

Tehkää perässä!

Datajournalismin tekijät kertovat, miten vetävä datajuttu syntyy. Perehdy esimerkkeihin ja tee rohkeasti omia kokeiluja – datajutun teko on helpompaa kuin miltä se kuulostaa.

Jessikka Aro, teksti



Äidinkielen ylioppilaskoe

Esa Mäkinen,
Helsingin Sanomat

3.

Kokeen vastaukset tallentuivat tietokantaan. Tietokannasta **Esa Mäkinen** laski, missä kielioppikysymyksissä suomalaiset tekevät eniten virheitä.

2.

31.5.2012 julkaistussa kokeessa kysyttiin muun muassa sitä, ovatko lauseet ”Aiotko räjäytellä raketteja uutena-vuotena?” tai ”On hyvä tietää että apu on lähellä.” oikein vai väärin.

1.

Helsingin Sanomien designtuottaja **Petri Salmén** teki kyselyn ulkoasun. Kielitoimittaja **Ville Eloranta** laati kysymykset.



DATAJOURNALISMIPROSESSIN KUVAT TEKIJÖILTÄ, TEEMU KUUSIMURTO, KUVA, HELI SAARELA, GRAFIikka.
HELSINGIN SANOMIEN ÄIDINKIELEN KOE: [HTTP://WWW.HS.FI/KULTTUURI/TESTAA+KIELITAITOSI+HSN+ÄIDINKIELEN+YLIOPPILASKOKEESSA/A1305572080547](http://www.hs.fi/KULTTUURI/TESTAA+KIELITAITOSI+HSN+ÄIDINKIELEN+YLIOPPILASKOKEESSA/A1305572080547)



4.

18. kesäkuuta 2012.
Mäkinen kirjoitti printtileh-
teen jutun kokeesta tallentunei-
den tietojen perusteella. Tiedot
vastauksista tallentuivat nimettö-
minä: vastaajilta kysyttiin vain
sukupuoli ja syntymävuosi. Koe on
edelleen verkossa ja ihmiset vas-
taavat siihen yhä.



helsingin Sanomien huippusuositettu äidin-
kielen ylioppilaskoe alkoi datajournalismin
tuottajan **Esa Mäkisen** omana kokeilu-
projektina. Ennen kokeen laatimista Mäki-
nen oli päivittänyt ohjelmointitaitojaan ja
opiskellut uusia ohjelmointikieliä.

Mäkinen oli pitkään miettinyt, millaisen
kyselyn hän voisi ohjelmoida ja teettää lu-
kijoilla. Äidinkielen yo-kokeen hän päätti
tehdä sen jälkeen kun *Suomen Kuvalehti*
julkaisi helmikuussa oman, hiukan yksinkertaisemman äi-
dinkielen kokeensa.

”Ajattelin, että voisin tehdä paremman kyselyn, joka tal-
lentaa tulokset. Halusin myös tehdä kyselystä pelimäisen”,
Mäkinen kertoo.

Idean saatuaan Mäkinen otti yhteyttä *HS:n* kielitoimitta-
jaan **Ville Elorantaan** ja designtuottajaan **Petri Sal-
méniin**. Eloranta laati kysymykset ja Salmén suunnitteli
ulkoasun.

Mäkinen rakensi kyselyyn koodia puolitoista kuukautta
luppohetkinään töissä ja kotonaan. Yhteensä kokeen raken-
tamiseen, ulkoasun sekä kysymysten suunnitteluun ja toteu-
tukseen Mäkinen arvioi kuluneen kolmelta toimittajalta
noin kymmenen työpäivää.

Kysely julkaistiin 31. toukokuuta, samaan aikaan lukio-
laisten ylioppilaskokousten kanssa. Kokeessa on 30 eri lausetta,
joista osa on oikeinkirjoitussääntöjen mukaan kirjoitettu,
osa ei. Vastausvaihtoehdot ovat oikein, väärin ja tyhjä.

”Kyselyyn ohjelmoitiin yo-kokeen pisteenlaskukaava.
Kun uusi käyttäjä vastasi kokeeseen, ohjelma laski uuden
pisterajan.”

Koetta puffattiin lehdessä, ja se jaettiin Facebookissa ja
Twitterissä. Se levisi verkossa nopeasti: Ensimmäisen tun-
nin aikana 5 000 ihmistä vastasi kokeeseen.

Kun 7 000 ihmistä oli vastannut, Mäkinen huomasi koo-
dissa yhden kirjaimen virheen, joka aiheutti tietojen vir-
heellisen tallentumisen.

”Se oli kuumottavaa, koska samaan aikaan niin moni teki
kyselyä. Sain virheen nopeasti korjattua. Tämä on esimerk-
ki siitä, että juttuja ei voi tehdä täydellisessä maailmassa.”

Helsingin Sanomien tuli kokeesta paljon palautetta.
Monet olivat tuohtuneita, koska kokeen mukaan moni kan-
sakoulussa opetettu kielioppiasia ei enää pitänyt paikka-
ansa.

”Kielikysymykset herättävät ihmisissä intohimoja”, pit-
kään kulttuuritoimittajana työskennellyt Mäkinen sanoo.

Vastaajia kertyi satatuhatta muutamassa päivässä. Kun
määrä alkoi kääntyä laskuun, vastanneita oli 135 000. Silloin
Mäkinen kirjoitti kokeesta tallentuneiden tietojen perus-
teella lehtijutun, jonka hän otsikoi: Ponttoni menee suoma-
laisilta väärin.

”Vaikka vastaajia oli 135 000, kyse ei ole tilastollisesti
pätevästä otoksesta. Mutta se kertoo, mitä yleisimmin kir-
joitetaan väärin.”

Hän teki jutusta myös verkkoversion, joka sai hiukan
alle 200 000 lukijaa.

Mäkinen yllättyi valtavasta käyttäjämäärästä. Etukäteen
hän arveli, että koe kerää enintään 25 000 vastaajaa.

”Perinteinen juttu kieliopista ei olisi kerännyt verkos-
sa pariasataa lukijaa. Perinteisellä jutulla ei myös-
kään olisi pystytty keräämään materiaalia, josta olisi tehty
uusi juttu.”

Kymmeniä datajuttuja tuottanut Mäkinen kehottaa data-
journalismista kiinnostuneita kokeilemaan omia projekteja.

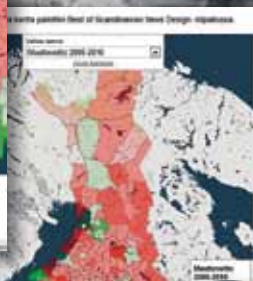
”Ensimmäisenä ei kannata miettiä, mikä olisi hyvä data-
journalismikurssi. Sen sijaan kannattaa katsoa esimerkkejä
ja toteuttaa oma projekti. Datajournalismin tekeminen on
yllättävän helppoa.”

JATKUU SEURAAVALLA AUKAAMALLA



Interaktiivinen kuntatalouskartta

Lauri Vanhala ja Juho Salminen,
Suomen Kuvalehti



2.

Verkkokartta on käytännöllisempi kuin printtikartta: tietoja voi verrata helpommin. Yksittäisiä kuntia voi tutkia zoomaamalla. Kuntien rajojen visuaalisen toteutuksen teki ohjelmointitaitoinen **Lauri Vanhala**.

3.

Kartan teossa haasteellista oli yhdistää Kuntaliiton, Tilastokeskuksen ja toimittaja Hynynsen tiedot. Tietoa oli paljon, sillä kuntia oli yhteensä 336 ja muuttujia kahdeksan.

1.

Idea interaktiiviselle kuntakartalle syntyi, kun toimittaja **Eeva-Liisa Hynynen** laati Suomen Kuvalehteen artikkelia kuntien taloudesta syyskuussa 2011

4.

16.9.2011
Kartta julkaistiin ja tiedot esitettiin myös painetussa SK:ssa.

DATAJOURNALISMIPROSESSIN KUVAT TEKIJÖILTÄ, HELI SAARELA, KUVA JA GRAFIikka
SUOMEN KUVALEHDEN INTERAKTIIVINEN KUNTAKARTTA: [HTTP://SUOMENKUVALEHTI.FI/JUTUT/KOTIMAA/SKN-INTERAKTIIVINEN-KUNTAKARTTA-POHJOISMAIDEN-PARHAIMMISTOA](http://suomenkuvalehti.fi/jutut/kotimaa/skn-interaktiivinen-kuntakartta-pohjoismaiden-parhaimmistoa)

Suomen Kuvalehden entinen verkkopalvelun toimitussihteeri **Juho Salminen** ja SK:n entinen verkkopalvelun toimittaja **Lauri Vanhala** ryhtyivät viime syksynä rakentamaan kuntien taloudellista tilannetta havainnollistavaa interaktiivista karttaa, kun lehteen oli teossa kuntien taloutta käsittelevä juttu.

Jutun kirjoittaneelle freelance-toimittajalle **Eeva-Liisa Hynyselle** oli prosessin aikana kertynyt valtava määrä numerotietoa Suomen kunnista. Toimituksessa syntyi idea: tieto pitää hyödyntää ja visualisoida verkkoon.

"Datan käyttö ja visualisointi oli SK:ssa silloin alkutekijöissään, eikä datan potentiaalia osattu aluksi havaita", kertoo Salminen, joka nykyään työskentelee Otavamedian sähköisen journalismin kehittäjänä.

Aluksi Salminen ja Vanhala miettivät, miten data olisi syytä esittää. Koska kyse oli kunnista, luontevimmaksi vaihtoehdoksi nousi Suomen kartta. Sen jälkeen he pohtivat, mitkä tunnusluvut esitetään. He valitsivat kahdeksan muuttujaa: muuttovoitto 2005–2010, tulokertymä, tulot ja

lainakanta (euroa per asukas), työpaikka kotikunnassa (prosenttia), syntyneet, huoltosuhde ja tulo-veroprosentti.

Osa tiedosta oli avointa ja ilmaista, osan oli SK:lle laskenut kunta-alan asiantuntija. Dataa oli paljon, mutta ei valtavasti, Salminen sanoo.

"Kuntia on yhteensä 336 ja muuttujia kahdeksan. Haastavaa oli erilaisten datasettien yhdistäminen: Kuntaliiton, Tilastokeskuksen ja Hynynsen tiedot piti saada yhtenäiseen muotoon", hän sanoo.

Kuntien rajat visualisoi Googlen ilmaiselle Suomen kartalle. Ajantasaiset kuntarajat saatiin Maanmittauslaitokselta, joka oli julkaissut ne avoimena datana. Rajat visualisoi koodaustaitoinen Vanhala, joka nykyään työskentelee freelance-toimittajana.

"Työläintä oli ohjelmoida sovellus, joka yhdistää eri lähteiden tilastotiedot ja piirtää ne Maanmittauslaitoksen dataa hyödyntäen Googlen karttapohjan päälle", Vanhala sanoo.

Haasteita aiheutti myös kuntien yhdistyminen: tilastot olivat eri vuosilta kuin uusimmat kuntarajat. Vanhalan ja Salmisen piti päättää, mitä tehdään kunnille, joita ei enää ole olemassa ja hankkia

uudet luvut yhdistyneille kunnille.

Ennen julkaisua tiedot tarkistettiin vielä kerran Excelissä.

Projekti vei arviolta yhteensä yhden työviikon kolmelta ihmiseltä eli Hynyseltä, Salmiselta ja Vanhalalta. Suurimman työn teki Vanhala, ja Hynynen toimi sparraajana. Prosessiin osallistui myös muita SK:n toimituksen jäseniä.

Työn tuloksena syntyi interaktiivinen kartta, josta näkee yhdellä silmäyksellä, miten hyvinvointi Suomessa jakautuu: mistä muutetaan pois ja minkälaiset huoltosuhteet ovat. Kartta palkittiin Best of Scandinavian News Design -kilpailussa.

"Kartta toimii monella tasolla. Karttaa voi zoomailla ja siitä voi tutkia oman kotikuntansa tietoja tai laajemmin useiden kuntien tietoja. Kartta palvelee sekä yleistä että spesifimpää kiinnostusta", Salminen sanoo.

Kartan julkaisusta on kulunut vuosi. Muuttaisiko Salminen jotain, jos tuottaisi kartan nyt?

"Täydellisessä maailmassa työkalulla voisi yhdistellä kuntia ja työkalu voisi laskea, mikä on taloudellinen tilanne, jos yhdistää esimerkiksi kolme kriisikuntaa."



Hyvinkään ampujan Facebook-verkoston analyysi

Teemo Tebest, Yle

1.

26.5.2012 **Teemo Tebest** selvitti epäillyn Facebook-kavereiden suhteet ja muutti ne verkostomuotoon. Nimet ja Facebook-ID:t on muutettu tunnistamattomiksi.

2.

26.5.2012 Verkostosta saatu data visualisoitiin ja käsiteltiin Gephi-ohjelmassa. Epäillyn kavereita edustavien pallojen koko on määritelty sen mukaan, miten keskeisiä he verkostossa ovat.

4.

29.5.2012 Yle Uutiset teki verkostoanalyysistä jutun verkkoon. Aihetta käsiteltiin myös illan televisiuutisissa.

3.

27.5.2012 Gephistä tulostettu, blogissa julkaistu verkostokuva. Punaiset pallot ovat naisia, siniset miehiä. Pallot sijoittuvat sitä lähemmäs toisiaan, mitä parempia kavereita niiden edustamat henkilöt ovat.

DATAJOURNALISMIPROSESSIN KUVAT TEKIJÖILTÄ: HELI SAARELA, KUVA JA GRAFIIKKA
HYVINKÄÄN AMPUJAN FACEBOOK-VERKOSTOANALYYSI: [HTTP://DATAJOURNALISMI.BLOGSPOT.FI/2012/05/HYVINKAAN-AMPUMISTAPAU.SHTML](http://datajournalismi.blogspot.fi/2012/05/hyvinkaan-ampumistapaus.html)

Kun nuori mies ampui ihmisiä Hyvinkään keskustassa viime toukokuussa, Ylen web-kehittäjä **Teemo Tebest** päätti tutkia epäillyn Facebook-kaveriverkoston. Tebest halusi selvittää, oliko epäillyllä yhteyksiä erikoisiin ryhmiin, kuten äärijärjestöihin.

”Tylsempi selitys analyysin teolle on, että minulla oli valmiina työkalu, joka mahdollisti Facebook-kaverisuhteiden selvittämisen. Halusin käyttää työkalua.”

Ensin Tebest googlasi epäillyn henkilöllisyyden. Sen jälkeen hän haki koodaamallaan ohjelmalla Facebookista epäillyn kaverit ja kaverien väliset suhteet sekä tulosti ne verkostomuotoon.

Analyysi paljasti, että epäillyn Facebook-kaveriverkosto oli hyvin tiivis ja koostui pääosin miehistä. Tästä Tebest päätteli, ettei epäilty ollut muuttanut paikkakunnalta toiselle – eri paikkakunnilla muodostetut kavereussuhteet olisivat näkyneet verkostossa omina ryppäinään.

”Sitä johtopäätöstä ei voi tehdä, etteikö epäillyllä olisi yhteyksiä ääritahoihin – vaikka ne eivät fb-verkostosta olekaan nähtävissä.”

Toimittajat voivat Tebestin mukaan käyttää verkostoanalyysiä tutkivan journalismin välineenä.

”Verkosto on esitysmuotona vaikealukuinen, mutta toimittaja voi hyödyntää analyysistä saatavaa tietoa. Jos verkostosta olisi löytynyt ryhmä ihmisiä – esimerkiksi muista verkostoista irrallisia henkilöitä, jotka nimensä perusteella eivät kuulu joukkoon – olisin toimittajana halunnut tutkia, keitä he ovat.”

Kun analyysi oli valmis, Tebest visualisoi tiedot Gephi-nimisellä ohjelmalla ja kirjoitti aiheesta datajournalismia käsittelevään blogiinsa. Myöhemmin hän vinkkasi aiheesta Ylen uutisille, ja samana iltapäivänä aiheesta tehtiin juttu Ylen verkkosivuille. Aihetta käsiteltiin myös illan tv-uutisissa.

Työkalun tekemiseen Tebest arvioi käyttäneensä päivän tai kaksi. Analyysin tekemisessä kesti noin 15 minuuttia.

”Koodin kirjoittamiseen vaaditaan teknistä orientoituneisuutta. Mutta valmis työkalu on 'paina nappulaa, nauti elämästä' -tyylinen.”

Samankaltaista työkalua käytettiin Ylellä viime keväänä, kun FST:n tutkiva ohjelma *Spotlight* tutki Facebookissa perussuomalaisten poliitikkojen yhte-

ysisiä äärioikeistolaiseen ryhmittymään Finnish Defence Leagueen.

”Analyysi paljasti, että perussuomalaiset ja FDL ovat tekemisissä Facebookissa. Kun asiasta kysyttiin perussuomalaisilta poliitikoilta, he eivät olleet kuulleetkaan FDL:stä. Analyysin perusteella heille voitiin osoittaa toisin.”

Työkalua voi käyttää kenen tahansa fb-verkoston selvittämiseen, vaikka kaveritiedot eivät olisi-kaan julkisia.

”Fb:ssa on porsaanreikä: vaikka henkilö olisi pääosin piilottanut tietonsa, Facebookissa voi henkilö kerrallaan selvittää, ketkä ovat kavereita keskenään. Prosessi vaatii ainoastaan, että analyysiin valitun henkilön kaverilista on tiedossa.”

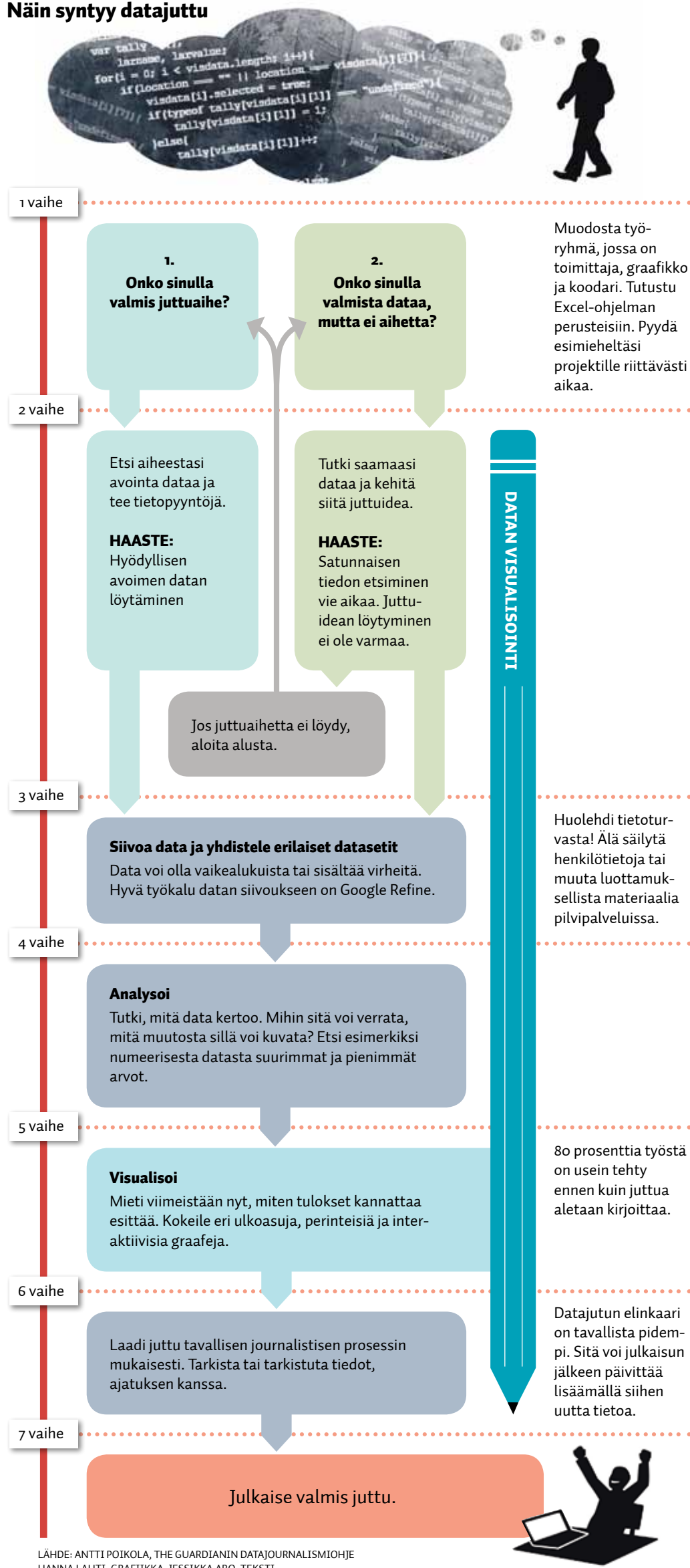
Valmiista verkostosta Tebest piilotti kaikkien nimet.

Jos Tebest tekisi analyysin uudestaan, hänen lähtökohtansa olisi journalistisempi.

”Tekisin jutun tiiviimmässä yhteistyössä uutistoimituksen kanssa ja hakisin juttuun laajemman näkökulman esimerkiksi epäillyn fb-tykkäyksen kautta. En osannut odottaa, että tekemäni analyysi olisi ollut niin kiinnostava kuin siitä tuli.”

JATKUU SEURAAVALLA SIVULLA

Näin syntyy datajuttu



LÄHDE: ANTTI POIKOLA, THE GUARDIANIN DATAJOURNALISMOHJE
HANNA LAHTI, GRAFIikka, JESSIKKA ARO, TEKSTI

Hans Rosling: Tee datajutusta helppokäyttöinen!

Hyvä datajournalismi on käyttäjäystävällistä, sanoo maailmankuulu ruotsalainen tilastotieteilijä **Hans Rosling**.

”Esimerkkiä datajuttujen tekemiseen voi katsoa *Angry Birds* -pelistä – sitä voi pelata lukematta ohjeita”, Rosling sanoo.

Roslingia kannattaa kuunnella, sillä hän tietää, miten ihmiset saa kiinnostumaan datasta: hänen globaalia väestökehitystä kuvaavia, animoituja ja interaktiivisia grafiikoita sisältäviä videoita on katsottu verkossa miljoonia kertoja.

Roslingin TED-konferenssissa vuonna 2006 pitämän puheen *Stats that Reshape your World View* on katsonut yli neljä miljoonaa ihmistä, ja tänä vuonna *Time*-lehti valitsi Roslingin yhdeksi maailman vaikutusvaltaisimmista henkilöistä.

Rosling johtaa Gapminder-säätiötä. Säätiö tunnetaan muun muassa Trendalyzer-ohjelmasta, jolla monimutkaisesta datasta voi luoda hyvännäköisiä ja kiinnostusta herättäviä infografiikoita.

Roslingin mukaan hyviä grafiikoita tarvitaan selittämään monimutkaisia ilmiöitä. Parhaiten se onnistuu, kun toimituksessa muodostetaan toimittajan, numeronmurskaajan ja kuvittajan tiimi.

”Toimittajat eivät yleensä hallitse Exceliä hyvin. Siksi on tärkeää, että tiimissä on mukana Excel-osaaja ja graafikko.”

Datajournalismin tekijän kannattaa myös unohtaa journalistille tyypillinen tarve luoda dramaattisia vastakkainasetteluita.

”Toimittajat etsivät juttuihin usein dramatiikkaa haastattelemalla ilmiöiden kannattajia ja vastustajia. Tämän sijaan toimittajien pitäisi haastatella henkilöitä, jolla on käsiteltävästä aiheesta totuudellinen, oikea tieto”, hän sanoo.

”Monilla on mielipiteitä ja näkemyksiä asioista. Etsi käsiisi se, joka tietää faktat.” ■

Jessikka Aro

ROSINGIN KUULUISA PUHE: [HTTP://WWW.TED.COM/TALKS/HANS.ROSING.SHOWS.THE.BEST.STATS.YOU.VE.EVER.SEEN.HTML](http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html)



Lukijat etusijalle. Maailmankuulun tilastonikkarin Hans Roslingin mielestä hyviä grafiikoita tarvitaan selittämään monimutkaisia ilmiöitä.

Haastateltavien parhaat datajournalismilinkit. **journalisti.fi**