

KIRJOITUKSIA DATASTA

Suomen Akatemian ja opetus- ja kulttuuriministeriön järjestämässä Tiedefoorumissa kesäkuussa käytiin keskustelu ”Rakasta dataa”. Siihen osallistuivat Jaakko Suominen, Reijo Sund ja Maria Kopsakan-gas-Savolainen. *Tieteessä tapahtuu* -lehti pyysi heiltä aiheesta kirjoituksia. Niissä vastataan kysymyk-siin: Mikä on dataa, mistä se on tullut ja miten sitä on käsitelty? Mitä tulevaisuuden trendejä ja rajoit-uksia on datan käytölle? Miten data mahdollistaa uudenlaiset energiamarkkinat?

KUINKA LAKKASIN OLEMASTA HUOLISSANI JA OPIN RAKASTAMAAN DATAA

JAAKKO SUOMINEN

Mikä on dataa ja miten datan historiallisuus vaikuttaa tulevaisuuteen?

Isäni on eläkkeellä, mutta työskentelee yhä säännöllisesti. Hän on pappi, joka on hoitanut kirkollisia toimituksia 1960-luvun lopulta lähtien. Hän on ollut osa vuosittaisessa tiedontuotannon systeemissä, jossa uskonnollinen organisaatio on kirjannut omiin ja myös valtiiovallan tarpeisiin tietoja ihmisistä, muun muassa heidän syntymistään ja kuolemistaan. Kirkonkirjat ovat henkilörekistereitä, joista Suomessa ollaan ylpeitä ja joita käytävät muutkin kuin viranomaiset. Kirkonkirjojen avulla selvitys- ja tutkimustyötään tekevät niin su-kututkijat, historioitsijat kuin jopa lääketieteen ja genetiikan tutkijat. Tietojen systemaattisuus, laa-juus ja valtavan pitkä ajallinen ulottuvuus tekevät kirkonkirjoista mielenkiintoista ja monipuolista tutkimusaineistoa. Niitä voi tutkia kokonaisuute-na tai toisaalta yksilöiden ja sukhistorioiden nä-kökulmasta.

Isäni ei ole ollut vain osa organisatorista tie-dontuotannon järjestelmää järjestelmän vaatimuk-sesta. Hän ei ole ainoastaan toimittanut tietoja virallisiin rekistereihin. Hänen tiedontuotannol-linen toimijuutensa on palvellut myös hänen omaa työhistoriaansa ja biografisesti rakentunutta ama-matillista identiteettiään. Isäni on kirjoittanut omiin mustakantisiin muistikirjoihinsa säännön-mukaisia yksityiskohtia kaikista kirkollisista toimi-tuksistaan: keitä hän on kastanut, vihkinyt ja hau-taan siunannut ja milloin. Hän on myös säästänyt omat paperiset vuosikalenterinsa, joista näkyvät hänen muutkin työaktiiviteettinsä. Niinpä vihki-essään jonkun parin hän voi viitata vaikkapa toi-mituksen kohteena olevan henkilön aikaisempaan kasteeseen tai kohtaamisiinsa henkilön sukulaisten kanssa. Viittaukset tuovat hauskan lisän puhei-siin, antavat taustaa ja kertovan ajan kulumisesta.

Suurta ja pientä

Isäni kirjaamat merkinnät eivät ole digitaalisia, eivät tietokoneelle tallennettuja nollia ja ykkösiä. Ne ovat olemassa käsin kosketeltavina ja muuten aistittavissa muistikirjoissa ja kalentereissa. Mutta niitäkin voi kutsua dataksi. Voi väittää, että isäni on oppinut pitämään datasta, jopa rakastamaan sitä, koska se on hänelle merkityksellistä ja hyödyllistä.

Voi myös väittää, että data on oppinut rakastamaan häntä. Isäni ja hänen datansa ovat osa teknologista ja sosioteknistä järjestelmää. Kuten kulutustutkija Mika Pantzar on todennut, me emme ainoastaan opi käyttämään teknologiaa, vaan se oppii käyttämään meitä. Isäni on oppinut osaksi kirkollisia tiedon rekisteröinnin ja hyödyntämisen käytänteitä, mutta on samalla kehittänyt itsenäisesti henkilökohtaisia toimintatapojaan. Michel Foucault olisi voinut käyttää isääni esimerkkinä omissa tietoon ja valtaan liittyvissä tutkimuksissaan, joissa yksilöiden puhettavat ja toimet sitoutuivat osaksi rakenteita ja niiden muutoksia.

Isäni vuosikymmenien kuluessa keräämä kirkollisten toimitusten data on määrällisesti vain hippunen verrattuna siihen, mitä digiaineista meistä kerääntyy tällä hetkellä joka minuutti erilaisten tietokonejärjestelmien kautta, jonnekin muistiin ja rekistereihin. Mutta tuo hippunen on tärkeää. Kuten todettu, se liittyy yhtä aikaa yksittäisiin ihmisiin ja sillä on arvoa myös kokonaisuutena. Dataa on erilaista ja -kokoista, eri muodoissa olevaa sekä eri tavoin käsiteltävää. Niinpä humanististen alojen tutkijatkin ovat käsitelleet aina dataa, vaikka he eivät ole sitä sillä sanalla kutsuneet eivätkä aina tuijottaneet datankäsittelykoneiden, tietokoneiden, kännyköiden, kameroiden ja mikrofilminlukijoiden ruutuja. Tulevaisuuden trendinä onkin erilaisten ja erikokoisten datojen yhdistäminen vaihtelevien menetelmien avulla.

Sähköaivoista tietokoneisiin

Matematiikan, tiedonkäsittelyn ja kalenterien historiaa tutkineen Arno Borstin mukaan sana data tulee latinan yksikkömuodosta *datum*, joka tarkoittaa jotain annettua. Datumilla viitattiin varhaiskeskiajalta eteenpäin päivään, jolloin esivallan antama pitkään valmisteltu määräys tai asetus annettiin tiedoksi. Sanakirjatiedon mukaan 1600-lu-

vun puolivälistä lähtien datalla alettiin viitata filosofiassa tiedettyihin asioihin, faktoihin, joita käytettiin järkeilyn tai laskennan pohjana. Myöhemmin sana yhdistyi näiden asioiden koneelliseen käsittelyyn. Data on sittemmin suomennettu myös tiedoksi tai informaatioksi puhuttaessa esimerkiksi tietokoneista tai tietojenkäsittelystä, mikä on tuottanut useita sekaannuksia.

Data-sanan suomentaminen liittyi alkujaan 1950-luvun lopulla keskusteluun uuden digitaalisen koneen nimeämisestä. Konttoriautomaation ammattilaisten Reikäkorttiyhdistyksessä ja sen sanastokeskustelussa mukana ollut Heikki Levas tiedusteli loppuvuodesta 1959 suomenkielistä vastinetta data-sanalle. Data oli hänen mielestään virheellisesti suomennettu tiedoksi muun muassa silloin, kun *Electronic Data Processing* -koneet oli nimetty elektronisiksi tietojenkäsittelykoneiksi, joita populaarijulkisuudessa kutsuttiin myös sähkö- tai elektroniaivoiksi. Levas ehdotti data-sanana suomennokseksi tiedos-sanaa. Kielilautakunta ei kuitenkaan katsonut ehdotuksen soveltuvan yleiseen käyttöön, mutta esitti, että elektronisesta tietojenkäsittelykoneesta käytettäisiin lyhyempää nimeä (elektroninen) tietokone. Tietokone vaikiintuikin sitten dataa käsittelevän koneen nimenä käyttöön 1960-luvulla.

Jotkut humanistitkin alkoivat innostua tietokoneista ja niiden mahdollistamista uudenaikaisista tutkimusaineiston käsittelytavoista jo 1960-luvulla. Silloin tietokoneiden pureskelema data oli pitkälti numeroita ja tutkimus liittyi erilaisiin tilastollisen käsittelyn tapoihin ja kvantitatiivisiin lähestymistapoihin. Seuraava, laajempi innostuksen aalto syntyi 1980-luvun alussa, jolloin tutkija pystyi sivuuttamaan aiemmat suurten keskustietokoneiden toimintatavat ja ATK-keskusten portinvartijat sekä käsittelemään dataa omalla henkilökohtaisella tietokoneellaan ja vielä kirjoittamaan samalla koneella tutkimusjulkaisujensa käsikirjoitukset itse. Nykyään arkipäiväiseltä tuntuva asia oli silloin ennenkuulumaton uutuus.

1990-luvulla kiihtynyt tietoverkottuminen toi mukanaan uudenlaisen globaalimman ja reaaliaikaisemman datasuhteen. Oma tietojenkäsittely alkoi näyttäytyä osana laajaa ihmisten ja koneiden jatkuvan datankäsittelyn verkostoa. Tietynlaisena aaltoliikkeenä data on taas ainakin osin karan-

nut yksittäisten tutkijoiden ulkopuolelle erilaisiin mystisiin ”pilviin”. Suurten datamassojen käsittely vaatii yleensä useamman erilaisen henkilön osaamisen yhdistämisen. Nyt olemme akateemisessa työssä ehkä siis jonkinlaisessa neljännessä digitalisaation vaiheessa.

Digitalisaation vaiheita on edeltänyt todella monia muita epädigitaalisia datan käsittelyvaiheita. Digitaaliset ja analogiset vaiheet tai aallot eivät ole toisiaan poissulkevia vaan elävät monesti edelleen rinnakkain ja päällekkäin.

Tämä päällekkäisyys tuo mukanaan ongelmia mutta samalla mahdollisuuksia. Vaikka isänikin pitää edelleen kirjaa toimituksistaan vihko- ja kalenterimuodossa, myös hän on uudistanut toimintatapojaan tietokoneiden, verkkojen ja mobiiliteknologian avulla.

Kirjoittaja on informaatioteknologian kulttuurihistorian tutkimukseen erikoistunut digitaalisen kulttuurin professori ja humanistisen tiedekunnan dekaani Turun yliopistossa.