

Käytännön lähestymistapa tilastojen ja paikkatiedon integroinnin kehittämiseen arvoketjujen ja tiedon laadun avulla

Title: Käytännön lähestymistapa tilastojen ja paikkatiedon integroinnin kehittämiseen arvoketjujen ja tiedon laadun avulla

Title in English: Practical approach in developing the integration of statistics and geospatial information through value chains and data quality

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. Arvoketjut.....	4
4. Tiedon laatukehikko ja laatukriteerit.....	4
5. Arvoketju ja laatukehikko tietoprosessissa	6
5.1. Esimerkki: Tilastokeskuksen yritysrekisterin (yritystietojärjestelmä) uudistus	8
6. Tiedon laadun hallinta tietoprosessissa	10
7. Lähteet.....	11

1. Yhteenveto

Tieto liikkuu lähteestään eri kohteisiin määriteltyjen tietoprosessien mukaisesti ja julkishallinnossa usein myös organisaatorajat ylittäen. Näiden tietovirtojen eri vaiheissa käyttöönotettava tieto ei välttämättä kaikilta osiltaan vastaa tarvetta, vaan tietoa joudutaan korjaamaan, täydentämään tai muuttamaan. Tietovirrassa voidaan havaita näitä arvon alenemia sekä yksittäisen organisaation omissa prosesseissa että tiedon virratessa organisaatiolta toiselle. Mitä paremmin eri vaiheiden tiedon käyttäjien tarpeet pystytään huomioimaan jo tiedon alkulähteillä, sitä vähemmän syntyy hukkaa myöhemmissä vaiheissa.

Tässä työssä syntyi ajatus kansallisen tiedon laatukehikon hyödyntämisestä arvoketjutarkastelussa ja tarkemmin arvonalenemiin pureutuen. Arvoketjunäkökulman avulla voidaan tiedon virtaamisen taitekohdissa ymmärtää, kuvata ja mitata tiedon laatua. Tiedon laatukehikon 11 laatuksiteeriä antavat systemaattisen ja yhteisen rakenteen tarkastella tiedon laatua ketjun taitekohdissa. Tämä tilanneymmärrys voidaan jakaa muiden toimijoiden tai muiden prosessivaiheen toimijoiden kanssa ja onnistuneimmillaan sopia tietyistä muutoksista, joilla arvonalenemia voidaan poistaa tai saada vähemmän virtausta estäviksi. Tiedon virtaus tehostuu, eli prosessit, joissa tieto virtaa, suoraviivaistuvat ja saadaan säästöjä niin ajassa kuin resurssien käytössä.

Tämä tarkastelu toteutettiin EU-rahoitteisessa Paikkatiedot Suomen tilastotuotannossa (GSFI) -projektissa (2/2023–1/2025). Esimerkissä tarkasteltiin Tilastokeskuksen Yritysrekisteri -tapaustutkimuksen kautta tiedon virtaamista ja sen arvossa tapahtuvaa muutosta.

2. Summary

Data flows from its source to various destinations according to defined information processes. Within public administration these flows often also transcend organizational boundaries. However, the information adopted at different stages of these data flows may not necessarily meet the intended purpose, requiring corrections or modifications. Within the information flow, value erosion can be observed both within an individual organization's processes and as information moves from one organization to another. The better the needs of information users of different stages can be taken into account already at the initial sources of information, the less waste is created in later stages.

This work led to the idea of leveraging the national information quality framework for value chain analysis, with a specific focus on addressing value erosion. The value chain perspective allows us to understand, describe, and measure the quality of information at critical points in its flow. The 11 quality criteria of the national information quality framework provide a systematic and shared structure for assessing information quality at these chain junctures. This situational understanding can be shared with other stakeholders or actors in different process stages. Ideally, this leads to agreements on specific changes to eliminate value erosion or reduce flow impediments. Information flow becomes more efficient, streamlining processes where information travels and resulting in savings in both time and resource utilization.

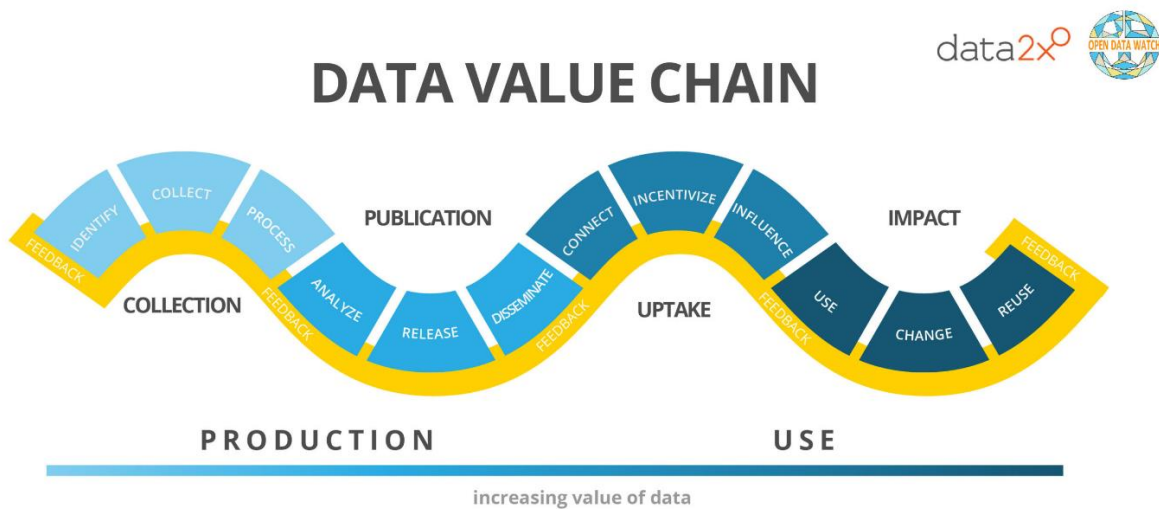
This analysis was conducted as part of the EU-funded Geospatial Information Finland statistical production project (GSFI) (2/2023–1/2025). In the example, the flow of information and changes in its value are examined through a case study of the Business Register of Statistics Finland. More about the study is available in the paper by Tammisto, Lowndes and Haakana (2024).

3. Arvoketjut

Yksi merkittävä haaste tilasto- ja paikkatietotietojen sujuvalle integroinnille ja yhteentoimivuudelle on saumattoman tiedonkulun varmistaminen eri toimijoiden välillä ja hallinnonalarajat ylittäen. Yhteentoimivuuden saavuttaminen edellyttää yhteistyötä niin organisaatioiden välillä kuin niiden sisällä sekä yhteistä ymmärrystä yksittäisistä kehittämistoimenpiteistä ja niiden eduista sekä sopimusta toteutuksesta.

Arvoketjunäkökulma auttaa kuvaamaan ja ymmärtämään tiedon virtauksen yhteydessä tapahtuvaa tiedon arvon kasvua (Kuva 1). Perinteisesti tieto tuotetaan tietyltä näkökulmalta ja tarjotaan käyttäjille ennalta määrättyjen tai muuten määriteltyjen periaatteiden mukaisesti. Alkuperäinen tiedon tuottaja toimii sovittujen käytäntöjen mukaisesti tai esimerkiksi lakisääteisten velvoitteiden pohjalta. Nykyaikaisissa tietoekosysteemeissä toimitaan kuitenkin monimutkaisemmin. Haasteena on, että kaikki tiedon käyttäjien tarpeet eivät välttämättä ole tiedon tuottajien tiedossa – tai ei ole keinoja kiinnittää niihin huomiota. Tieto virtaa silti, mutta tiedon laatu kärsii, mikä aiheuttaa turhaa ja päällekkäistä työtä kokonaisuuden kannalta.

Tässä työssä on arvoketjunäkökulmaa sovellettu tilastotuotannon tarvitsemien ja tuottamien tietojen virtauksen tarkasteluun tilastoprosessissa. Tarkastelun keskiössä on siis tilastoprosessista vastaava taho. Muita toimijoita arvoketjussa ovat tiedon alkuperäiset tuottajat ja tiedon jalostajat ennen tilastoprosessin käynnistymistä sekä tiedon hyödyntäjät ja jatkojalostajat tilastoprosessin jälkeen.



Kuva 1. Tiedon arvoketju kuvaa tiedon luomis- ja käyttöprosessia ensimmäisestä tiedontarpeen tunnistamisesta sen lopulliseen käyttöön ja mahdolliseen uudelleenkäyttöön. Arvoketju pilkkoutuu useisiin eri vaiheisiin ja koko prosessin ajan, arvoketjun päästä toiseen ja takaisin, tuottajien ja sidosryhmien välillä tulisi olla jatkuvaa palautetta. (Open Data Watch, 2018)

4. Tiedon laatukehikko ja laatukriteerit

Tiedon tehokkaampaa hyödyntämistä varten tarvitaan panostusta tiedon käyttäjän tarpeisiin soveltuvuuden eli tiedon laadun arviointiin. Tähän tarpeeseen on luotu kansallinen tiedon laatukehikko (2022). Laatukehikko koostuu tiedon laatua kuvaavista laatukriteereistä ja näille konkreettiset arvot tuottavista mittareista (Kuva 2). Laatukriteerien ja mittariston käyttöä tukee soveltamisohje.

Miten tieto kuvaa todellisuutta?



Miten tieto on kuvattu?



Miten tietoa voi käyttää?



Kuva 2. Tiedon laatukehikko (2022) on luotu tietoaaineistojen laadun arvioinnin tueksi. Laatukehikon ytimessä ovat tiedon laadun eri näkökulmia sanoittavat 11 laatuksiteeriä.

Tiedon laatuksivauksset antavat tietoa tietoaaineistoista sekä tietoaaineiston omistajalle että tiedon käytöstä kiinnostuneille. Kuvausten tuottaminen kannustaa myös yhtenäisten käsitteiden käyttöön ja siten edistää osaltaan tiedon yhteistoimivuutta. Yhtenäinen sanoitus tiedon laadun näkökulmille helpottaa laatuun liittyviä keskusteluja. Yhtenäisten kuvausten avulla tiedon käyttäjä pystyy vertailemaan eri tietoaaineistojen laatua ja valitsemaan käyttötarkoitukseensa soveltuvimmat aineistot. Toisaalta tiedon tuottaja pystyy tarpeet tunnistaessaan kehittämään tietoaaineistojensa laatua. Alla taulukossa 1 on esimerkki laatuksivauksen tuella käydyn tarkastelun tuloksista. Esimerkissä tarkastelun kohteena on kappaleessa 5.1 esitelty Tilastokeskuksen yritysrekisterin uudistus.

Taulukko 1. Esimerkki tiedon laatukehikkoon nojaavasta tarkastelusta Tilastokeskuksen yritysrekisterin uudistuksen kohdalla.

Kriteeri	Havainnot
Virheettömyys	Tietopalvelun kautta tullut palaute sijaintitietojen laadusta on vähentynyt.
Kattavuus	Lähes kaikille osoitteellisille toimipaikoille pystytään johtamaan sijaintitieto. Osoitetieto löytyi 94,3 prosentilta EBS-asetuksen mukaiselta toimipaikalta tilastovuonna 2021.
Ajantasaisuus	Sijaintitiedot perustuvat DVV viikkoaineistoon.
Johdonmukaisuus	Pysyvää rakennustunnusta vastaa aina johdonmukainen sijainti.
Suosituksen mukaisuus	Päätelyprosessi on sijaintitietoasiantuntijoiden kanssa kehitetty ja prosessin tuottamat tiedot ovat suositusten mukaisia.
Alkuperäisyys	Prosessissa poimitaan sijaintitietovarannosta pysyvää rakennustunnusta vastaava sijaintitieto.
Metatietojen ymmärrettävyys	Sijaintitietoihin liittyvä metatieto keskitetty yhteen paikkaan.
Käyttöoikeudet	Käyttöoikeudet prosessiin ja kaikkiin tietoihin ovat Tilastokeskuksen käyttövaltuushallinnon mukaisesti hoidettu.
Koneluettavuus	Prosessi ja sen tuottamat sijaintitiedot on kutsuttavissa https- ja sql-rajapintojen kautta.
Oikea-aikaisuus	Prosessi tuottaa oikea-aikaiset sijaintitiedot tilastojen viiveet ja kuntaliitokset huomioiden.

5. Arvoketju ja laatukehikko tietoprosessissa

Tilastotuotanto perustuu lukuisten tietojen yhdistämiseen ja käsittelyyn. Tässä merkityksellistä on, mitä saatava tieto sisältää ja miten se on tarjolla tiedon tuottajan toimesta. Perinteisesti tilastojen tuotanto on perustunut kyselyihin eli suoriin tiedonkeruuihin, jotka on kohdennettu tietyille otokselle, esim. vapaa-ajan tutkimus, tai kaikille tiedonantajille, esim. oppilaitostiedonkeruu. Tällöin tilastotuotannon tietotarve määrittelee suoran tiedonkeruun sisällön ja tietoja saadaan kohdennetusti vastaamaan oletettua tai selvitettyä tietotarvetta.

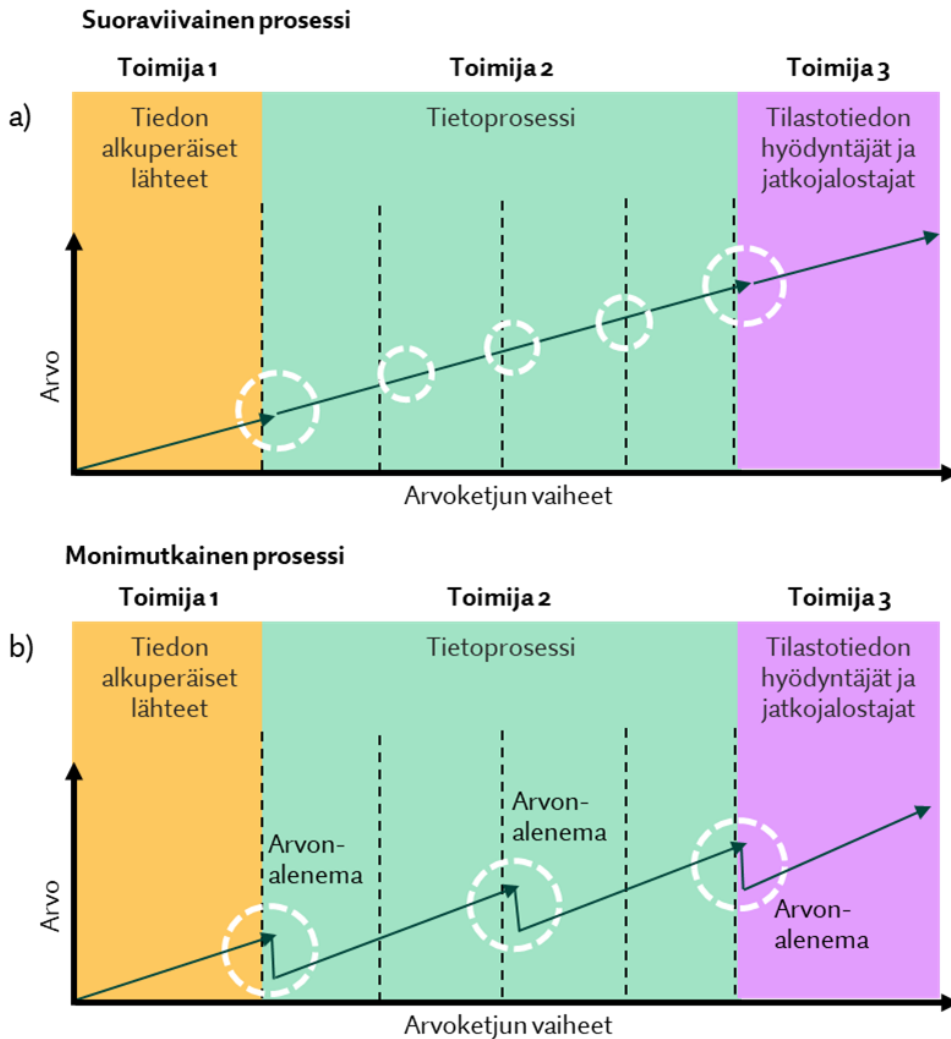
Suomessa tilastotuotanto nojaa kuitenkin jo pääosin olemassa oleviin laajoihin kansallisten tietovarantojen tietoihin eli ns. hallinnollisiin aineistoihin. Mm. vakiintuneet, luotettavat toimijat ja lainsäädäntö ovat mahdollistaneet muutoksen. Tiedoista koostetaan tilastotuotannon tarvetta vastaava tietojen kokonaisuus, jota tilastotuotannossa edelleen käsitellään ja yhdistellään. Tavoitetilassa tilastotuotannon käyttöön saamat tiedot jatkavat virtaamistaan tietoprosessissa ilman muutoksia. Tieto on joustavasti saatavilla käyttöä varten, tiedon sisältö vastaa tietotarvetta ja tieto on muutenkin laadukasta eikä siihen kohdistu korjaavia toimenpiteitä.

Käytännössä tieto ei aina virtaa tavoitetilan kaltaisella tavalla. Tieto voi olla teknisesti tai käyttöoikeuksien takia hankalasti saatavilla. Joudutaan tekemään erillisiä toimenpiteitä tiedon saamiseksi, tai kenties sopimaan erikseen yksityiskohdista tietojen saamiseksi toistuvasti. Esimerkiksi laissäädäntö edellyttää tarvekuvausta aineistosopimuksiin. Tiedon rakenne ei välttämättä vastaa hyödyntämisessä tarvittavaa rakennetta ja vastaanottava osapuoli aloittaa oman prosessinsa tietojen käsittelyllä, jotta rakenne saadaan jatkokäsittelyä varten oikeanlaiseksi. Hankalin tilanne on, kun saatavassa tiedossa on suoranaisia virheitä, jotka tulevat ilmi, kun tietoja yhdistetään tilastoprosessissa muihin tietoihin.

Kaikki edellä mainitut ongelmat ovat harvoin laiminlyönneistä johtuvia. Voi olla, että tiedon tuottajalla on vankka usko tarjoamiensa tietojen hyvään laatuun. Tiedon tuottaja ei kuitenkaan tiedä kaikkia tiedon käyttötarkoituksia tai tiedon käyttäjien erityistarpeita. Tiedon laatu eli soveltuvuus käyttäjän tarkoituksiin on siis oletettua alhaisempi. On otettava myös huomioon, että jokainen toimija pyrkii varmistamaan ensi sijassa omaan toimintaansa liittyvät keskeiset tiedot ja niiden laadun. Usein on myös niin, että tiedon tuottajalle ei välity tieto vastaanottajan prosesseissa tehdyistä tietoon kohdistuvista korjaavista toimenpiteistä. Esimerkiksi tilastolaki ei mahdollista yksittäisiin tietoihin liittyvistä korjauksista tai huomioista viestimisen tietojen toimittajalle. Keskustelu yleisistä huomioista, virheistä ja tietojen laadun parantamisesta on kuitenkin mahdollista ja suositeltavaa.

Arvoketjukuvauksen avulla on mahdollista luoda yhteinen näkemys tiedon laadusta ja virtaamisesta sekä saada ymmärrys erityisesti ketjun taitekohtien, toimijalta toiselle tai prosessin vaiheesta toiseen siirtymisen tilanteista (Kuva 3). Yhteinen kieli arvoketjujen toimivuuden tai toimimattomuuden tarkastelulle löytyy tiedon laatukehikosta. Laatukehikon 11 näkökulmaa laatuun antavat rakennetta keskustelussa siitä, mitkä ominaisuudet tiedossa ja sen tarjoamisessa on hyvin tai mitä pitäisi kehittää.

Arvoketjuja kuvaamalla saadaan tiedon virtaukseen laaja näkemys – tiedon alkuperäiselle tuottajalle tulee mahdollisuus ymmärtää tiedon etenemistä jatkojalostajien ja hyödyntäjien käsissä. Teknisesti pienilläkin muutoksilla voidaan tukea ja tehostaa koko tiedon hyödyntämisen ketjua. Koko ketjun näkeminen johdattaa todennäköisesti usean eri hallinnonalan läpi ja osoittaa, että erillisissä silloissa toimiminen rajoittaa tiedon virtaamista ja siten tiedon arvoketjun optimaalista kulkua ja aiheuttaa myös lisäkustannuksia.



Kuva 3. Tiedon arvoketju a) suoraviivaisessa ja b) monimutkaisessa prosessissa.

5.1. Esimerkki: Tilastokeskuksen yritysrekisterin (yritystietojärjestelmä) uudistus

Kansalliset tilastolaitokset käyttävät helposti saatavilla olevia hallinnollisia tietolähteitä tilastojen tuotannossa ja turvautuvat kalliisiin suoriin kyselyihin, kun muut lähteet eivät riitä. Riippumatta lähdeaineistosta, mittaaminen kohdistuu perinteisesti valmiiden tilastotuotteiden laatuun, jolloin kaikki saattaa näyttää hyvältä, mutta ei paljasta koko totuutta (Taulukko 2).

Taulukko 2. Esimerkki arvoketjutarkastelun esille nostamista haasteista verrattuna valmiiden tilastotuotteiden laadun tarkasteluun. Tarkastelun kohteena Tilastokeskuksen yritysrekisterin geokoodausprosessi.

Näkymä valmiin tilastotuotteen laatutarkastelussa	Arvoketjutarkastelun havainnot
Kaikki näyttää olevan kunnossa. Prosessi hyödyntää kattavaa hallinnollista tietolähdettä	<p>Kokonaiskuvassa nähdään, että hallinnollinen tietolähde ei aivan vastaa tilastotarpeita ja lisäkyselyjä tarvitaan ja osoitetiedot eivät ole aivan riittäviä tilastotarpeisiin.</p> <p>Puutteet tietolähteissä aiheuttavat tehottomuutta ja manuaalisia muokkaustarpeita tilastoprosessissa.</p> <p>Lisää puutteita hallinnollisissa tietolähteissä, esimerkin osoitetietojen heikko laatu aiheuttaa tehottomuutta.</p> <p>Joissakin tapauksissa tehottomuudet ulottuivat aina tilastotuotteisiin, erityisesti tutkimuksessa käytettäviin tietoihin ja muihin maksettuihin tarkan tason tilastotuotteisiin.</p>

Tilastokeskuksen yritysrekisterin äskettäin (2020–2021) uudistettu prosessi osoittautui erinomaiseksi esimerkiksi tarkastella geokoodausprosessia arvoketjujen avulla (Kuva 4). Kun esimerkkiprosessia kuvattiin arvoketjuna, ilmeni, että hallinnollisten tietolähteiden kohdalla seurasi arvon laskua, kun Tilastokeskus vastaanotti aineistot. Tämä arvon aleneminen kompensoidaan kalliilla suorilla tiedusteluilla ja erilaisilla muokkausprosesseilla. Lisäksi arvon alenemisesta aiheutuva työ jää helposti organisaation sisäiseksi, jolloin arvon alenemisesta aiheutuvat seurannaisvaikutukset moninkertaistuvat muiden organisaatioiden tehdessä vastaavia korjaustoimenpiteitä. Esimerkkinä käytetty yritystietojärjestelmän tarkempi kuvaus ja sen geokoodausprosessin tarkastelu on esitelty erillisessä projektin raportissa (EFGS and Eurostat, 2023 ja 2024).

Tiedon alkuperäiset lähteet	Tietoprosessi				Tilastotiedon hyödyntäjät ja jatkojalostajat
Saatavilla olevien tietojen laadun arviointi tietotarpeen näkökulmasta	Vastaanotettavien tietojen laadun mittaaminen	Yksikköaineiston laadun mittaaminen ennen tietovarantolatausta	Tietovarantojen laadun mittaaminen ennen aihealueiden aineistojen koontia	Valmiiden tietotuotteiden ja tilastojen laadun mittaaminen	Tietotuotteiden ja tilastojen laadun arviointi tiedon hyödyntäjien näkökulmasta
Paikkatietoaineistot eivät vastaa tilastointitarpeita. Niistä puuttuu standardoidut osoitteet ja toimipaikkojen tiedot. → Kansallinen osoiterekisteri ja toimipaikat YD:n tietoekosysteemiin ja täältä tulorekisteriin.	Kaikki aineistot tiedonhankinnan läpi, mutta hallinnollisista aineistoista puuttuu toimipaikkojen tiedot. → Toimipaikkatiedot suorien tiedonkeruiden varassa. → Tiedonkeruurasite pienempi, yhteysosoitteiden laatu parempi	Kaikki lähdeaineistot perusyksikötasoisissa tietovarannoissa, mutta osoitetiedoissa paljon puutteita. → Paljon manuaalista ja koneellista korjaustyötä. → Vähemmän manuaalista osoitteiden korjaustyötä. Rekisteripohjaiset työsuhdetiedot työssäkäynti-tilastolle.	Tilastojen ja KTMT:n paikkatiedot pohjautuvat keskitettyihin sijaintitietovarannon paikkatietoihin.	Tilastojen laaturaporttien perusteella kaikki hyvin. → Laadukkaammat tilastot paikkatietojen osalta.	Tietopalvelutuotteiden ja tutkijapalveluiden asiakaspalautteen perusteella kaikki hyvin. → Kattavimmat paikka- ja aluetiedot tutkijoille ja tietopalveluasiakkaille.

Kuva 4. Käytännön esimerkki tarkastella geokoodausprosessia arvoketjujen avulla. Tarkastelun kohteena Tilastokeskuksen yritysrekisterin prosessi.

6. Tiedon laadun hallinta tietoprosessissa

Tavoitteena on ymmärtää, miten luoda edellytyksiä edellä mainittuun tiedon arvoketjun ymmärtämiseen ja hallintaan tietoprosessissa ja laajemmin tietoekosysteemeissä. Tässä vaiheessa on tunnistettu kolme keskeistä osa-aluetta, joiden avulla tiedon laadun hallintaan voidaan tarttua arvoketjutarkastelun avulla (Kuva 5). Nämä hallinnan alueet ovat:

- Ohjaus, tarpeiden määrittely ja sopiminen: Tarvitaan vahva sitoutuminen prosessien eri toimijoiden välille laadun ymmärtämiseen ja tiedon virtaamisen tunnistamiseen ja tehostamiseen. Tietoprosessi tulee nähdä kokonaisuutena ja ymmärtää eri toimijoiden roolit siinä. Prosessin hallinnan tulee olla sitoutunut sen kaikilla tasoilla, ohjauksesta käytännön toimiin, jotta käytännön yksityiskohdat voidaan muuntaa ohjaavan tason sopimuksiksi.
- Taitekohtien ja mittauspisteiden seuranta ja palaute: Prosessit elävät jatkuvassa muutoksessa. Tiedot, teknologiat, vastuut muuttuvat, joten tiedon virtaamisen monitorointia tarvitaan, jotta ymmärretään mitata oikeita asioita muutoksista huolimatta. On myös tärkeää, että seuranta tuottaa jatkuvaa palautetta toimijoille tiedon ylävirrassa.
- Laadun arvioinnin automatisointi prosesseissa: Laadun mittaamisesta on tarpeen tehdä vakiintunutta toimintaa. Käytännössä vain automatisoinnilla voidaan taata, että laadun mittaaminen on tiedon hyödyntämisen prosessissa mukana vakiintuneena osana prosessia.



Kuva 5. Tiedon arvoketjunäkökulman edellyttämät hallinnan alueet.

Yleisesti tiedon laatukehikon käyttöönoton hyötyjä ja edellytyksiä on kuvattu kattavasti Tiedon laatukehikon käyttöönoton suunnittelu -soveltamisohjeessa (Tiedon laatukehikko, 2022).

7. Lähteet

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

EFGS and Eurostat, 2024. Paikkatieto Tilastokeskuksen yritystietojärjestelmässä - tapaustutkimus. GSFI-projekti. EU, 101112903 - 2022-FI-GEOS-GSFI.

Open Data Watch, 2018. The Data Value Chain: Moving from Production to Impact. Prepared for Data2X by Open Data Watch. [Linkki Open Data Watch -sivustolle: The Data Value Chain.](#)

Tammisto, Lowndes and Haakana, 2024. Practical approach in developing the integration of statistics and geospatial information through value chains and data quality. European Conference on Quality in Official Statistics 2024, Estoril, Portugal 4.-7.6.2024.

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