

Tiedon laatu

Johdanto

Tiedon määrä kasvaa eksponentiaalisesti yhteiskunnassa. Samalla nousevat esille kysymykset tiedon käytettävyydestä, saatavuudesta ja tiedon hinnasta. Onko tieto avoimesti saatavilla kaikkien käyttöön? Onko tieto käytettävää ja riittävää esimerkiksi päätöksenteon pohjaksi? Mikä hinta tiedosta on maksettava ja kenelle?

Tiedon käytettävyyteen vaikuttaa olennaisesti tiedon laatu. Laadukkaan ja luotettavan tiedon avulla voidaan tehdä johtopäätöksiä yhteiskunnan, kansalaisten, ja yritysten toiminnasta, taloudesta, vaikutuksista ympäristöön, osallistumisesta, hyvinvoinnista jne.

Suomalaisessa yhteiskunnassa luottamus viranomaisiin ja heidän hallussaan olevaan tietoon on hyvällä tasolla, sillä kansalaiset ja yritykset antavat hyvin tietonsa viranomaisten käyttöön. Kansainväliset tutkimukset osoittavat, että viranomaisten, kansalaisten ja yritysten luottamus toisiinsa ja kerättyyn tietoon ei ole kaikissa maissa yhtä korkealla tasolla kuin pohjoismaissa.

Markkinoille on syntymässä uusia tietoaaineistoja, jotka muuttavat tietomarkkinoita merkittävästi. Tällaisia aineistoja ovat yksityisen sektorin hallussa olevat ns. suuret tietoaaineistot (big data), jotka koskevat yleensä kolmansia osapuolia esimerkiksi asiakkaita, toimittajia, kansalaisia. Myös kansalaisia koskevan oma-datan (my data) lisääntyminen muuttaa tietomarkkinoita tuoden uuden tietolähteen tiedon jatkokäyttöön.

Tietomarkkinoiden murroksen myötä tärkeäksi taidoksi tulee relevantin, luotettavan ja laadukkaan tiedon tunnistaminen valtavasta tieto- ja aineistomäärästä. Tosiasiassa vain osa syntyvästä tiedosta on laadukasta jatkokäsittelylle mutta osa tiedosta on turhaa. Myös tekoälyn ja koneoppimisen kehittämisessä on tunnistettu laadukkaan tiedon merkitys. Koneoppiminen edellyttää suuria tietomääriä, joiden pohjalta tekoäly kykenee tulkitsemaan tietoihin liittyviä lainalaisuuksia. Huonolaatuinen tieto voi johtaa väärin johtopäätöksiin yhteiskunnassa vallitsevista periaatteista.

Tiedon laatu ja relevanttius vaikuttaa osaltaan myös tiedon hintaan sillä laadun varmistaminen tuotantoprosessin aikana ja aineiston muokkaaminen jatkokäytön edellyttämään muotoon vaatii resursseja ja ajantasaisia menetelmiä.

Laatukriteerit

Laatua voidaan tarkastella useasta näkökulmasta erilaisia laatukriteereitä hyödyntäen. Yleisesti tuotteen tai palvelun laatu määritellään siten, miten hyvin se vastaa asiakkaan tarpeita ja odotuksia. Myös tiedon laadun määrittelee lopulta se, kuinka hyvin se palvelee käyttäjien tarpeita.

Tilaston tuottajat ovat pitkään käyttäneet laatukriteereitä tuotantonsa ja tuotteiden varmistamisessa. YK:n Euroopan talouskomissio (UNECE) valmisti virallisen tilaston perusperiaatteet, jotka hyväksyttiin jo vuonna 1992. Euroopan tilastojen käytännesäännöt (European Statistics Code of Practice, CoP) määrittävät eurooppalaisten tilaston tuotantoa ja periaatteita varmistuen riippumattomien, rehellisten ja vastuullisten tietojen tuotanto.

Laatukriteereissä keskeisinä nousevat esille tiedon ja/tai aineiston kattavuus, tarkkuus, ajantasaisuus, oikea-aikaisuus, yhtenäisyys ja vertailukelpoisuus. CoP

laajentaa kriteerit lisäksi yhtenäisiin käsitteisiin ja luokituksiin, läpinäkyviin menetelmiin, kustannustehokkuuteen ja kohtuulliseen vastausrasitteeseen.

Kattavuus, tarkkuus ja luotettavuus tarkoittavat sitä, että tiedot kuvaavat hyvin kuvauskohdettaan, niiden tilaa ja muutoksia. Esimerkiksi väestön hyvinvointia kuvaavissa tiedoissa on huomioitu väestö kaikkien ikäryhmien mukaan ja hyvinvointia kuvaavat muuttujat ovat asiakkaan kannalta relevantteja.

Tietojen ajantasaisuus ja oikea-aikaisuus liittyy siihen, että tiedot kuvaavat asian tilaa tuoreimman mittaushetken mukaisena. Yhtenäisyys ja vertailukelpoisuus puolestaan varmistavat eri mittaushetkellä tehtyjen mittausten tuloksien vertailukelpoisuutta toisen mittaushetken tuloksiin tai mahdollisuuden vertailla tuloksia eri alueiden välillä. Vertailukelpoisuus myös edellyttää, että käytössä ovat standardit käsitteet, luokitukset ja tiedon rakenteet. Esimerkiksi kuntien valtion osuuksia määrittäessä tarvitaan ajallisesti (useamman vuoden seurantajakso) ja paikallisesti (yhtenäiset alueluokitukset) vertailukelpoisia tietoja.

Laadun moniulotteisuus on tarpeen ymmärtää käytettäessä yhteiskunnassa syntyvää tietoa. Tiedon tuottajan ja tiedon käyttäjän näkökulmat hyvälaatuisen tietoon voivat kuitenkin poiketa toisistaan. Esimerkiksi tiedon tuottajalle hyvä laatuisen tiedon kokoaminen edellyttää laadun varmistusta tuotantoprosessin jokaisessa vaiheessa (suunnittelu, keruu, käsittely, analysointi, julkaisu) mikä saattaa hidastaa tiedon ja/tai aineistojen julkistamista. Tiedon käyttäjälle puolestaan nopeus voi olla yksi laatulementti.

Laadukkaan tiedon varmistamiseksi ovat vastuut tiedon (jatkuvasta ja säännöllisestä) keräämisestä, ylläpidosta, jakelusta ja hallinnasta määriteltävä selkeästi. Tämä tarkoittaa mm. tiedon omistajien määrittämistä. Omistajalla on oltava vastuu tiedon jatkokäytön mahdollistamisesta käyttämällä yhtenäisiä tunnisteita, standardeja, käsitteitä ja luokituksia sekä tietorakenteita.

Laadun mittaaminen

Laatua voidaan tarkastella objektiivisesti laatumittareilla tai subjektiivisesti käyttäjän omaan kokemukseen nojaten. Objektiivisesti tiedon laatua voidaan mitata useilla mittareilla, joita on kehitetty kansainvälisessä (EU/OECD/YK) yhteistyössä. Tällaisia laatumittareita ovat mm. kattavuus, oikea-aikaisuus, vastausaste sekä käytetyt menetelmät ja eri virhelähteiden tunnistaminen, mittaaminen ja raportointi.

Subjektiivinen mittaaminen puolestaan tehdään kvalitatiivilla kyselyillä, käyttäjien näkemysten ja kokemusten keräämisellä. Esimerkkejä tällaisista kyselyistä ovat asiakaskyselyt ja laatuarviointit.

Metatiedot, tietojen yhdisteltävyys ja yhtenäisyys

Tietojen yhdisteltävyys ja yhteentoimivuus asettaa uusia vaatimuksia tiedon metatiedoille. Jotta tietoa voidaan yhdistää eri lähteistä, on huolehdittava tunnistetiedoista vaarantamatta tietosuoja. Yhteisesti käytettävälle tiedolle on määritettävä metatietojen minimivaatimukset ja standardit, joilla kaikki käyttäjät voivat varmistaa aineiston laadun ja kelpoisuuden käyttötarkoitukseen.

Yhtenäisiä periaatteita sanastoille, tietomalleille, tietorakenteille, koodistoille, luokituksille ja metatiedoille tarvitaan, jotta yhteiskunnassa syntyvät tietoaineistot saadaan parempaan käyttöön.

Viranomaisen laadun varmistajana

Julkinen sektori on kerännyt omaa toimintaansa varten runsaasti tietoa yhteiskunnasta, yritystoiminnasta ja kansalaisista. Kukin viranomainen on

lähtenyt lainsäädännön antamien valtuuksien nojalla keräämään tarvitsemansa tiedot. Tiedonkäsittely ja -hallinnointi on ollut virasto –tai hallinnonalakohtaista.

Yhtenäistä toimintatapaa ja tiedon yhdisteltävyyttä on edistetty ottamalla käyttöön yhtenäiset yksilöitä/yksiköitä koskevat tunnisteen (y-tunnus, henkilötunnus, pysyvä rakennustunnus), yhtenäiset paikkatiedot (alueluokitukset, alueen määrittäminen) ja eräitä luokituksia (esim. COICOP, HS/CN, toimialaluokitus). Erilaisissa hankkeissa on edistetty tietojen yhtenäisyyttä, esimerkiksi JUHTA-työryhmä edisti kuntia koskevien tietojen standardointia.

Tällä hetkellä käynnissä on yhteinen tiedonhallinta (YTI) hanke, jossa luodaan edellytyksiä julkisen hallinnon yhtenäiselle tiedonhallinnalle. Hankkeessa kehitetään mm. yhteentoimivuuden menetelmiä ja välineitä sekä sanastoja, koodituksia ja luokituksia.

Työtä yhtenäisyyden ja yhteentoimivuuden eteen tarvitaan paljon, jotta yhtenäiset sanastot, koodistot jne otetaan käyttöön laajasti yhteiskunnassa. Tämän myötä aineiston käsittelyn kustannukset ja samalla tiedon hinta voivat alentua.

Mitä huomioitava tiedon laatua ja sopivaa hintaa arvioitaessa?

- määritä selkeästi tiedon tarve eli mihin aiot tietoa käyttää ja käytä selkeitä (kansainvälisesti tai kansallisesti hyväksytyjä) käsitteitä
- selvitä tietojen saatavuus, tuotantomenetelmä, tuotannon jatkuvuus, tietojen vertailukelpoisuus (yli ajan, yli alueiden)
- selvitä metatiedot ja mahdolliset lasketut laatumittarit
- selvitä tiedon hinta tuottajalta ja arvioi sitä laatuun nähden.