kansallisiin rekistereihin 2.5 Määrittele yhteiset paikkatiedon laadun määritelmät, jotka huomioivat alueellisen ja ajallisen johdonmukaisuuden	Asian esiin tuominen PRH:n ja Tulorekisterin kanssa. Toteutus riippuvainen kansallisesta osoiterekisteristä.	PRH\YD, Vero \Tulorekisteri	Johdonmukaisuus Alkuperäisyys Virheettömyys Ajantasaisuus Kattavuus Suositusten mukaisuus

Käsittely	Tilastokeskuksen Sijaintitieto-järjestelmää tai tämän palveluiden toimintaa ei tunneta laajasti Tilastokeskuksessa. Sijaintitietojärjestelmän toimintalogiikka ja laadunvarmistustyö jäävät järjestelmän rajapintapalveluiden käyttäjille vieraksi.	Sijaintitietojärjestelmän tuntemuksen parantaminen	2.3 Varmista geokoodauksen tulosten laatu 5.4 Helpota tiedon löydettävyyttä paremman luetteloinnin ja ohjeistuksen avulla	Esitellään Sijaintitietojärjestelmän sisäistä toimintaa ja prosesseja Tilastokeskuksen palvelualueiden ja osastojen infoissa.	-	Johdonmukaisuus Alkuperäisyys Metatietojen ymmärrettävyys Suositusten mukaisuus
Käsittely	Sijaintitietojen laadun määritelmiä ei ole kirjattu ylös.	Sijaintitietojen laadun määritelmien kirjaus	2.5 Määrittele yhteiset paikkatiedon laadun määritelmät, jotka huomioivat alueellisen ja ajallisen johdonmukaisuuden	Lisää sijaintitietojärjestelmän tehtävälisälle sijaintitietojen laatuvaatimusten kuvaus yhteistyössä järjestelmän käyttäjäkunnan kanssa.	-	Suosituksen mukaisuus Metatietojen ymmärrettävyys
Jakelu	Yritysrekisterin tietopalvelussa ei ruututietoa tai prt-tietoa.	Yritystietopalveluiden tietosisällön kasvattaminen mm. prt- ja ruututietojen osalta.	Periaate 3: Yhteiset aluejaot tilastojen jakelussa. 3.2 Vakiinnuta olemassa olevien	Lisää YTY-tietojärjestelmän kehitysjonoon tehtävä ruututietojen lisäämiseksi YTYyn	-	Kattavuus Koneluettavuus Suositusten mukaisuus

			tilastoruudukoiden käyttö ja tutki uusia kehittyviä maailmanlaajuisia tilastoruudukkoratkaisuja	sekä yritysrekisterin tietopalvelun tehtävälliställe prt- ja ruututietojen lisäämiseksi osaksi tietopalvelua		
Jakelu	Tilastokeskukselta puuttuu avoin rajapinta tilastolain 18 §:n mukaisiin julkisiin yritysrekisterin tietoihin näiden laajempaa hyödyntämistä varten. Tietojen jakelu tapahtuu maksulliset Yritysrekisterin yksikkötietopalvelu kautta.	Rajapinta tilastolain 18 §:n mukaisiin julkisiin yritysrekisterin tietoihin.	4.3 Julkaise tiedot kerran ja mahdollista niiden uudelleenkäyttö alkuperäisestä lähteestä	Kehittämisehdotus Toimi yritysrekisterin yksikkötietopalvelun rajapinnan toteutusta varten	-	Koneluettavuus Oikea-aikaisuus Johdonmukaisuus