

Keskustelua – Diskussion

Digitaalisten tietoaineistojen arvo ja saatavuus

Tutkimusaineistojen saatavuus on ollut esillä viimeaikaisissa tiedeyhteisön keskusteluissa. Hannu Katajamäki (2002) nostaa ongelmaksi Tilastokeskuksen keräämien aineistojen huonon saatavuuden – siis tutkijan näkökulmasta katsottuna niiden korkean hinnan. Erkki Niemi (2002) vastaa Katajamäen kritiikkiin, jälleen Tilastokeskuksen näkökulmasta.

Kysymyksessä on kuitenkin paljon maantieteilijöiden ja Tilastokeskuksen välistä kaupantekoa laajempi ongelmakenttä. Monissa tieteissä, myös maantieteessä, nojaututaan nykyään uusiin tietekniikan tarjoamiin menetelmiin. On mielekäs ajatella, että tutkimuksen käytössä olisivat aina parhaat mahdolliset aineistot. Muussa tapauksessa tutkija on kuin mestarikokki eilisen kalan äärellä: parasta lopputulosta ei voi saavuttaa.

Tieteellisessä tutkimuksessa voidaan itse kerätävien aineistojen lisäksi käyttää kahdentyyppisiä digitaalisia aineistoja: (1) sellaisia, jotka ovat saatavilla julkisesti maksua vastaan ja (2) sellaisia, joilla ei ole määriteltyä hintaa eikä markkinapaikkaa. Ensimmäiseen ryhmään kuuluvat tyypillisesti kunnallis- ja valtionhallinnon sekä yritysten tuottamat aineistot. Jälkimmäiseen ryhmään voidaan lukea joukko tutkimuslaitosten, -hankkeiden ja yksittäisten tutkijoiden keräämiä aineistoja.

Julkisin varoin kerätyt aineistot ovat esimerkiksi kyselyiden tuloksia tai kartoituksia, joiden koostaminen on ollut osa yhteiskunnan normaalia toimintaa ja työnjakoa. Monet niistä ovat jo digitaalisessa muodossa jossakin valtion yksikössä kuten Tilastokeskuksessa, Maanmittauslaitoksessa tai Metsäntutkimuslaitoksessa. Myös digitointi ja tarvittavien tietojärjestelmien kehitys lienee toteutettu pääosin julkisin varoin. Vaikka joihinkin työvaiheisiin voi sisältyä paljonkin ulkopuolista rahoitusta, tämäkin on usein julkisen sektorin sisäistä eurojen kierrätystä.

Digitaalisten tietoaineistojen oikeuksista ja hinnoittelusta on keskusteltu esimerkiksi paikkatietoalan *Positio*-lehdessä jo pitkään. Antti Rainio (1993) peräsi tietojenkäsittelyn automatisoinnin tuomien säästöjen heijastumista hinnoitteluun ja kaipasi korjausta tottumukseen, ettei kalliita tietoja käytetä lainkaan. Lähes vuosikymmentä myöhemmin Risto Kalliola (2001) ja Kirsi Mäkinen (2002)

kirjoittivat tietoaineistojen arvosta ja käytöstä yhteiskunnassa. Viesti on sama: tietojen saatavuus on yhteiskunnan kokonaisedun kannalta liian rajoitettua.

Niemi (2002: 85) korostaa, että maksullisten aineistojen lisäksi on muita aineistoja, jotka olisivat helposti saatavilla, mutta jäävät käyttämättä. Esimerkiksi yliopistoissa on julkisin varoin rahoitettuja tutkimusprojektien yhteydessä koottuja raaka-aineistoja. Kun tulokset on julkaistu, arvokas data hautautuu helposti kaappeihin ja kovalevyille. Aineistojen kartuttamisen sijaan alkuperäistietoja hukataan, vaikka ne voisivat palvella muitakin tutkimuksia. Uusiokäyttö olisi mahdollista jos aineistot olisivat jossakin saatavilla. Tutkijoiden tulisi myös motivoitua luovuttamaan aineistojaan käytettäviksi edelleen.

Mielestämme tulisi etsiä keinoja olemassa olevien tietovarantojen hyvän saatavuuden ja mahdollisimman laajan käytön varmistamiseksi. Länsimaisen kulttuurin menestys on paljolti perustunut tietojen kartuttamiseen ja jakamiseen, ja tätä perinnettä tulisi tietoyhteiskunnassakin kunnioittaa. Julkisen sektorin olisi suotavaa näyttää esimerkkiä, tuleehan sen toiminnan perustua nimenomaan yhteiskunnan kokonaisedun tavoitteluun.

Miksi hinnoitellaan ja mitä siitä seuraa?

Käsitykset julkisen sektorin tehtävistä ja toiminnasta yhteiskunnassa ovat viime vuosina muuttuneet nopeasti. Nykyilmapiirissä valtiontalous nähdään verraten itsenäisten tuloyksiköiden kokonaisuutena. Rönsyt leikataan pois yksityistämällä. Jäljelle jäävien osien toivotaan kehittävän keinoja menestyä huolimatta siitä, että budjettirahoitusta kiristetään. Kun rahoitus tulee järjestää osin omilla tuloilla, seurauksena on hinnoittelujärjestelmän kehittäminen. Aiheuttamisperiaatteen mukaan kunkin tulee maksaa tilaamistaan palveluista.

Aina ei kuitenkaan ole yksiselitteistä, mikä on uuden tilauksen aiheuttamaa kustannusta ja mikä osuus on sillä aineellisella tai aineettomalla pääomalla, joka on vuosikymmenien aikana budjettilaroin jo rahoitettu. Harmaalla alueella ovat aineistojen tuotannon ja ylläpidon kustannukset, tuotekehityksen osuus, laitteistot ja muu infrastruk-

tuuri sekä henkilöstökulut. Tilattujen aineistojen suoranaiset irtiottokulut ovat asiakkaan kannalta hyväksyttäviä, mutta silloinkin täysimääräisinä vain, jos julkinen palvelu ei kuulu kyseisen yksikön tehtäviin.

Ajatellaan, että yhteiskunnassa A julkisella virastolla B on perinteisesti ollut tehtävänä tuottaa, hallita ja välittää aineistoa C. Kokonaisuutena ajatellen C:stä vastaaminen on ollut B:n tehtävä, jotta A toimisi hyvin. Nykyinen budjettimuoti kuitenkin aiheuttaa B:lle paineita ulosmitata monopoliasemaansa C:n hallitsijana.

Tällaisessa asetelmassa tiedon saanti voi muuttua elitistiseksi ja arkisissa tilanteissa voidaan joutua tinkimään laadusta arvaamattomin seurauksin. Jo tulosityksiköiden perustaminen julkisen hallinnon sisällä voi johtaa surkukupaisiin ongelmiin aineistojen saatavuudessa. Esimerkiksi kaupungin virastoissa aineistot saattoivat taannoin liikkua hallinnonalalta toiselle automaattisesti, mutta viime aikoina virastojen välinen tililiikenne on yleistynyt. Sen tuloksena itse tuotettujen aineistojen saatavuus voi heikentyä jo kuntaorganisaation itsensä sisällä.

Suomen ympäristön tilaa ja hoitoa koskevien tietojen kerääminen ja ylläpito on ympäristöhallinnon tärkeä tehtävä. Vuosikymmenien työhön perustuvia digitaalisia aineistoja on runsaasti ympäristöhallinnon sisäisessä verkossa, mutta ulkopuoliset eivät niitä saa nähdä – eivät yliopistojen tutkijat, opiskelijat, tietoja tarvitsevat kansalaiset tai yrittäjät. Mitä ympäristöhallinnon intressiä loukkaa, jos Natura-alueiden viralliset rajat, tiedot eläimistä ja kasveista tai mittaustulokset järvien ja rannikon tilasta olisivat vapaasti koko maailman käytettävissä? Eikö päinvastoin olisi ympäristön hyvän hoidon kannalta tavoitteellista varmistaa tällaisten tietojen mahdollisimman laaja käyttö yhteiskunnassa?

Vastaavasti digitaalisten kartta-aineistojen saatavuutta rajoittavat Maanmittauslaitoksen tuloksellisuuspainet. Kun digitaaliset aineistot on hinnoiteltu korkealle ja niitä tarvitsevat puutavara- ja teleyhtiöt maksavat silti, näytettävää tulosta syntyy. Samaan aikaan kunnissa, maakuntahallinnoissa, yliopistoissa ja tutkimuslaitoksissa pohditaan mahdollisuuksia säästää aineistomaksuissa. On huoletuttavaa, jos tämän asetelman seurauksena lisääntyvät paineet käyttää huonolaatuisia lähteitä.

Ilmaisten kartta-aineistojen käyttö voi aiheuttaa arvaamattomia ongelmia (Kalliola 2000). Esimerkiksi Mapinfo-ohjelman mukana tuleva digitaalinen maailmankartta on kelvoton: siinä Länsi-Eurooppa on esitetty melko tarkan näköisesti, mutta entisen itäblokin maat on digitoitu kulmikkaasti, Laatokka puuttuu ja Suomen ainoa järvi on mer-

kitty Kuusamon tuntumaan. Kenenkään ei tulisi käyttää tällaisia aineistoja, mutta ne ovat piinallisen helposti saatavilla. Mainittua laatua on jo nähty esimerkiksi maantieteen väitöskirjassa, useissa tieteellisissä julkaisuissa ja Genimapin ilmoituksessa *Helsingin Sanomissa* (12.8.2002).

Ilmaisia ja helposti saatavilla olevia tietoja kopioidaan ja levitetään tehokkaasti, mutta huuhaan päälle on vaarallista rakentaa. Jos huonon pohjakartan yhteyteen digitoidaan uusia teemoja ja lopulta käytetty alusta hävytetään, tuntemattomat sijaintivirheet ovat jo saastuttaneet uusiakin aineistoja. Ongelmallisten karttakerrosten tunnistaminen ja korjaaminen voi käydä kalliiksi, kun niitä käytetään yhdessä laadukkaiden aineistojen kanssa. Jos haluamme välttää turhia ongelmia ja edistää Niemen (2002) ehdottamaa ajatusta Suomesta avointen tietolähteiden maana, se tulisi tehdä tarjoamalla käyttöön laadukkaita pohja-aineistoja niin helposti, että niiden valitseminen käyttöön olisi yhtä vaivatonta kuin ilmaistenkin aineistojen. Tämä olisi helppo toteuttaa Internet-aikana: tahto ja siitä johdettu politiikka ratkaisevat.

Raaka-aineistot talteen

Harvat organisaatiot tuottavat aineistoja päätoimeenaan – usein raakadatan kerääminen on vain välivaihe. Näin on esimerkiksi ympäristöhallinnossa: sen päätehtävä on tutkia ja vaalia ympäristön tilaa, ei tuottaa digitaalisia tietoaineistoja omiin arkistoihinsa. Kunnan virastot hoitavat julkisen palvelun tehtäviään ja käyttävät tähän työhön tietoaineistoja. Tutkijat keräävät aineistoja voidakseen julkaista tieteellisiä tutkimuksia.

Kun digitaalisia aineistopankkeja on koottu, olisi suotavaa että karttunutta tietoa ei hukattaisi. Mutta kun jopa julkisen hallinnon sisäisessä kulttuurissa on etäännytty aineistojen vapaan käytön kulttuurista, on vaikeaa ajatella, että organisaatioiden, alueiden tai valtioiden välille syntyisi suotuisa ilmapiiri avoimemman aineistopolitiikan kehittämiseksi. Protektionismista kärsivät kuitenkin kaikki.

Nykyinen aineistopolitiikka teettää tarpeetonta päällekkäistä työtä. Esimerkiksi Maanmittauslaitos ja ympäristöhallinto koostavat kuntakohtaisista kaavoista omia alueellisia kaavayhdistelmiään. Eikö olisi mielekkäämpää yhdistellä kaavoja vain yhdessä paikassa ja antaa työn tulos kaikkien tarvitsijoiden käyttöön? Nyky-yhteiskunnassa projektityöskentely on usein lyhytjänteistä: rahoituskaudet ovat lyhyitä, ryhmiä syntyy ja hajoaa ja asian tuntijoiden toimenkuvat muuttuvat toistuvasti. Päätyvissä projekteissa arvokkaat raaka-aineistot ovat vaarassa kadota, jos niiden säilymisestä

ei huolehdi ensisijaisten julkaisu- tai muiden tavoitteiden täyttymisen jälkeen. Vaikka dataa varastoidaan, käsittelemättömät tiedot hautautuvat helposti. Joskus niiden tuottajat voisivat olla suopeita aineistojensa jakamiselle, mutta ei ole olemassa selvää ohjetta, jonka mukaan toimia. Tieto aineistojen olemassaolosta jää muistin ja kuulopuheiden varaan. Tutkijoiden näkökulmasta on myös tärkeää varmistua siitä, että tieteellisen meritoitumisen ja kiittämisen mekanismeja kunnioitetaan.

Pyrittäessä tehostamaan digitalisten tietoaisteiden uuskäyttöä ensimmäinen haaste on aineistojen listaaminen sekä tämän tiedon välittäminen siitä kiinnostuneille. Hyvin laaditut metadatan (tietoaisteiston kuvaus) sisältävät tietokannat hakutyökaluineen voivat tarjota tehokkaan väylän oivaltavaa yhteistoiminnan mahdollisuuksia. Myös päällekkäistä työtä voidaan välttää, kun saatavilla on oikeaa ja ajantasaista tietoa siitä, millaisia aineistoja on olemassa ja kuka niitä hallinnoi.

Mutta millä edellytyksillä aineistoja voi saada käyttöönsä? Ratkaisua voidaan hakea tapaus tapaukselta, kuten nykyisin paljolti tapahtuu. Tämä on kuitenkin tehotonta ja johtaa helposti sekalaisen ratkaisumallien kehittymiseen. Helpointa olisi, jos vallitsevana periaatteena olisi julkisin varoin koostettujen perusaineistojen mahdollisimman pitkälle viety maksuttomuus sekä tähän tukeutuva selkeä välitysmekanismi. Yhteistyöhön perustuva kustannustehokkuus lienee monessa tilanteessa taloudellisempaa kuin rahastukseen keskittyminen.

On aihetta keskustella yhteiskunnan kannalta edullisista ratkaisuista kokonaisuutena, ei siivu siivulta. Erinäisiä julkisten aineistojen saatavuutta ja käyttöä selvittäviä hankkeita onkin käynnissä sekä kansallisesti (esimerkiksi Suomen paikkatietoasiain neuvottelukunta) että kansainvälisesti. On tarpeen tarkastella asioita kokonaisuuden, ei vain sen osien kannalta. Siten keskustelussa tarvitaan varmasti myös maantieteilijöiden kokonaisvaltaista ajattelutapaa.

Kokonaistyyttöä voidaan vähentää, kun minimoidaan aineistojen hakemiseen, käyttöoikeuksien tarkastamiseen ja siirtoon käytettävä työaika. Keskitetty välitysmekanismi, *clearing house* -järjestelmä, voi vastata molempiin tarpeisiin. Siinä aineistojen luovuttajille ja niiden tarvitsijoille tarjotaan yksi paikka, johon olla yhteydessä. Aineiston tarvitsija tietää, mistä selata aineistoja, ja tuottaja tietää minne aineistonsa luovuttaa – vain yhden kerran. Suunnitelmia ja prototyyppisiä tiedonvälitystoreista on jo kehitetty eri puolilla maailmaa, mutta yleiseksi malliksi soveltuvia esimerkkejä ei tietyvästi vielä ole luotu.

Alueellinen tietoaarkisto kehitteillä

Turun yliopiston maantieteen laitoksella tehdään tietoaisteistojen saatavuuden parantamiseen tähtäävää tutkimusta Maj ja Tor Nesslingin säätiön rahoituksella. Hankkeen taustalla on vuonna 1999 käynnistetty paikkatietoalan alueellinen yhteistyö, johon osallistuvat myös Varsinais-Suomen liitto ja Lounais-Suomen ympäristökeskus. Projektiluontoisesti toiminut hanke vakiinnutettiin vuonna 2002 kiinteäksi osaksi kaikkien kolmen organisaation toimintaa.

Syntyneen paikkatietoalan palvelukeskuksen (www.varsinais-suomi.fi/lounaispaikka) yksi ydintoiminto on alueellinen digitaalisten tietoaisteistojen arkisto, jonka ajatusmalli muistuttaa Tampereen yliopiston ylläpitämää Yhteiskuntatieteellistä tietoaarkistoa (www.fsd.uta.fi). Kehitysvaiheen teemaksi tässä työssä valittiin ympäristöön ja luonnon monimuotoisuuteen liittyvät aineistot, sillä ne ovat paikkaan sidottuja. Lähes kaikki biodiversiteetti- ja ympäristötiedot voidaan tallentaa ja analysoida paikkatietojärjestelmillä.

Arkiston yhtenä tavoitteena on tarjota tie ympäristöä kuvaavien tietoaisteistojen tehokkaaseen löytämiseen – arkiston Internet-palveluun sisältyy myös karttakäyttöliittymä. Tavoitteena on myös tuottaa toimiva malli arkiston toimintamekanismeista. Arkiston toivotaan kehittyvän aineistoliikenteen solmukohtaksi, joka kerää, jäsentää ja jakaa digitaalisia tietoaisteistoja. Sen sijaan aineistojen tuottaminen ei kuulu arkiston tehtäviin, joskin arkistoituja aineistoja voidaan tarvittaessa muokata asiakkaan tarpeiden mukaisiksi. Tällöin voidaan asiakkaalta veloittaa työn hinta.

Arkistotoiminnassa on kaksi päätapahtumaa, jotka vaativat neuvotteluja ja sopimuksia: aineiston siirtäminen tuottajalta arkistoon ja aineiston siirtäminen arkistosta asiakkaalle. Näissä tapahtumissa on kiinnitettävä erityistä huomiota lainmukaisuuteen ja tekijänoikeuksien haltijan oikeus turvaan. Aineisto luovutetaan arkistoon luovutus sopimuksella, jossa määritellään oikeudet aineistoihin, hyväksytyt jatkokäyttäjärühmät, jakelutapa ja muita yksityiskohtia. Vastaavasti aineiston siirtäminen jatkokäyttäjälle vaatii käyttösopimuksen, jossa määritellään kenelle ja mihin käyttötarkoitukseen aineistot luovutetaan. Kaikissa vaiheissa digitaalista aineistoa seuraa aineistokuvaus, johon sisältyy teknisten tietojen lisäksi keskeisiä tietoja luovutussopimuksesta.

Joidenkin aineistojen kohdalla on tarkoituksemukaisempaa siirtää arkiston esitettäväksi ainoastaan metadatan ja jättää itse aineisto tuottajan haltuun. Tällaisessa tapauksessa arkisto palvelee tietojen näyteikkunana, joka tarjoaa näkyvyyden

lisäksi aineiston haltijan yhteystiedot. Ympäristötietoarkisto toimii vapaan ja avoimen tietojenvälityksen periaatteen mukaisesti. Kaikkia arkiston välittämiä aineistoja koskevat yleiset käyttöehdot, joiden periaatteisiin kuuluvat ei-kaupalliset käyttötarkoitukset, aineiston oikeuksien säilyminen tuottajalla, riittävästä tietoturvallisuudesta huolehtiminen sekä aineiston alkuperän mainitseminen.

Paraneeko aineistojen saatavuus?

Laajoja digitaalisia tietoaineistojen välitysmekanismeja on esimerkiksi Internetin kautta saatavilla vasta vähän, vaikka tässä kirjoituksessa kuvatut tarpeet ovat maailmanlaajuisia. Monet nykyisistä palveluista ovat lähinnä metatietoa välittäviä alueellisia keskuksia. Aineistoja jakelevia palveluita löytyy vasta vain kourallinen, ja nekin ovat usein yksittäisten projektien suppeita aineistopankkeja tai palveluita, jotka perustuvat dokumentoimattomiin toimintatapoihin ja -ympäristöihin.

Digitaalisten tietoaineistojen saatavuutta pyritään kuitenkin parantamaan monin tavoin. Näkyvimmat hankkeet, kuten esimerkiksi luonnon monimuotoisuutta koskevan biodiversiteettitiedon kansainvälinen *clearing house* -mekanismi ovat maailmanlaajuisia. Ylhäältä alaspäin etenevä vaikuttaminen on kuitenkin hidasta, etenkin kun paikallisen tason toimintaan ei ole selviä malleja (Laihonen ym., *arvioitavana*). On nähtävissä, että tiedon saatavuuden kysymyksiin on kiinnitettävä yhä enemmän huomiota kaikilla yhteiskunnan tasoilla paikallisista globaaleihin toimintaympäristöihin.

Digitaalisten aineistopankkien toteuttaminen on niin yhteiskunnallisesti, oikeudellisesti kuin teknisestikin haasteellinen tehtävä. Aineistojen yhteiskäyttö on vielä melko ongelmallista puuttuvien standardien vuoksi. Sama koskee korostetusti mahdollista suorakäyttöä tietokannoista. Oman ongelmansa muodostavat tekijänoikeudet, sillä monesti aineistot koostuvat osista, joiden alkuperäisten tekijöiden ja oikeudenomistajien jäljittäminen voi olla raskas prosessi. Usein aineistojen luovuttami-

nen jatkokäyttöön kariutuu liian monimutkaiseen tekijöiden ketjuun.

Tietoarkistojen ja niihin liittyvän käyttökulttuurin kehittäminen on kuitenkin askel kohti avoimempia aineistomarkkinoita. Omalla toiminta-alueellamme hyvä valmius alueelliseen paikkatietoyhteistyöhön on avaamassa mielenkiintoisia yhteistyön mahdollisuuksia Lounaispaikka-konaisuuden ympärille. Tähän kehitykseen liittyy paljon maantieteellisiä näkökulmia. Toivottavasti samantapaista kehitystä tapahtuu kaikkialla ja kaikilla aloilla.

Lopuksi pohdittavaa *Terran* lukijoille: olisitko Sinä valmis antamaan omien tutkimuksiesi raaka-aineistoja muiden käyttöön, kun et enää käytä niitä itse? Haluatko siinä tapauksessa luovuttaa aineiston itse jokaiselle kiinnostuneelle käyttäjälle erikseen, vai luovuttaisitko mieluummin aineiston vain kerran – yhteiseen tietoarkistoon?

KIRJALLISUUS

- Kalliola, R. (2000). Quality, information and maps in the information society. *Foresight* 2: 2, 223–229.
- Kalliola, R. (2001). Laadukasta tietoa kaikille. *Positio* 10: 1, 23.
- Katajamäki, H. (2002). Kansallisen maantieteen kunnia on palautettava. *Terra* 114: 1, 1–2.
- Laihonen, P., M. Rönkä, H. Tolvanen & R. Kalliola (*arvioitavana*). Geospatially structured biodiversity information as a component of a regional biodiversity clearinghouse. *Biodiversity and Conservation*.
- Mäkinen, K. (2002). Tiedon arvo. *Positio* 11: 1, 3.
- Niemi, E. (2002). Kommentteja Hannu Katajamäelle kansallisen maantieteen kunniaa, tietojen maksullisuudesta ja suomalaisen yhteiskunnan erityisluonteesta. *Terra* 114: 2, 83–85.
- Rainio, A. (1993). Tieto maksaa. *Positio* 2: 2, 3.

HARRI TOLVANEN
RISTO KALLIOLA
*Maantieteen laitos,
Turun yliopisto*