

## **Koodari ja koronavirus**

“Kiinassa leviää joku uusi virus. Sitä voisi alkaa seurata”, toimituspäällikkö Esa Mäkinen totesi kävellessään työpöytäni ääreen Helsingin Sanomien toimituksessa tammikuussa 2020. “No eiköhän aleta sitten”, vastasin.

Tehtävänanto ei ollut heti ilmeinen. Raportteja ja tutkijoiden kommentteja uudesta viruksesta oli jo alkanut kertyä, mutta ei ollut selvää, mitä datajournalismiin keskittyvä Helsingin Sanomien Datadeski asialle voisi tehdä. Jostain tarvittaisiin helposti tietokoneen ymmärrettävää tietoa ja jokin esitys, jolla taudin leviämistä olisi helpompi ymmärtää.

Tällä kertaa olin onnekas. Yhdysvaltalainen Johns Hopkinsin yliopisto oli alkanut seurata koronavirukseksi tunnistetun viruksen leviämistä ja jakoi tietoa vapaasti kenen tahansa käytettäväksi. Kirjoitin nopeasti ohjelmiston, joka päivittäin haki viimeisimmät tiedot taudin leviämisestä. Yhdessä Datadeskin toimittajien ja graafikoiden kanssa pyörittelimme dataa, ja päädyimme alustavasti esittämään taudin leviämistä maailmankartalla eri värein ja käyrin kuvaamaan leviämismuutoksia. Julkaisimme ensimmäiset versiot tammikuun lopussa.

Seurannan alkuvaiheessa haasteeksi muodostui tiedon muodon jatkuva muuttuminen. Johns Hopkins julkaisi kyllä joka päivä viimeisimmät tietonsa koronaviruksen etenemisestä maailmalla, mutta lähes päivittäin jokin osa tiedoista oli muuttunut edelliseen verrattuna. Tämä on ohjelmoijalle ja ohjelmistolle hankalaa, sillä tietokoneet ovat vain tehokkaita laskimia, ja syöttämällä niille sopimatonta dataa vastauksesi saa vain kaatumisia. Työrutiinikseni muodostui herätä aamulla kello viiden aikaan, tarkastaa mikä tällä kertaa tiedoissa on muuttunut jo ohjelmistoissani räjähtänyt, ja korjata ne mahdollisimman nopeasti (tiedot päivittyivät noin kello kolmelta aamusta).

Jatkoimme erilaisten visualisaatioiden ja analyysien tekemistä maailmalta tihkuvista tiedoista, kun Euroopasta ja Suomesta alettiin raportoida ensimmäisiä havaintoja koronaviruksesta. Eurooppaa varten ei onneksi tarvinnut etsiä uutta tietolähdettä, vaan tietojen kytkeminen Euroopan karttaan riitti. Helmikuun lopulla Italia oli suurin koronavirustartuntojen keskittymä Aasian ulkopuolella, ja taudin saapuminen Suomeenkin vain ajan kysymys.

Automaattisen taudin leviämisen seurannan kannalta haasteeksi muodostui se, että Suomen tilanteesta täytyisi saada ajantasaisempi ja tarkempi kuva, kuin maailman tilanteesta. Johns Hopkinsin tietoja kerättiin eri lähteistä ja päivitettiin takautuvasti sitä mukaa, kun tietoja kertyi. Suomen tietoja ei siellä vielä ollut, ja tärkeää olisi saada tiedot tuoreeltaan eikä kansainvälisten tietojen joukosta.

Kun tauti alkoi levitä Suomessa, ajantasaista tietoa kerättiin päivittäisillä soittokierroksilla sairaaloihin ja listaamalla tiedot omiin järjestelmiimme. Laajensin maailman tilannetta seurannutta ohjelmistoamme yhdistämään käsin kerätyt tiedot mahdollisimman hyvän tilannekuvan luomiseksi. Työtä helpotti se, että Johns Hopkinsin kansainväliset tiedot olivat alkaneet vakiintua muodoltaan, eikä niitä tarvinnut enää päivittäin valvoa (joskin aina silloin tällöin piti ratkoa ongelmia esimerkiksi merentakaisien departementtien vaikutuksia lukuihin).

Toimittajien käsin keräämistä tiedoista alkoi myös yksi koko koronaviruspandemian ajan kestänyt sivujuonne. Aloimme pian julkaista keräämiämme tietoja vapaasti käytettäväksi, sillä muuta helposti saatavaa lähdeä Suomen tautitilanteelle ei ollut vielä olemassa. Tiedoille löytyi nopeasti erilaisia käyttäjiä, kun monet ohjelmoijat tekivät omia seurantatyökalujaan ja erilaiset tahot alkoivat käyttää tietoja omissa analyyseissaan. Sanon erilaiset tahot, koska sain silloin tällöin sähköpostia esimerkiksi kaupunkien kanslioista ja yksityishenkilöiltä, kun tiedoissa oli epäselvyyksiä tai muuten vain kysymyksiä. Ympyrä sulkeutui, kun EU:n viranomaiset ottivat keräämiämme tiedot omaan seurantaansa, joka taas oli Johns Hopkinsin yliopiston tietolähde Suomen tietoja raportoidessaan.

Ajoittaiset sähköpostit eivät olleet ainoa yhteydenottokeino, muutamat lukijat myös etsivät puhelinnumeroni käsiinsä ja soittelivat antaakseen palautetta milloin mistäkin tekemästani visualisaatiosta. Journalismin ulkopuolelta uutisetulinjan syvään päähän eksyneenä kokemus oli mielenkiintoinen.

Maaliskuun puolessavälissä sattui myös tapahtumasarja, joka kuvasi tekemisen tahtia ja tapaa koronavirustilanteen pahentuessa Suomessa. Koska elin ja hengitin koronavirusuutisointia ja tietolähteitä, selasin lauantaina kansainvälisiä medioita ja törmäsin Washington Postin sivustolla yksinkertaistettuun simulaatioon siitä, miten sosiaalinen eristäytyminen vaikuttaa taudin leviämiseen ihmispopulaatioissa. Jutun luettuani ajattelin välittömästi, että tämä kannattaa toteuttaa meillekin.

Käytin lauantain ja sunnuntain simulaation ohjelmointiin (konsultoin viikonloppuna muitakin Datadeskin jäseniä, sillä en ollut ainoa, joka velloi koronavirustietolähteissä), maanantai käytettiin varsinaisen jutun kirjoittamiseen toimittajan ja Datadeskin graafikoiden kanssa ja juttu julkaistiin tiistaina. Mainittu tiistai oli myös sama päivä, jolloin koronavirusta koskevat valmiuslait astuivat voimaan. Jutun kirjoittanut toimittaja itse oli sairaana. Lukijoita kertyi nopeasti yli puoli miljoonaa. Tämä kaikki etätöinä, joihin toimitus oli siirtynyt edellisenä perjantaina.

Simulaatiojuttu oli minulle se hetki, jolloin aika alkoi koronaviruspandemian aikana vääristyä. Etätyöpäivät venyivät, kun valtavaan tiedontarpeeseen vastattiin uusilla visualisatioilla ja analyyseillä. Olimme myös kerryttäneet teknistä velkaa – erilaisia oikoteitä ja ratkaisuja, joilla asioita oli ratkottu ennemmin nopeammin kuin oikein – joiden ylläpitäminen vaati myös huomiota. Yhä suurempi osa Datadeskistä, ohjelmoijista toimittajiin ja graafikoihin, työskenteli aiheen parissa. Koko toimitus järjestettiin koronaviruksen ympärille.

THL alkoi myös julkaista keräämäänsä tietoa koronaviruksen leviämisestä. Saatoimme korvata tietojen käsin keräämisen, mikä osaltaan helpotti seurantaohjelmistojen kirjoittamista. Jatkoin kuitenkin oman julkisen tietolähteen ylläpitoa, koska THL:n tiedon muoto ei ollut helpoin mahdollinen luettava.

THL:n päivittäin julkaisemista tiedoista syntyi nopeasti uutiskilpailun väline. Tiedot päivittyivät noin kello 11:35 päivittäin, ja eri mediat kisasivat ilmoitusten lähettämisestä lukijoilleen ja juttujen julkaisusta. Sekuntitason viritys ohjelmistoillemme oli uusi haaste, mutta sekin saatiin alkukankeuksien jälkeen toimimaan luotettavasti.

Suomen paikallisten tietojen automaattisen julkaisun myötä siirryttiin koronaviruspandemian alkuvaiheen epämääräisten tietolähteiden ajasta tilanteeseen, jossa voitiin keskittyä automaatioon ja hiontaan jatkuvan datalähteiden korjauksen sijaan. Rakensin kesän aikana toimitukselle esimerkiksi työkalun, jolla se pystyi itse luomaan haluamiaan visualisatioita tarjolla olevasta datasta, mikä vapautti työaikaa taas uusien visualisatoiden ja muiden uusien asioiden tekemiseen.

Uutta arkea oli korjailta ja parannella olemassa olevia visualisatiota, käydä kriittisiä keskusteluita siitä, mitä tietoa ja miten sitä pitäisi esittää (kävimme esimerkiksi pitkiä keskusteluita siitä, miten liukuva keskiarvo lasketaan) ja soittaa aika ajoin THL:n viestintään selvittääkseen, miksi heidän tietonsa eivät olleet päivittyneet. Jälkikäteen tuntuu vaikealta sanoa, kuinka kauan tätä kesti – koronaviruksen etätyöaikavääristymässä päivät ja kuukaudet alkoivat sulaa yhteen. Välillä taas keksittiin uusia kulmia aiheeseen ja niitä kokeiltiin. Osa toimi ja osa ei. En ole koskaan työskennellyt kasvuyrityksessä, mutta kuvittelisin sen muistuttavan tämän tarinan kaarta alun käynnistymisepäselvyyksiä seuranneisiin pitkiin työpäiviin, joilla tekeminen oli välillä veitsen terällä tanssimiselta uusia asioita kokeillessa ja välillä vanhojen virheiden kiroilemista.

Oma roolini koronaviruksen seurannassa alkoi myös vähentyä tilanteen asettuessa uuteen epänormaaliin tasapainoonsa. Datadeskin muut ohjelmoijat alkoivat ottaa vastuun seurantaohjelmistostamme kehittämisestä ja parantamisesta. Teimme uuden simulaatiojutun rokotteiden vaikutuksesta ja laumasuojasta tammikuussa 2021, jonka

jälkeen koronaviruseurannasta tuli päätyöni sijaan sivutyö. Automaattinen koronaviruseuranta jatkui Datadeskissä aina joulukuulle 2022, jolloin se lopetettiin sattumalta syntymäpäivänäni.

Omalta osaltani koronaviruseurannan loppuhetki oli kuitenkin huhtikuussa 2022, kun tapasimme THL:n asiantuntijoita käydäksemme läpi poikkeusvuosien tapahtumia ja tavataksemme toisemme kasvokkain ensimmäistä kertaa. Jo pelkkä tapaaminen oli aikakauden loppu, ja käymämme keskustelut pandemiaan liittyvästä mallintamisesta ja median roolista tiedonvälityksessä olivat terapeuttisia. Jossain vaiheessa iltaa joku lohkaisi vitsillä: “Seuraavassa pandemiassa hoidetaan nämä asiat sitten paremmin”.

Helpottuneella naurullani oli hermostunut pohjavire.