

Eetu Halonen

Miksi datajournalismista tulisi (Suomessakin) kiinnostua

Miksi datajournalismista tulisi (Suomessakin) kiinnostua

Eetu Halonen
Opinnäytetyö
Kevät 2016
Viestinnän tutkinto-ohjelma
Oulun ammattikorkeakoulu

TIIVISTELMÄ

Oulun ammattikorkeakoulu
Viestinnän tutkinto-ohjelma, journalismin suuntautumisvaihtoehto

Tekijä(t): Eetu Halonen

Opinnäytetyön nimi: Miksi datajournalismista tulisi (Suomessakin) kiinnostua

Työn ohjaaja: Pertti Sillanpää

Työn valmistumislukukausi- ja vuosi: Kevät 2016

Sivumäärä: 88 + 6

Tämä tutkielma pyrkii perustelemaan, miksi datajournalismista tulisi Suomessakin kiinnostua. Datan analyttinen käyttäminen journalismin tekemiseen on Suomessa tiettävästi kahden kauppa: vain Yle ja Helsingin Sanomat tekevät datajournalismia rutiininomaisesti. Maailmalla ja erityisesti Yhdysvalloissa datajournalismilla on pitkät perinteet: vastaavia työmenetelmiä on käytetty jo 1800-luvulla osana tutkivaa journalismia. Tuolloin sitä vain ei kutsuttu datajournalismiksi, eikä sitä pitäisi kutsua datajournalismiksi tänäkään päivänä.

Tämän tutkielman tarkoitus on paitsi esittää vahvoja perusteluja, miksi datan käyttämisestä tulisi toimituksissa kiinnostua, myös selittää, miksi koko datajournalismin termistä tulisi luopua. Siinä missä tietokoneavusteinen journalismi on kuulostanut hassulta viimeistään 1990-luvun lopulta lähtien, on datajournalismista puhuminen tänä avoimen datan ja loputtoman informaation aikakautena turhaa. Datan kerääminen, analysoiminen ja esittäminen tulisi olla ihan kaikenlaisen journalismin kulmakivi, eikä eriytnyt journalismin osa-alue. Muun muassa näille väitteille esitetään tämän tutkielman sivuilla perusteluja.

Tutkielmassa käsitellään laajasti datan journalistista käyttämistä koskevaa kirjallisuutta, artikkeleita ja raportteja niin Suomesta kuin muualta maailmalta. Lisäksi on haastateltu datajournalistisia työmenetelmiä vakituisesti käyttäviä toimittajia Suomesta, Ruotsista ja Englannista. Tämän aineiston pohjalta etsitään perusteluja, miksi datan käyttäminen journalismissa kannattaa. Useiden juttuesimerkkien ohella tehdään tarkempi katsaus Miami Heraldin Pulitzerilla palkittuun Hurrikaani Andrew'ta käsittelevään juttuun vuodelta 1992. Se on yhä tänä päivänä, lähes neljännesvuosisata ilmestymisensä jälkeen erinomainen esimerkki data-analyysin ja perinteisen tutkivan journalismin yhdistämisestä.

Lopuksi käsitellään joulukuussa 2015 tehtyä kyselytutkimusta, johon vastasi yhdeksän suomalaista maakuntalehteä. Tutkimuksella kartoitettiin datajournalististen työmenetelmien yleisyyttä maakuntalehdistössä. Tuloksista kävi ilmi, että toimituksissa on selvää kysyntää datajournalismille, mutta taitoa ja aikaa kehittää näitä työmenetelmiä puuttuu.

Asiasanat: journalismi, data, avoin data, datajournalismi, tietokoneavusteinen journalismi, sanomalehti, maakuntalehti, Suomi

ABSTRACT

Oulu University of Applied Sciences
Degree Programme in Communication, Option of Journalism

Author(s): Eetu Halonen

Title of thesis: Why newsrooms should be interested in data journalism (in Finland as well)

Supervisor(s): Pertti Sillanpää

Term and year when the thesis was submitted: Spring 2016 Number of pages: 88 + 6

This thesis is trying to answer why newsrooms – in Finland as well – should be interested in data journalism techniques. The term 'data journalism' was coined and brought into a more common discourse in the end of the first decade of the 21st century. As we are well into the second, it is timely for a study like this to take place.

The purpose of this study is to present strong arguments as to why data analysis should be an integral part of news organization's journalistic practices. In addition, this thesis is trying to explain why the term data journalism itself is somewhat obsolete in this grand age of data and information. As was the case with the term and techniques related to 'computer-assisted reporting' in the end of the 20th century, the techniques for gathering, analyzing and presenting data as a part of contemporary journalism have and should become axiomatic. We don't speak of pencil-assisted reporting or telephone driven journalism, and we shouldn't speak of computer-assisted reporting or data driven journalism either. Furthermore, as this study shows, the use of data to enhance journalism has a long history spanning over centuries. When it comes to data visualization, one must only consider the very first maps crafted hundreds of years ago.

To address these issues, this thesis draws from existing literature covering the topic of using data in journalism. Additionally, few professional journalists from Finland, Sweden and England, who regularly use data in their work, were picked to be interviewed for this study. Among many examples of stories and projects where data has been used successfully as a part of journalism, a closer look is taken at the Miami Herald's Pulitzer-prize winning story about Hurricane Andrew in 1992. It is still today, nearly a quarter of a decade later, a great example of combining data analysis to traditional shoe-leather reporting.

Finally, a survey was conducted to determine the use of data journalistic techniques in 9 regional newspapers in Finland. The results showed that although there is a clear demand for these practices in the regional press and some of the papers even use these techniques to some extent, the dearth of skills and time is preventing these news organizations from developing their data practices further.

Keywords: journalism, data, open data, data journalism, computer-assisted reporting, newspapers, Finland

SISÄLLYS

| | |
|--|----|
| 1 JOHDANTO | 6 |
| 2 KÄSITTEIDEN MÄÄRITTELYÄ..... | 8 |
| 2.1 Data | 8 |
| 2.1.1 Avoin data | 10 |
| 2.1.2 Avoin data Suomessa ja muualla | 13 |
| 2.2 Journalismi | 17 |
| 2.2.1 Journalistinen totuus..... | 18 |
| 2.2.2 Objektivisuus | 19 |
| 2.2.3 Itsenäisyys..... | 22 |
| 2.3 Datajournalismi | 25 |
| 3 DATA, JOURNALISMI JA TIETOKONEET – HISTORIALLINEN NÄKÖKULMA | 32 |
| 3.1 Ei mitään uutta auringon alla? | 32 |
| 3.2 Kohti tietokoneavusteista journalismia | 36 |
| 3.3 CARista datajournalismiin..... | 44 |
| 4 MIKSI JOURNALISMIN TULISI OLLA KIINNOSTUNUT DATASTA | 54 |
| 4.1 Dataa, dataa kaikkialla..... | 54 |
| 4.2 Parempaa journalismia | 56 |
| 4.3 Kuinka se tehdään oikein – esimerkkinä Miami Herald & hurrikaani Andrew..... | 61 |
| 5 TYÖKÄYTÄNNÖT SUOMALAISESSA MAAKUNTALEHDISTÖSSÄ..... | 63 |
| 5.1 Kysely maakuntalehtiin datan käyttämisestä | 65 |
| 5.2 Kyselyn tulokset..... | 66 |
| 5.3 Johtopäätökset | 73 |
| 6 LOPUKSI | 77 |
| LÄHTEET | 80 |
| LIITTEET | 89 |

1 JOHDANTO

Tämän tutkielman nimi on siinä mielessä harhaanjohtava, että suuri osa näiden sivujen sisällään pitämästä tekstistä käytetään kyseenalaistamaan datajournalismin käsitteellinen mielekkyys. Tämä termihän on noussut yleisempään keskusteluun ja ehkäpä tietoisuuteen ylipäätään vasta 2010-luvun tienoilla. Datajournalismilla on tarkoitettu monenlaisia asioita, ja sillä on jopa hieman mystinen kaiku: tämä on jotain uutta ja ihmeellistä. Ihmeellistä se ehkä onkin, mutta ei missään nimessä uutta: datajournalismia on tehty suurin piirtein yhtä kauan kuin journalismia.

Tämä tutkielma pyrkii osaltaan hälventämään datajournalismin ympärillä leijuvaa savuverhoa. Tavoitteena on nähdä pelkän uuden muotitermin taakse ja tuoda esiin näkemys, että eriytyneen journalismin lajityypin sijasta olisi tässä tapauksessa mielekkäämpää puhua ihan vain journalismista. Datan käyttämistä journalismin tekemisessä ei pitäisi enää – jos sitä koskaan on pitänytään – tällä vuosituhanella erikseen mainita. Datajournalismin termin syntyyn ovat vaikuttaneet kaikenlaisen datan parantunut saatavuus, datan analysoimiseen käytettävien työkalujen monipuolistuminen ja ylipäätään tietoteknisen osaamisen yleistyminen yhteiskunnassa. Nämä seikat eivät kuitenkaan tee dataan liittyvistä journalistisista työtavoista erikoisia, vaan päinvastoin: niiden tulisi olla arkipäivää.

Tutkielma aloittaa datajournalismin purkamisen siihen liittyvien kahden käsitteen, datan ja journalismin määrittelyllä, minkä jälkeen käydään läpi itse datajournalismille annettuja ja tarjottuja määritelmiä. Samalla herätellään kysymystä, onko datajournalismin ja journalismin välillä edes mitään käsitteellistä eroa.

Luvussa kolme tehdään aikamatka satojen vuosien taakse menneisyyteen ja sieltä takaisin tälle vuosituhanelle. Tämän tripin tarkoitus on osoittaa, että dataa, tietokoneita ja journalismia on yhdistelty toisiinsa varsin menestyksekkäästi jo kauan. Vaikka työmenetelmien periaatteet ovat pysyneet samoina ensimmäisistä datan visualisoinneista ja 1800-luvun tutkivasta journalismista lähtien, ovat tämän päivän tietotekniset harppaukset, internet ja valtavat datamäärät muuttaneet sen yhteiskunnan, jossa journalismi toimii. Journalismi ei kuitenkaan itsessään ole muuttunut, mutta sillä on käytössään enemmän resursseja kuin koskaan – niitä on vain haluttava käyttää.

Luvussa neljä päästetään ääneen journalismin ammattilaisia niin Suomesta kuin muualta maailmalta. Puheenvuorojen tarkoitus on koota yhteen perusteluja, miksi journalismin tulisi olla kiinnostunut datasta. Muiden tapausesimerkkien ohella käydään tarkemmin läpi yksi Pulitzer-palkittu datajournalistinen juttu vuodelta 1992.

Luku viisi käsittelee Suomen maakuntalehdistöön vuoden 2015 joulukuussa lähetettyä kyselytutkimusta, jolla kartoitettiin datan käyttöä ja siihen liittyviä näkemyksiä maakuntalehdissä. Tutkimuksen tulosten perusteella lehdissä olisi kysyntää datajournalistisille työmenetelmille, mutta syvempi ymmärrys datan käyttömahdollisuuksista puuttuu.

Tämä tutkielma ei ole käytännön opas datajournalismista. Näiltä sivuilta ei siis löydy ohjeita taulukkolaskentaohjelmien käyttämisestä, tilastollisesta analyysistä tai koodaamisesta. Näitä voidaan käyttää apuna journalismin tekemiseen, mutta kenenkään toimittajan ei ole pakko osata tätä kaikkea pystyäkseen käyttämään dataa.

Tämän tutkielman tarkoitus on johdatella lukija datan journalistiseen käyttämiseen historian ymmärtämisen kautta: kyseessä ei ole mikään uusi pinnallinen muoti-ilmiö, vaan pitkän perinteen työmenetelmä, jolla voidaan saada aikaan yksinkertaisesti parempaa journalismia.

2 KÄSITTEIDEN MÄÄRITTELYÄ

Tässä luvussa yritetään selvittää, mitä datajournalismilla oikeastaan tarkoitetaan. Termi koostuu kahdesta sanasta: 'datasta' ja 'journalismista'. Näin ollen katsaus lienee paras aloittaa näiden sanojen määrittelemisellä. Tätä kautta voidaan lähestyä itse datajournalismin määritelmää, joka osoittautuu hieman hankalaksi tapaukseksi. Herää kysymys, tulisiko koko termiä edes määritellä, ja puhua ihan vain pelkästä journalismista?

Tällä pyörittelyllä halutaan paitsi selvittää, mistä datajournalismissa on kysymys, myös ottaa kantaa viime vuosina puhjenneseen datajournalismin buumiin. Tarkoitus on tuoda esiin näkökulma, että tietyn, eriytyneen journalismin lajityypin sijasta olisi ehkä hedelmällisempää puhua työkaluista ja -menetelmistä, jotka nykyajan toimittajien tulisi ottaa laajemmin käyttöönsä.

2.1 Data

Sana *data* tulee alun perin latinan kielestä: *dare* tarkoittaa suomeksi 'antaa', *datum* tarkoittaa 'annettua' ja *data* on sen monikkomuoto eli 'annetut' (Niiniluoto 1996, 28). Monet filosofiset teoriat tuntevat termin 'miehelle annettu'; se on jotakin sellaista, joka on mielessä, mutta ei ole riippuvaista mielestä itsestään. Tällaisina 'annettuina' asioina on pidetty esimerkiksi havaintojen aistimussisältöjä, todellisuuden metafysisen rakenteen perusteita tai logiikan lakeja. Laajemmassa filosofisessa mielessä 'annetuksi' voidaan kutsua mitä tahansa ajattelun lähtökohtaa. (Korkman & Yrjönsuuri 1998, 440.) Tällainen lähestymistapa ei ole kaukana määritelmästä, jonka mukaan *data* on faktapohjaista informaatiota, jota käytetään järkeilyn, keskustelun tai laskemisen perustana (Merriam-Webster 2016, viitattu 3.2.2016).

Suomen kielessä termit *data*, *tieto* ja *informaatio* ovat ainakin arkikäytössä lähestulkoon synonyymeja (Poikola, Kola & Hintikka 2010, 13). Kielitoimiston sanakirja (Kotimaisten kielten keskus 2016, viitattu 8.2.2016) määrittelee tiedon olevan muun muassa "asia - - joka jostakin tiedetään tai annetaan, saadaan tietää; informaatio, *data*". *Datan* määritelmä on kuitenkin hitusen rajatumpi: "[jonkin] asian säännönmukainen esitys viestitettävässä tai käsittelykelpoisessa muodossa; tieto, tiedot" (ibid.). Myös Sanastokeskus määrittelee, että *data* on nimenomaan koneellisesti luettavassa, viestittävässä tai käsiteltävässä muodossa olevaa tietoa. Määritelmässä

kuitenkin huomautetaan, ettei yleiskielessä sanoilla data, informaatio ja tieto ole useinkaan selvää eroa, ja että tieto-sanalla voidaan viitata myös dataan ja informaatioon. (Sanastokeskus TSK:n termipankki 2016, viitattu 9.2.2016.)

Ilmeisesti ainakin vielä 1990-luvulla Kotimaisten kielten keskus suositteli datan käännökseksi sanaa *anne* (Niiniluoto 1996, 29). Niiniluoto kuitenkin huomauttaa, että *anne* on terminä huomattavasti vaikkapa tietoa kömpelömpi. Vai miltä 'automaattinen anteiden käsittely', 'annekone', 'anteiden siirto' tai 'annesuoja' kuulostaisivat? Sillä olisi toisaalta ollut se etu, että se ei ainakaan olisi mennyt päällekkäin minkään muun käsitteen kanssa: vaikka 'data' suomennettaisiin 'tiedoksi', ei tieto aina käänny takaisin dataksi. Esimerkiksi filosofinen tieto-oppi ei ole 'dataoppia', tietokilpailu ei ole 'datakilpailu', eikä Suomen tietotoimisto ole 'Suomen datatoimisto'. (Ibid.)

Filosofian tohtorin Veli Valpolan Suuressa sivistyssanakirjassa datalle on annettu kaksi määritelmää. Ensimmäisen mukaan datalla tarkoitetaan erityisesti tiettyihin ajankohtiin liittyviä tosiasioita ja tietoja sekä havaintojen ja kokeiden välittömiä, suoria tuloksia ilman teoreettisia tai yleistäviä tulkintoja. Toinen määritelmä sanoo, että data on varsinkin tiedonsiirrossa tiedon, käskyn tai muun sellaisen täsmällinen esitys sovitun säännönmukaisen merkkijärjestelmän (koodin) mukaisessa muodossa. (Valpola 2000.)

Valtion- ja kunnallishallinnon tietohallintoa koskevat menettelytavat, ohjeet ja määrittelyt löytyvät kootusti JHS-järjestelmästä. Järjestelmän suositus numero 189 koskee julkishallinnon avointen tietoaineistojen käyttöä. Siinäkin määritellään data koneellisesti käsiteltävässä muodossa olevaksi tiedoksi, minkä lisäksi suosituksessa mainitaan vielä erikseen *raakadata*:

Kun raakadataa käsitellään ja tulkitaan, siitä voi syntyä informaatiota ja lopulta tietoa. Raakadataa ovat esimerkiksi puheen äänneet, painetun tekstin kirjaimet, bitit tietokoneessa, ilmiötä koskeva tallennettu numerotieto, biologinen data ja yksittäiset tosiseikat. Tilastot, julkaisut, videotallenteet, kuvat, kartat ja 3d-mallit sisältävät raakadataa. Raakadatan ja informaation ero on suhteellinen. Raakadataa käsitellään monesti useassa vaiheessa, ja yhden vaiheen käsitelty data saattaa olla raakadataa seuraavalle vaiheelle. (JHS 189 avoimen tietoaineiston käyttöä, viitattu 9.2.2016.)

Suomen julkishallinto tekee näin pesäeron datan, informaation ja tiedon välillä (vrt. Poikola ym. 2010, 13). Näiden termien merkitysten keskinäinen hierarkia kulkee kyseisessä määrittelyssä seuraavasti: data → informaatio → tieto, missä tieto on jotakin sellaista, joka saavutetaan vasta,

kun datalla ja informaatiolla on ensin tehty jotain. Samainen JHS-suositus määrittelee tiedon olevan paitsi dataa ja informaatiota, myös näiden perusteella syntyvää *tietämystä* (ibid.).

Myös informaatiotutkimuksen piirissä puhutaan tiedon arvoketjusta, jossa irrallisesta informaatiosta jalostuu ihmiselle käytännössä hyödyllistä tietoa. Tämä arvoketju vie edellä mainittujen termien merkitysten välisen hierarkian vielä pidemmälle; siinä datan ja informaation pohjalta saatu tieto jalostuu tietämykseksi ja edelleen *viisaudeksi* (Haasio & Savolainen 2004, 14).

Erityisesti tieteellisessä käytössä datan, informaation ja tiedon merkityksillä on eroja. Se, mitä vaikkapa tiedolla tarkoitetaan informaatiotutkimuksessa, ei ole välttämättä sama asia kuin tiedotusopissa tai tietojenkäsittelyopissa. (ibid.) Informaatiotutkimuksen piirissä datalla tarkoitetaan merkeistä ja symboleista koostuvaa potentiaalista informaatiota; data on ikään kuin raaka-ainetta, josta *saattaa* tulla informaatiota. Ja edelleen data voidaan mieltää asiayhteydestään irrallisiksi faktoiksi, joista ihmisen toiminnan seurauksena muodostuu informaatiota, jota voidaan vaikka välittää muille ihmisille. (Haasio & Savolainen 2004, 15.)

Poikola ym. (2010) ovat määrittäneet datan tarkoittavan kaikkea digitaalisesti tallennettua informaatiota. Lisäksi he korostavat, että esimerkiksi julkishallinnon tulisi tarjota informaatiota tietokoneiden avulla käsiteltävässä muodossa, ja viittaavat Euroopan tietoverkkorikollisuutta koskevaan yleissopimukseen. Sen mukaan data tarkoittaa ”tosiseikkojen, tietojen tai käsitteiden esitystä sellaisessa muodossa, että se soveltuu käytettäväksi tietojärjestelmässä, mukaan lukien ohjelmat, joiden avulla tietokone pystyy suorittamaan jonkin toiminnon”. (TEM 2009, Poikolan ym. 2010, 14 mukaan.)

2.1.1 Avoin data

Vaikka datan määrittely osoittautuu ehkä luultua hankalammaksi, on *avoin data* käsitteenä suoraviivaisempi. Kuutin (2012, 13) mukaan avoin data on julkishallinnolle ja muille organisaatioille kertynyttä jalostamatonta informaatiota, johon ulkopuolisilla on vapaa pääsy. Datan avoimuuteen vaikuttavat sen tekninen saatavuus, maksuttomuus, uudelleenkäytön sallivat käyttöehdot, löydettävyys ja ymmärrettävyys. (Ibid.)

Tekninen saatavuus edellyttää, että data on ylipäättään julkaistu sellaisessa muodossa, että sitä on mahdollista käsitellä tietoteknisesti (ks. Poikola ym. 2010). *Maksuttomuus* taas mahdollistaa datan hyödyntämiseen liittyvät kokeilut ilman budjettibyrokraatia. Julkaisijan on sallittava datan *uudelleenkäyttö* ja tuotava tämä oikeus selkeästi esille aineiston käyttöehdoissa. Data-aineiston *sijainti* ja tieto aineiston *olemassaolosta* on oltava yleisesti tunnettu; data voi sijaita esimerkiksi julkisessa datakatalogissa. Lisäksi datan rakenne ja merkitys on kuvailtava käyttäjille selväsanaisesti. Ilman tällaista kuvailua data uhkaa jäädä merkityksettömäksi, eikä sitä voida hyödyntää. (Kuutti 2012, 13.)

(Huom. Edellä mainittujen määritelmien mukaan avoin tarkoittaa tässä yhteydessä nähdäkseni samaa asiaa kuin julkinen. Tulen käyttämään jatkossa myös tätä sanaa avoimen sijasta. Tällä tavalla vältetään käyttämästä liikaa avoin-sanana johdannaista *avoimuus*, jonka voi mielestäni aivan hyvin korvata sanalla julkisuus.)

Data-aineiston julkisuutta voidaan arvioida vertaamalla sitä tiettyyn avoimen datan kriteeristöön. Tämä kriteeristö pohjautuu yleisesti hyväksytyyn *avoimen tiedon määritelmään* (engl. Open Knowledge Definition). (Poikola ym. 2010.) Avoimen tiedon määritelmä on puolestaan johdettu amerikkalaisen ohjelmoijan Bruce Perensin *avoimen lähdekoodin* määritelmästä (Open Knowledge 2016, viitattu 11.2.2016).

Avoimen tiedon määritelmän mukaan (Open Knowledge Finland 2016, viitattu 11.2.2016) avoimen ja vapaan tiedon pääperiaatteita ovat:

- Aineiston tulee olla kokonaisuudessaan saavutettavissa ja ladattavissa julkisessa tietoverkossa.
- Tiedon on oltava uudelleenjaettavissa ilman käyttöehtorajoituksia. Näin mahdollistetaan aineistojen nopea hyödyntäminen ja tehokas yhdistely.
- Tiedon on oltava uudelleenkäytettävissä. Näin sallitaan aineistojen esteetön ja innovatiivinen käyttö edistyksellisiin ja yllättäviinkin tarkoituksiin.
- Aineiston on oltava vapaa teknisistä rajoitteista niin, ettei yllämainittujen kohtien mukaiselle toiminnalle ole teknisiä esteitä.
- Aineiston on oltava vapaa sosiaalisista ja organisatorisista rajoitteista, niin ettei henkilön työ, sijainti, asuinpaikka, organisaatiomalli (kuten kaupallinen tai ei-kaupallinen organisaatio), uskonto, poliittinen suuntautuneisuus tai etnisyyden rajoita pääsyä tietoon.

Lisäksi avoimen tiedon määritelmässä todetaan: ”Avoin tarkoittaa että kuka tahansa voi avoimesti päästä tietoon käsiksi, käyttää, muokata ja jakaa tietoa mihin tahansa käyttötarkoitukseen [sic]” (Open Knowledge Finland 2016, viitattu 11.2.2016).

Vaikka avoimella datalla tarkoitetaan usein juuri julkisen sektorin tietovarantoja, jotka on tuotettu veronmaksajien rahoilla, voi tällainen määritelmä olla liian laaja (Lehtonen 2011, 6). Esimerkiksi henkilöiden yksityisyyttä koskevaa tietoa ei voida julkaista avoimesti kaikille saatavaksi. Näin ollen *kaikki* julkishallinnon tieto ei voi olla avointa.

Tietovarantojen maksuttomaan ja vapaaseen julkaisemiseen verkossa liittyy erilaista maakohtaista lainsäädäntöä ja muuta sääntelyä. Tämän tutkielman puitteissa ei ole kuitenkaan olennaista syventyä lakipykäliin sen tarkemmin. Seuraavalla sivulla olevassa kuvassa on listattu lainsäädäntö, suositukset ja direktiivit, jotka Suomessa vaikuttavat julkishallinnon avoimeen dataan. Kuvan taulukkoa apuna käyttäen lukija voi halutessaan hakeutua kyseisten lakien ja säädösten pariin.

| | |
|----------------------------|---|
| Julkisuuslaki | Julkisuuslailla tarkoitetaan vuonna 1999 voimattomia lakia ja asetusta viranomaisten toiminnan julkisuudesta (1999/621) ja (1999/1030). |
| Tietosuojalainsäädäntö | Yleisnimityksellä tietosuojalainsäädäntö tarkoitetaan ainakin ensisijaisesti henkilötietolakia (523/1999) ja myös muita yksityisyyden suojaan liittyviä säädöksiä kuten rikoslain (24. luvun) pykälää yksityiselämän loukkaamisesta sekä lakia yksityisyyden suojasta työelämässä (2004/759). Viimeaikoina erityisesti on puhuttanut sähköisen viestinnän tietosuojalaki (2004/516) ja sen Lex Nokiaksi ristitty uudistus HE 48/2008 . |
| Maksuperustelaki | Valtion maksuperustelaki (1992/150) |
| Kansainväliset suositukset | OECD Recommendation of the Council concerning Access to Research Data from Public Funding (OECD 2006) OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information (OECD 2008) |
| Tekijänoikeuslaki | Tekijänoikeuslaki (1961/404) |
| Kilpailulainsäädäntö | Kilpailulainsäädännön lähtökohtana on taloudellisen kilpailun ensisijaisuus verrattuna markkinoiden säätelyyn. Nykyisin voimassaoleva laki on vuonna 1992 säädetty laki kilpailunrajoituksista (1992/480). |
| EU Direktiivit | INSPIRE direktiivi (2007/2/EY) tähtää paikkatietojen käytön tehostamiseen, viranomaisten yhteistyön lisäämiseen ja monipuolisten kansalaispalvelujen syntymiseen. Suomessa direktiivin toteuttamisesta on säädetty laissa (421/2009) ja asetuksessa (725/2009) paikkatietoinfrastruktuurista. PSI (Public Sector Information) direktiivi (2003/98/EY) julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä tähtää julkisen datan kaupallisen hyödyntämisen lisäämiseen. |

KUVA 1. Julkishallinnon avoimeen dataan liittyvät suositukset, direktiivit ja lainsäädäntö Suomessa (Poikola ym. 2010, 46)

2.1.2 Avoin data Suomessa ja muualla

Lohtanderin (2013, 5) mukaan käsite 'avoin data' liittyy yleisempään yhteiskunnalliseen keskusteluun avoimen, vapaan tai ilmaisen ideologioista. Ajatus vapaasta tiedosta ei ole mitenkään uusi. Jo vuonna 1984 julkaistun *hakkerietiikan* kolmas periaate kuuluu: "informaatio tahtoo olla ilmaista". (Ibid.)

Valtiot ovat kuitenkin vasta aivan viime vuosina alkaneet julkaista tietovarantojaan verkossa. Yhdysvallat avasi oman julkisen datan portaalin vuonna 2009, Norja vuonna 2010, Espanja 2011 ja Intia vuonna 2012. (Lohtander 2013, 5.)

Open Knowledge -yhteisön (OKF) avoimen datan indeksi pyrkii selvittämään, kuinka paljon julkishallinnot maailman eri puolilla ovat avanneet data-aineistojaan. Indeksien tiedot kerätään suoraan joukkoistamalla kansalaisilta ja järjestöiltä, mikä lisää OKF:n mukaan indeksin luotettavuutta, koska valtiot eivät ole itse esittämässä väitteitä tietovarantojensa julkisuudesta. (Open Knowledge 2015, viitattu 12.2.2016). Vuoden 2015 avoimen datan indeksiin oli kerätty tiedot 122 maasta. Katsauksen mukaan tietoaineistot olivat kyseisenä vuonna kaikkein julkisimpia Taiwanissa. Toisena oli Iso-Britannia ja kolmantena Tanska. Suomi oli sijalla viisi. (Open Data Index 2015, viitattu 12.2.2016.)

Myös World Wide Web -säätiö on julkaissut vuosina 2013 ja 2014 avoimen datan barometrin. Barometrillä on yritetty selvittää avoimen datan yleisyyttä ja todellisia vaikutuksia eri maissa. (Web Foundation 2015, viitattu 12.2.2016.) Barometrin esittelyssä kerrotaan:

[Avoimen datan barometri] analysoi maailmanlaajuisia kehityssuuntia ja tarjoaa vertailukelpoista tietoa eri maista sekä alueista syvällisten tutkimusmenetelmien avulla. [barometrissa] yhdistellään kontekstuaalista tietoa, teknisiä arviointeja ja toissijaisia indikaattoreita avoimen datan valmiuksien, toteutusten ja vaikutusten tutkimiseksi. (ibid.)

Vuoden 2014 avoimen datan barometrissa oli mukana 86 maata. Kärkikolmikko oli järjestyksessä Iso-Britannia, Yhdysvallat ja Ruotsi. Suomi oli sijalla 12. (Huom. Open Knowledge -yhteisön avoimen datan indeksin vuoden 2015 ykkönen Taiwan ei ollut barometrissa mukana lainkaan.) Suomen sijoitus barometrissa paranee merkittävästi, jos vertaillaan maiden ottamia edistysaskelia julkisen tiedon saralla. Verrattaessa vuosien 2013 ja 2014 arvosanoja, on Suomi parantanut kokonaisarvosanaansa kolmanneksi eniten. Vielä enemmän dataansa olivat avanneet Chile ja Indonesia. (Ibid.) Barometrin tuloksia erittelevässä raportissa todetaan, että Suomen kokonaisarvosana oli noussut merkittävästi verrattuna vuoden 2013 barometriin. Esimerkiksi tietoisuus avoimesta datasta on kasvanut Suomessa niin median ja kansalaisyhteiskunnan toimijoiden keskuudessa kuin tietyissä yritysmaailman piireissäkin. Avoimen datan vaikutus näytti ylipäätään olevan Suomessa kasvussa. (Ibid.)

Suomessa on tehty julkishallinnon tasolla lukuisia linjauksia avoimen datan suhteen. Viestintä- ja liikenneministeriön alaisuudessa vuosina 2008–2011 toiminut *Arjen tietoyhteiskunnan neuvottelukunta* nosti jo vuonna 2008 esiin kysymyksen julkisten tietovarantojen käyttömahdollisuuksista. (Lehtonen 2011, 4.) Valtioneuvosto on tehnyt 3.3.2011 ja 5.10.2011 periaatepäätökset, joiden mukaan julkishallinnon keräämän tiedon tulisi olla vapaasti saatavilla (Lehtonen 2011, 4; Kauhanen-Simanainen & Suurhasko 2015, 8). Pääministeri Jyrki Kataisen hallitusohjelmassa todettiin muun muassa näin:

Julkisin varoin tuotettuja tietovarantoja avataan kansalaisten ja yritysten käyttöön. Tavoitteena on julkisen sektorin hallinnoimien digitaalisten tietoaineistojen saattaminen helposti uudelleenkäytettävässä muodossa tietoverkkojen kautta kansalaisten, yritysten ja yhteisöjen, viranomaisten, tutkimuksen ja koulutuksen hyödynnettäväksi. (pääministeri jyrki kataisen hallituksen ohjelma 2011, 50.)

Julkisen hallinnon ICT:n strategiassa 2012–2020 mainitaan yhtenä alueena avoin tieto ja tiedon yhteiskäyttö. Strategian mukaan julkisia tietoja avataan Suomessa kaikissa hallinnon organisaatioissa systemaattisesti ja mahdollisimman nopeasti. Digitaalisen talouden tulevaisuutta pohtineen ICT 2015 -työryhmän raportissa puolestaan ehdotettiin *avoimen tiedon ohjelman* käynnistämistä ja kansallisen *avoimen datan portaalin* perustamista. (Kauhanen-Simanainen & Suurhasko 2015, 8.) Valtiovarainministeriö asettikin vuonna 2013 avoimen tiedon ohjelman koordinoimaan julkishallinnon tietovarantojen avaamista. Ohjelman toimikausi oli kaksivuotinen, ja sen loppuraportti julkaistiin syyskuussa 2015. Ohjelman myötä Suomessa otettiin vuonna 2014 käyttöön yllä mainittu kansallinen avoimen datan portaali *avoindata.fi*. (Ibid., 9; Valtion tieto- ja viestintätekniikkakeskus Valtori 2016.)

Julkinen data on mainittu myös Juha Sipilän hallituksen ohjelmassa: “Luodaan avoimella datalla ja tietovarantojen paremmalla hyödyntämisellä edellytyksiä uusille liiketoimintaideoille” (Pääministeri Juha Sipilän hallituksen strateginen ohjelma 2015, 27).

Julkisesta datasta on niin ikään tehty linjauksia Euroopan Unionin tasolla. Euroopan komissio julkaisi vuonna 2011 *eurooppalaisen avoimen datan strategian*. Sen tavoitteena on edistää julkisen datan hyödyntämistä jäsenmaissa muun muassa tutkimusohjelmien ja infrastruktuurihankkeiden avulla. Tiedotteessaan komissio arvioi strategian merkitsevän Euroopan Unionin taloudelle 40 miljardin euron kasvua joka vuosi. (Kauhanen-Simanainen & Suurhasko 2015, 8; Euroopan komissio 2011, viitattu 13.2.2016.)

Julkishallinnon datasta on säädetty EU:n direktiivillä vuonna 2003. Tämän julkishallinnon tietojen uudelleenkäyttöä koskevan niin sanotun PSI-direktiivin (2003/98/EY) tarkoitus on ollut helpottaa jäsenmaiden tietovarantojen käyttöä myös muihin tarkoituksiin kuin niihin, joita varten tiedot on alun perin kerätty. (Kauhanen-Simanainen & Suurhasko 2015, 8.) Direktiiviin annettu muutos (2013/37/EU) tuli voimaan 17.7.2013. Muutoksessa painotettiin, että kaikkia direktiivin soveltamisalaan kuuluvia tietoja tulee voida käyttää uudelleen sekä kaupallisiin että muihin tarkoituksiin, kunhan näiden tietojen saatavuutta ei ole muutoin kansallisesti rajoitettu. Soveltamisalaa myös laajennettiin koskemaan kirjastoja, arkistoja ja museoita. Direktiivin virallinen täytäntöönpanon määräaika oli jäsenmaissa 18.7.2015. (Ibid.)

Valtiovarainministeriön avoimen tiedon ohjelman loppuraportin mukaan Suomen lainsäädäntö mahdollistaa PSI-direktiivin tavoitetason saavuttamisen. Suomessa on vuodesta 2011 avattu viranomaisten hallussa olevia maastotietoja, sää-, ilmasto- ja meridataa, ilmastomalleja, liikennedataa, ajoneuvotietoja, tilastoaineistoja, julkisia verotietoja, maaperätietoja, talousdataa ja kulttuuriaineistoja. Tietovarantojen avaamista on vuodesta 2012 lähtien käsitelty valtion julkisen talouden suunnitelman valmistelun yhteydessä. Valtiovarainministeriö on pyytänyt muilta ministeriöiltä suunnitelmat, mitä tietovarantoja hallinnossa avataan ja mitkä ovat avaamisen taloudelliset ja yhteiskunnalliset vaikutukset. (Ibid.)

Myös useat kunnat ovat alkaneet avata data-aineistojaan. Pääkaupunkiseutu on ollut tässä edelläkävijä Helsinki Region Infoshare -palvelullaan, joka julkaistiin vuonna 2011. Lisäksi Helsinki, Espoo, Vantaa, Tampere, Turku ja Oulu ovat vuonna 2014 käynnistyneen 6Aika-hankkeen myötä avanneet tietovarantojaan. (Ibid.; 6Aika, viitattu 15.2.2016.)

2.2 Journalismi

Journalismi on joukkoviestintää. Joukkoviestintä tai joukkotiedotus on puolestaan lyhyen määritelmän mukaan sanomien lähettämistä ennalta rajaamattomalle ja suhteellisen suurelle vastaanottajajoukko. Joukkoviestinnässä käytetään teknisiä apuvälineitä, ja se on ainakin lyhyellä aikavälillä yksisuuntaista. (Kunelius 2009, 17; Hemánus 1995, 27–28.)

Perinteisen määritelmän mukaan journalismi on ajankohtaista ja faktapohjaista joukkoviestintää. Lisäksi nykyjournalismi määritellään yhtäältä itsenäisen journalistisen harkinnan myötä syntyväksi joukkoviestinnäksi, mutta toisaalta sen pitäisi myös edustaa suurta yleisöä. (Kunelius 2009, 17, 23.) Journalismi on sekä työprosessi että tuon prosessin lopputulos; jotkut tekevät journalismia, ja toiset ottavat tuon työn hedelmät vastaan (Hemánus 1995, 28).

Kuutin (2012, 68) mukaan sanan 'journalismi' alkuperä viittaa "päiväkirjan pitämiseen yhteiskunnallisista tapahtumista"; ranskan sana 'le jour' tarkoittaa 'päivää', ja 'journal' viittaa päiväkirjaan. Journalismin määrittely nimenomaan *ajankohtaiseksi* joukkoviestinnäksi lienee lähtöisin juuri tällaisesta ajattelutavasta. Tosin esimerkiksi Kunelius (2009, 21) huomauttaa, että ajankohtaisuus on kriteerinä melko väljä: esimerkiksi talvisota voi hyvinkin olla ajankohtainen useita vuosikymmeniä päättymisensä jälkeen.

Journalismiksi voidaan myös laskea laajemmin journalististen *toimintaedellytysten* – kuten taloudellisten toimintaedellytysten – luominen yksittäiselle tiedotusvälineelle, eli journalistista tuotantoa koskeva kustannus- ja lähetystoiminta (Kuutti 2012, 68). Edelleen väljästi määriteltynä tietynlainen kirjallisuus voidaan katsoa journalismiksi. Tällaista kirjallisuutta on erityisesti ajankohtainen, dokumentaarinen ja faktapohjainen yhteiskunnallinen kirjallisuus myös sellaisissa tapauksissa, joissa kirjoittaja ei ole ammattijournalisti. (Ibid.)

Journalismi voidaan myös määritellä sille annettujen tehtävien kautta. Näitä ovat Kuutin (ibid.) mukaan totuudenmukaisen ja olennaisen tiedon välitys, yhteiskunnan kriittinen tarkkailu ja yhteiskunnallisen vallankäytön valvonta sekä sananvapauden ja julkisen keskustelun edistäminen. Journalismi voidaan nähdä maailman tapahtumien ja kehityskulkujen tarkkailijaksi, kuvaajaksi, selittäjäksi ja arvioijaksi pyrkimyksensä kuvata todellisuutta mahdollisimman hyvin.

Journalismin tärkeimmäksi tehtäväksi esitetään usein juuri päätöksentekojärjestelmän valvontaa (Kunelius 2009, 193). Journalismia on usein kutsuttu 'neljänneksi valtiomahdiksi'. Ajatus juontuu vallan kolmijako-opista, jossa julkinen valta on jaettu lakien säätämisen (eduskunta), toimeenpanon (valtioneuvosto) ja tulkinnan (tuomioistuimet) osalta toisistaan riippumattomille elimille. Neljäntenä valtiomahtina journalismin pitäisi valvoa kolmen muun valtiomahdin toimintaa. (Kunelius 2009, 194.)

Edellä kuvaillut määritelmät journalismista ovat tämän tutkielman puitteissa riittäviä, eikä termin problematisointi ole tässä yhteydessä tarpeellista. Myöskään 'totuuden' tai 'tiedon' laajempi filosofinen pohdinta esimerkiksi tieto-opin tai tietoteorian näkökulmasta ei ole tutkielmassa olennaista. Seuraavaksi käydään kuitenkin läpi sitä käsitystä totuudesta, jonka kanssa journalismi on tekemisissä ja niitä käytäntöjä, joiden kautta journalismi pyrkii kohti totuutta.

2.2.1 Journalistinen totuus

Kuten edellä sanottiin, journalismi perustuu faktapohjaisten eli *tosiasiallisten* viestien välittämiseen. Määrittely ei ole kuitenkaan täysin ongelmaton. Sukeltamatta sen syvemmälle faktapohjaisuuden, faktan, tiedon ja totuuden käsitteisiin todettakoon ainoastaan, että nämä käsitteet voivat olla erittäin hankalia määrittellä. (Hemánus 1995, 29.) Kuinka journalismi siis on kiertänyt tämän ongelman?

Lyhyt vastaus on, että kiinnittämättä siihen liiemmin huomiota. Journalismi ei nimittäin edes pyri tavoittelemaan totuutta sanan absoluuttisessa tai filosofisessa mielessä (American Press Institute 2016, viitattu 18.2.2016). Pikemminkin journalismi etsii "käytännönläheistä ja toimivaa muotoa totuudesta"; se pyrkii kohti sellaisia totuuksia, joiden parissa journalistien on mahdollista työskennellä päivittäisessä työssään (ibid.). Toisin sanoen: journalistinen totuus voidaan mieltää maalaisjärkeen perustuviksi käsityksiksi siitä, mikä on totta (Kuutti 2015, 11).

Journalistiseksi totuudeksi riittää siis käsitys ulkoisesta maailmasta. Tässä mielessä yksittäinen lehtijuttu on totta, jos se kuvailee uskollisesti kohdettaan ja asioiden tilaa todellisessa maailmassa ja jos sillä on selkeä vastaavuus tämän todellisen ympäröivän maailman kanssa. (Ibid.) Tällainen käsitys journalistisesta totuudesta nojaa totuuden *koresspondenssi-* eli vastaavuusteoriaan, jonka mukaan mikä tahansa väite on totta silloin ja vain silloin, jos sen kuvaama asiantila vastaa

todellisuutta. Teorian heikkous on kuitenkin siinä, että todellisuuden ehdoton hahmottaminen on mahdotonta; ei ole olemassa absoluuttista pistettä, josta todellisuus avautuisi täydellisen alastomana, jolloin sitä pystyisi vain havainnoimaan ilman ongelmia. Todellisuus on aina enemmän tai vähemmän hämärän peitossa, ja tästä syystä myös mikä tahansa todellisuutta kuvaava väittäjä on totuusarvoltaan hämärä. Tämä ei kuitenkaan tarkoita, etteikö tajunnastamme riippumatonta todellisuutta olisi olemassa: esimerkiksi ympäristö saastuu, vaikka joku muuta väittäisikin. (Hemánus 1995, 35.)

Hemánus totesikin Tampereen yliopiston jäähyväisluennossaan, että ”journalismi toimii ja on tuomittu toimimaan sellaisen otaksunnan varassa, että niin sanottua todellisuutta on [olemassa]” (Hemánus 1997, 8). Ongelmia syntyy, kun tätä todellisuutta pitäisi jotenkin ryhtyä kuvaamaan. Sen suhteen journalismin keinot rajoittuvat Hemánuksen mukaan jonnekin arki ajattelun ja tieteen välimaastoon, mutta kuitenkin selvästi lähemmäs arkista ajattelua. Esimerkiksi niinkin yksinkertaiselta kuulostavan asian *syvällinen* uutisoiminen kuin kahden auton yhteentörmäys ja tämän kolarin *ymmärtäminen* edellyttäisivät Hemánuksen mukaan asian taustoittamista hyvin monesta suunnasta. Huomioon tulisi ottaa muun muassa auton tekniset ominaisuudet, niiden kehitys ja historia; liikennepsykologia ja sen ongelmat; ihmisen psykofysiikan mahdollisuudet ja rajoitukset auton ajamisessa; liikennesääntöjen juridiikka ja niiden kehitys; auton ja autoilun symbolinen merkitys ja niin edelleen. Kolarin syvällinen kuvaaminen vaatisi enemmän paperia kuin maailman metsävarat sallisivat, ja juuri tässä on Hemánuksen mukaan journalismin yksi ongelma pähkinänkuoressa: se pyrkii todellisuuden kuvaamisen suhteen tavoitteisiin, jotka ovat mahdottomia. Journalismi pystyy näin ollen tuottamaan vain karkeita *likiarvoja* todellisuudesta. (Ibid.)

Vaikka niin sanottua täydellistä totuutta ei olisikaan mahdollista paljastaa journalistisen toiminnan kautta, voidaan asioiden todellisen tilan selvittämistä kuitenkin lähestyä. Epätodennäköisiä tai vähemmän mahdollisia selityksiä kerrottavien asioiden tilasta voidaan karsia muun muassa tiedonhankinnan, tietojen arvioinnin ja arvottamisen pohjalta. (Kuutti 2015, 11.)

2.2.2 Objektiivisuus

“Journalistin velvollisuus on pyrkiä totuudenmukaiseen tiedonvälitykseen”, sanotaan Journalistin ohjeiden kohdassa numero kahdeksan. (Julkisen sanan neuvosto 2016, viitattu 17.2.2016.) Muita

tietojen oikeellisuuteen liittyviä journalistin eettisiä ohjeita ovat kohdat 10, 11, 12, 13, 15 ja 16 sekä joissain määrin kohdat 2, 3, 4, 5, 6 ja 7. Tämä eettinen koodisto on yksi apuväline journalistien päivittäiseen työhön liittyvien ongelmien ratkaisemisessa, mutta se tarjoaa kuitenkin melko vähän työkaluja totuuden jäljittämiseksi: ohjeissa todetaan, että totuutta on tavoiteltava, mutta siellä ei kerrota, kuinka tähän tavoitteeseen käytännössä päästään.

Ohjeiden kohta 13 on totuuden näkökulmasta erityisten mielenkiintoinen: "Uutisen voi julkaista rajallistenkin tietojen perusteella. Raportointia asioista ja tapahtumista on syytä täydentää, kun uutta tietoa on saatavissa. Uutistapahtumia on pyrittävä seuraamaan loppuun saakka." (Ibid.) Journalistinen ammattietiikka antaa täten – ainakin Suomessa ja nimenomaan uutisten kohdalla – jonkin verran liikkumatilaa journalismin totuusvaatimuksen suhteen. Tästä huolimatta totuuteen on joka tapauksessa *pyrittävä*, eivätkä 'rajalliset tiedot' myöskään ole sama asia kuin esimerkiksi 'väärät tiedot'.

Journalistit ovat kehittäneet erilaisia strategioita ja työvälineitä totuutta kohti kulkevan ristiretkensä tueksi. Esimerkiksi objektiivisuuden ihanne, tieteellisen metodin matkiminen, läpinäkyvyyteen pyrkiminen ja kahden tai useamman lähteen vaatimus tietojen tarkistamiseksi ovat keinoja, joiden avulla journalistit voivat elää totuuden huteran perustan kanssa. (Kuutti 2015, 12.)

Perti Hemánus ja Ilkka Tervonen ovat aikanaan perehtyneet varsin syvällisesti ja varmaankin kaikista kuviteltavissa olevista näkökulmista journalismin ja joukkoviestinnän objektiivisuuteen teoksessaan *Objektiivinen joukkotiedotus* (1980). Kyseisen tutkimuksen tarkempi erittely ei kuitenkaan liene tämän tutkielman puitteissa aiheellista. Hémanus on myös jälkeempään todennut, että ainakin osa tuon tutkimuksen esityksistä on jo aikansa elänyt, tai kaipaisi vähintään päivittämistä (Hemánus 1995, 37). Sanottakoon silti sen verran, että ko. teoksen määritelmä joukkoviestinnän objektiivisuudesta noudattaa pitkälti totuuden korrespondensiteoriaa. Eli Hemánuksen ja Tervosen (1980, 104) mukaan minkä tahansa sanoman, viestin tai väitteen on vastattava ulkoista todellisuutta ollakseen objektiivinen. Tosin tämäkään määritelmä ei ole heidän mielestään täysin tyydyttävä. Jos joukkotiedotus – ja journalismi – mielletään työprosessiksi ja tuon prosessin lopputulokseksi (ks. Hemánus 1995, 28; American Press Institute 2016, viitattu 18.2.2016), voidaan ottaa erikseen tarkasteluun sekä tämän prosessin objektiivisuus että sen lopputuloksen objektiivisuus, eivätkä ne ole välttämättä sama asia. Esimerkiksi prosessin ei tarvitse olla ainakaan sanan kaikissa merkityksissä objektiivinen, jotta lopputulos voisi olla sitä.

Vastaavasti prosessin objektiivisuus ei automaattisesti tarkoita lopputuloksen objektiivisuutta. (Hemánus & Tervonen 1980, 104.)

Kuutin (2012, 137) mukaan objektiivinen joukkoviestintä, johon journalismikin luonnollisesti osana kuuluu, tarkoittaa asioiden raportointia ilman journalistin henkilökohtaista subjektiivista katsomusta, näkemystä tai asennoitumista. Journalismin objektiivisuudessa on kyse siitä, kuinka todenmukaisesti, yleispätevästi, asiallisesti, puolueettomasti tai ei-värityneesti journalismi pystyy kertomaan ympäröivästä todellisuudesta. (Ibid.) Hemánus (1995, 37–38) on tosin huomauttanut, että objektiivisuudella ei ole juuri mitään tekemistä *esitystavan* kanssa. Esimerkiksi provosoiva pilapiirros tai törkeäsanainen televisiosketsi voi “ - - avata todellisuuteen todemman ja paremmin olennaisessa pitäytyvän näkökulman kuin sovinnainen journalismi ja tavanomainen ajattelu, jota sovinnainen journalismi myötäilee.” Yritys lähestyä todellisuutta ja *kertomisen ilo* eivät siis sulje toisiaan pois (Hultén 1993, Hemánuksen 1995, 38, mukaan).

Westerståhl (1983, Kuutin 2015, 12, mukaan) on jakanut objektiivisuuden joukkoviestinnässä *asiallisuuteen* eli kerrottavien tietojen todenmukaisuuteen ja olennaisuuteen sekä *puolueettomuuteen* eli tietojen tasapainoiseen ja neutraaliin esitystapaan. Todenmukaisuudella tarkoitetaan jo edelläkin mainittua tiedonvälityksen sisällön ja todellisuuden keskinäistä vastaavuutta, joka on kuitenkin Westerståhlin mukaan jotain muuta, kuin tiedonvälityksen antama *vaikutelma* todellisuudesta. Olennaisuus taas merkitsee keskittymistä todellisuuden tärkeimpien so. olennaisimpien puolien kertomiseen ilman mielivaltaisesti tai tarkoitushakuisesti valittuja sinänsä todenmukaisia yksityiskohtia. (Ibid.)

Toisaalta on myös olemassa koulukunta (American Press Institute 2016, viitattu 22.2.2016, ks. myös Kuutti 2015, 21), jonka mukaan journalisti ei voi olla objektiivinen, koska tämän on tehtävä työssään tiettyjä valintoja. Sen sijaan journalistiset menetöt, joilla näitä valintoja tehdään, voivat olla objektiivisia. Tässä ajattelussa puolueettomuus ei ole yksi journalismin peruskivistä. Objektiivisuuden ihanne ei tämän periaatteen mukaan alun perinkään tarkoittanut, että journalistit olisivat vapaita omista ennakoasenteistaan. Sen sijaan sillä tarkoitettiin johdonmukaista tietojen testaamista siten, että henkilökohtaiset ja kulttuurisidonnaiset asenteet eivät vähentäisi työn tarkkuutta. (Ibid.)

Yleensä toimittajat ovat ratkaisseet työhönsä liittyvän totuusongelman sisällyttämällä journalistisiin tuotoksiinsa kaikki mahdolliset totuutta koskevat väitteet. Kun journalisti ei pysty

kertomaan, *mikä on totta*, hän voi vain taltioida totuutta koskevia väittämiä mahdollisimman tarkasti. Tällöin journalismin totuusvaatimus siirtyy väitteen ja todellisuuden välisestä suhteesta väitteen ja sen *kuvauksen* väliseen suhteeseen. Journalisti ei siis voi taata, että esitetty *väite* on totta, vaan ainoastaan sen, että tällainen väite on *sellaiseksi sanottu*. Tällainen toimintamalli perustuu tarkkuutta koskeviin standardeihin ja vapauttaa toimittajan selvittämästä totuutta. (Dunwoody 2005; Partanen 2007, Kuutin 2015, 12–13, mukaan.)

Näinkin menetellessään journalistien olisi kuitenkin pyrittävä arvioimaan, kuka puhuu totta ja kuka ei. Jos ja kun journalismi on valikoima erilaisia väittämiä ja näkökantoja, julistaa se lopulta vain totuuden olevan 'tuolla jossain'. Kahden täysin vastakkaisen näkökannan sisällyttäminen journalistiseen juttuun voi antaa käsiteltävästä asiasta liian mustavalkoisen kuvan ja johtaa yleisöä petollisen yksinkertaiseen tulkintaan. Yleisö ei välttämättä edes ymmärrä tämän tapaista näkökulmien painotusta ja tulkitsee jutun siten, että kukaan ei tiedä totuutta asiasta. (Dunwoody 2005, Kuutin 2015, 13–14 mukaan.)

Oman höysteensä journalismissa esitettyjen tietojen paikkansapitävyyteen on tuonut sosiaalinen media. Koskaan aikaisemmin eivät yhtä moni ihminen ja organisaatio ole osallistuneet tietojen tarkistamiseen ja varmistamiseen. Myöskään virheen paljastuminen ei ole koskaan ollut niin herkässä kuin nykyään. Journalisteilta vaaditaankin voimakasta itseuria ja julkaistavien tietojen varmistamisen nostamista yhdeksi tärkeimmistä journalismin arvoista ja käytännöistä. (Silverman 2012, Kuutin 2015, 14, mukaan.)

2.2.3 Itsenäisyys

Niin kuin edellä mainittiin, yksi journalismille annetuista määritelmistä on sen *itsenäisyys*. Tällä viitataan siihen, että journalistien ammattikunta pyrkii suojelemaan työtään manipulointi- ja vaikutusyrityksiltä; toimituksellisen työn ja harkinnan katsotaan olevan tae julkaistujen tietojen luotettavuudesta. (Kunelius 2009, 23.) Näin ollen itsenäisyyden voidaan katsoa olevan jonkinlaisessa suhteessa journalismin pyrkimyksiin kohti todenmukaista tiedonvälitystä ja objektiivisuutta. Itsenäisyyden vaaliminen liittyy Kuneliuksen (ibid.) mukaan myös *edustavuuteen*, joka on niin ikään yksi journalismin perusaineista: vain itsenäinen maine voi taata uskottavuuden, kun tarkoitus on edustaa suuren yleisön ääntä ja näkökulmaa julkisuudessa.

Syvemmillä tasolla itsenäisyydellä voidaan viitata journalistin henkilökohtaiseen itsenäisyyteen, avarakatseisuuteen ja uteliaisuuteen, jotka kaikki auttavat tarkastelemaan maailmaa omasta yhteiskunnallisesta asemasta, rodusta, etnisestä taustasta, uskonnosta, sukupuolesta tai egosta riippumatta (American Press Institute 2016, viitattu 22.2.2016). Journalistinen itsenäisyys ei kuitenkaan ole sama asia kuin puolueettomuus. Vaikka esimerkiksi lehtien pääkirjoitukset tai kolumnit eivät ole puolueettomia, tulisi niiden uskottavuuden perustua tarkkuuteen, älylliseen rehellisyyteen ja pyrkimykseen evästää lukijaa sen sijaan, että näissä kantaaottavissa teksteissä kumarrettaisiin tiettyihin eturyhmiin päin tai pyrittäisiin edistämään jotakin määrättyä tavoitetta ja lopputulemaa. (Ibid.)

Suomessa Journalistin ohjeiden kohdat 2 ja 3 liittyvät erityisesti journalismin itsenäisyyteen: "Tiedonvälityksen sisältöä koskevat ratkaisut on tehtävä journalistisin perustein. Tätä päätösvaltaa ei saa missään oloissa luovuttaa toimituksen ulkopuolisille.", sekä: "Journalistilla on oikeus ja velvollisuus torjua painostus tai houkuttelu, jolla yritetään ohjata, estää tai rajoittaa tiedonvälitystä." (Julkisen sanan neuvosto 2016, viitattu 18.2.2016.)

Yhteiskunnassa vaikuttavia journalismin ulkopuolisia joukkoviestinnän muotoja ovat muun muassa propaganda, disinformaatio ja organisaatioiden pr-toiminta (Kuutti 2015, 15–16). Nämä kaikki asettavat haasteita journalismin itsenäisyydelle ja sitä kautta journalistiselle totuudelle.

Propaganda on yleisen mielipiteen muokkaamista propagandan harjoittajan tavoitteiden mukaiseksi. Keinovalikoimaan kuuluvat mustamaalaus, yleistyksiset, perustelemattomat väitteet, vastakkainasettelut, vastaanottajan imartelu, oman maineen korostaminen, vastaanottajaan samaistuminen ja vastustajien leimaaminen. Arvosidonnainen ajattelutapa, vastakkaisten näkemysten ja tulosten salaaminen sekä pyrkimys yhdenmukaistaa vastaanottajan uskomuksia ja näkemyksiä ovat propagandan tyypillisiä piirteitä. (Ibid.) Disinformaatio puolestaan muistuttaa propagandaa, mutta on Kuutin (ibid.) mukaan sitä vaarallisempaa. Siinä synnytetään ja levitetään julkisuuteen tarkoituksellisesti harhaanjohtavaa tietoa. Disinformaatio naamioidaan usein luotettavan vaikutelman antavaksi uutiseksi, joka ujutetaan huomaamattomasti tiedotusvälineisiin. Pr-toiminnassa (engl. public relations, suom. tiedotus- ja suhdetoiminta) journalistiseen muotoon verhoilluilla viesteillä tuetaan perusteettomasti jutun käsittelemää tahoa. Taustalla voi olla mainoskytkentä tai hyvien suhteiden ylläpitäminen. Viestien uskottavuus kasvaa nimenomaan sillä, että ne saadaan osaksi journalistisia sisältöjä. (Ibid.)

Kaupallisen journalismin juttutuotannon on oltava paitsi ennakoitavaa ja turvattua myös taloudellisesti tehokasta. Tästä syystä monet tiedotusvälineet suosivat lähteitä, joilla on tietoja mahdollisimman valmiissa muodossa. (Kunelius 2009, 92) Esimerkiksi viranomaistiedotteet ovat suosittuja uutislähteitä, koska ne tarjoavat helposti uutisiksi taipuvaa materiaalia. Jotkut ovatkin kärjistäneet, että mediat saavat näin valtaapitäviltä yhteiskunnallisilta instituutioilta epäsuoraa taloudellista tukea. (Ibid.) Kuutti (2015, 17) sivuaa samaa aihetta ja vie vertauksen pelkästä taloudellisesta tuesta pidemmälle. Esimerkiksi erilaisten tiedotusaineistojen journalistisessa käsittelyssä ei välttämättä oteta huomioon, että näitä aineistoja tuottamalla viranomaiset ja muut organisaatiot pystyvät käytännössä itse laatimaan itseään koskevia uutisia. Yleisön käsitys tietojen alkuperästä hämärtyy, kun tiedot julkaistaan toimituksen omana tuotantona ja jopa toimittajan omista nimistä. Yleisö kuitenkin olettaa, että journalismin välittämät tiedot olisivat käyneet läpi toimituksellisen arvioinnin ja että niiden paikkansapitävyys olisi tarkistettu. Taloudelliset paineet vaikuttavat myös tässä: kun toimittajille ei anneta aikaa tarkistaa faktoja tai etsiä omia juttuaiheita, he ovat entistä alttiimpia pr- ja muulle tiedotustoiminnalle (Aro 2011, Kuutti 2015, 17, mukaan).

Kuneliuksen (2009, 92) mukaan journalismin luonne yksityisenä liiketoimintana vapaaseen kilpailuun perustuvassa markkinataloudessa herättää kysymyksiä tällaisen talousjärjestelmän seurauksista journalismin sisällöille. Voidaan esimerkiksi kysyä, tuottaako tiukka kilpailu monipuolista sisältöä. Kun eri tiedotusvälineet kilpailevat samasta yleisöstä, alkavat niiden tuotteet enemmän muistuttaa toisiaan kuin eriytyä. Tiukassa kilpailutilanteessa eri mediat seuraavat tarkasti toistensa tekemisiä, välittävät ainakin samat uutiset kuin kilpailijansa ja kopioivat menestyneitä toimintamalleja. Toisaalta Kunelius (ibid.) myös huomauttaa, että tämä ei automaattisesti tarkoita, että vähempi kilpailu merkitsisi moniarvoisempaa journalismia.

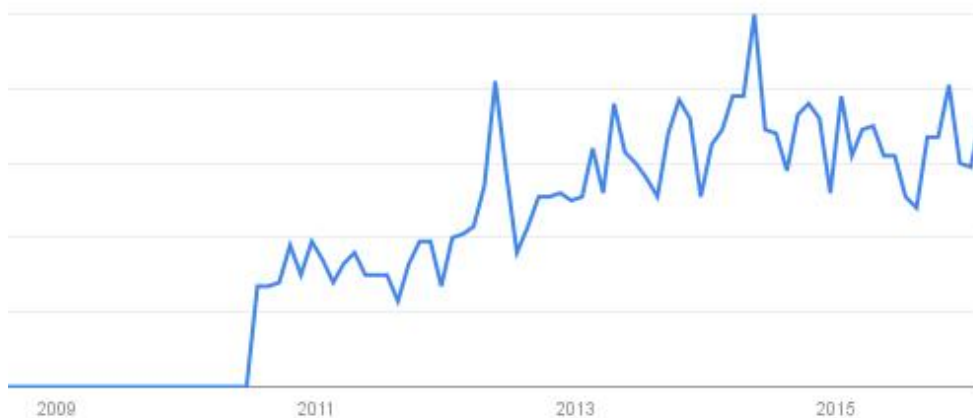
Joukkoviestinnän kaupallistuminen on johtanut siihen, että journalismia koskevia ratkaisuja tehdään yhä enemmän liiketaloudellisin perustein. Toimiessaan osana pörssiyhtiötä perinteinen maakuntalehtikin asettuu samalla sijoittajien ja eri analyytikkojen kriittisen silmän arvioitavaksi. Jos journalismin tekeminen ja sen laadun kehittäminen katsotaan yhtiössä menoeräksi, edellyttää menojen karsiminen vaikkapa yhtiön tuloksen parantamiseksi usein myös journalismista karsimista. Kun liiketaloudellinen tehokkuus menee kaiken edelle, jäävät journalismin tavoitteet helposti sen jalkoihin. Toimitusten kokoa pienennetään, työtahti kiihtyy, virheet lisääntyvät ja esimerkiksi tutkivaan journalismiin ei panosteta. (Ibid.)

Vehkoo (2011, 31–32) on samoilla linjoilla ja toteaa, että suomalaiset mediatilat ovat taloudellisesta taantumasta huolimatta jakaneet korkeita osinkoja osakkeenomistajilleen. Voitto-odotusten laskeminen voisi kuitenkin olla järkevää, sillä näin vapautuisi rahaa journalismin ja uusien digitaalisten palvelujen kehittämiseen, mikä olisi Vehkoon mukaan sijoitus tulevaisuuteen. (Ibid.) Pörssikapitalismin toimintalogiikka ei kuitenkaan usein toimi niin, että mediayhtiö ilmoittaisi jakavansa vähemmän osinkoja ja sijoittavansa journalismiin, elleivät osakkeenomistajat sitten sattuisi olemaan journalismin ystäviä ja demokratian puolestapuhujia. Todennäköisemmin he kuitenkin myisivät osakkeensa pois ja yhtiön osakekurssi laskisi. (Ibid.)

Kysymys journalismin itsenäisyydestä on yksi viestinnän tutkimuksen ikuisista kiistakapuloista (Kunelius 2009, 24). Yhtäältä koetaan, että journalismilla on median täyttämässä yhteiskunnassa paljon valtaa, mutta vähän vastuuta. Toisaalta näyttäisi siltä, että journalismin kyky ajatella irrallaan omasta ajastaan ja sen taloudellisista sekä poliittisista kytköksistä olisi rajallinen. Journalismi on esimerkiksi enemmän tai vähemmän kiinni siitä tiedosta, jota yhteiskunnan muut toimijat tuottavat tai ylipäättään hyväksyvät tiedoksi. Lisäksi journalismin itsenäistä ja niin sanoen vapaata ajattelua rajoittaa sen luonne yhä useammin markkinataloudessa toimivana yksityisenä liiketoimintana. (Ibid.)

2.3 Datajournalismi

Kaikki leipänsä (data-)journalismin parista hankkivat tai hankkineet ammattilaiset eivät pidä koko käsitteen määrittelyä järin tarpeellisena (ks. mm. Holovaty 2009, viitattu 24.2.2016; Liski, haastattelu 17.2.2015; Mäkinen, Niemisen 2015, mukaan). Datajournalismi on joka tapauksessa alkanut 2010-luvulla kiinnostaa ihmisiä. Tämän tutkielman puitteissa on olennaista selvittää, mistä puhutaan (ja tehdään Google-hakuja), kun puhutaan datajournalismista.



KUVA 2. Hakusanalla 'data journalism' tehdyt haut alkoivat yleistyä Googlen hakukoneessa vasta vuoden 2010 kesäkuun jälkeen (Google Trends 2016, viitattu 25.2.2016)

Datajournalismin määritelmä riippuu pitkälti siitä, mikä on datan määritelmä. Jos 'data' ymmärretään – kuten se puhekielessä usein ymmärretään (ks. Poikola ym. 2010, 13) – synonyymiksi tiedon, informaation tai faktojen kanssa, on se sanan 'journalismi' etuliitteenä ehkä hieman turhahko tai vähintään kyseenalainen. Sillä jos journalismi on pohjimmiltaan nimenomaan faktapohjaista joukkoviestintää, eikö se silloin terminä riitä kattamaan kaikenlaisen faktoihin, informaatioon, tietoon tai dataan perustuvan joukkoviestinnän ilman, että sitä pitäisi erikseen mainita? Miksi siis puhutaan datajournalismista? Onko datajournalismin ja 'pelkän' journalismin välillä jokin perustavaa laatua oleva ero?

Eräs määritelmä, jossa datajournalismin käsite on yritetty tiivistää yhteen virkkeeseen, kuuluu näin: ”Datajournalismi on työprosessi, jonka aikana hankitaan, muokataan ja analysoidaan dataa ja jonka lopputuloksena julkaistaan 'datajournalistinen juttu' ja mieluiten myös jutussa käytetty data” (Harju & Sirkkunen 2015, viitattu 23.2.2016). Määrittely on kuitenkin melko suurpiirteinen, eikä se tee selvää eroa datajournalismin ja journalismin välille. Jos sanan 'data' korvaisi tässä määrittelyssä esimerkiksi sanalla 'tieto', oltaisiin pitkälti pysytty edelleen pelkän journalismin määritelmässä. (ks. Liski, haastattelu 17.2.2015.)

Honkonen & Lankinen (2012, 143) ovat määritelleet datajournalismin “[prosessiksi], jossa laajoja tietoaaineistoja yhdistelemällä, suodattamalla, analysoimalla ja visualisoimalla tuotetaan uutisia.” Tässä määrittelyssä painoarvo on 'laajoilla tietoaaineistoilla', jotka viittaavat johonkin muuhun kuin esimerkiksi pelkkään haastattelulausuntojen pohjalta tehtyyn journalismiin. Mutta edelleen: mistä

kumpuaa tarve erottaa 'laajoihin tietoaineistoihin' pohjautuva journalismi 'pelkästä' journalismista kutsumalla sitä datajournalismiksi?

Anna-Sofia Nieminen puolestaan on ehdottanut pro gradu -tutkielmassaan 'Uuden ajan journalismia' datajournalismille määritelmää "- - journalistinen prosessi, jossa tuotetaan datan avulla uutta tietoa verkkoon tai verkko edellä ja johon liittyy usein visuaalisia ja interaktiivisia elementtejä" (Nieminen 2015, 71, viitattu 23.2.2016). Tämäkään määritelmä ei riitä juurikaan kuvaamaan, kuinka esitetty datajournalismi eroaa journalismista. Niemisen määritelmä sivuaa lähinnä *verkkojournalismia* (ks. esim. Kuutti 2012, 217), mikäli journalismin julkaisemista internetissä on enää vuonna 2010-luvulla edes mielekästä eritellä omaksi lajityypikseen.

Ehkäpä ainakin edellä mainittuja ehdotuksia tarkempi määritelmä datajournalismille olisi kutsua sitä "[journalismin] suuntaukseksi, jossa yritetään löytää mistä tahansa [journalistisesta] jutusta ne asiat, jotka ovat luokiteltavissa, laskettavissa ja verrattavissa, ja joista tietotekniikan avulla saadaan irti jotakin kertomisen arvoista sekä sellaista, jota ei muuten saataisi mitenkään tietää" (Thibodeaux 2011, viitattu 24.2.2016). Vaikka tässäkin määrittelyssä ei tuoda selvästi esille, kuinka datajournalismi eroaa journalismista, painotetaan siinä vahvasti tietoteknisiä työmenetelmiä sellaisten tietojen kaivamisessa esiin, joita olisi ilman tietotekniikkaa mahdotonta löytää.

Brittiläinen toimittaja Paul Bradshaw määrittelee (2012, 2) datajournalismin nimenomaan datan ja sen kasvaneen määrän sekä saatavuuden kautta. Datajournalismi on hänen mukaansa perinteisen journalistisen 'uutisvainun' yhdistämistä valtavaan määrään digitaalista informaatiota, joka on nykyään saatavilla. Lisäksi Bradshaw (2015, viitattu 25.2.2016) on sanonut, että datajournalismi on mitä tahansa journalismia, joka sisältää jollakin tavalla jäsenettyä tai strukturoitua dataa.

Kuutin (2012, 26, 198) mukaan datajournalismi taas tarkoittaa samaa asiaa kuin *tietokantajournalismi*:

Tietokantajournalismi [on] journalistinen suuntaus, jossa (sähköisen) tiedonhankinnan kohteena ovat erityisesti viranomaisten hallinnoimat tietokannat ja rekisterit kokonaisuudessaan tai niiden valikoidut yksittäiset sisällöt. Toimintamallin ideana on, että toimitustyössä tarkasteltavia asioita tai tapahtumia voidaan määrittää ja nimetä tietoteknisesti luokiteltaviksi ja mitattaviksi suureiksi. Tietokantojen ja rekisterien sisältämää

dataa voidaan kerätä, tallentaa, paloitella, analysoida ja hyödyntää useita kertoja. (Kuutti 2012, 26, 198)

Määritelmä käy järkeen nimenomaan silloin, kun on kyse tietokannoista ja niihin liittyvästä journalistisesta tiedonhankinnasta. Dataa on kuitenkin olemassa myös tietokantojen ulkopuolella (ks. mm. Lindenbergh 2012, 124–125), minkä johdosta tätä Kuutin määritelmää tietokantajournalismista ei voida pitää kaiken kattavana määritelmänä datajournalismista.

Kuutti (2012, 198) myös toteaa, että tietokantajournalismi on osa laajempaa käsitettä *tietokoneavusteinen journalismi* (engl. *computer-assisted reporting; CAR*), jolla tarkoitetaan journalistista suuntausta, jossa hyödynnetään tietotekniikkaa ja tietoverkkoja. Tietokoneavusteinen journalismi voidaan edelleen katsoa osaksi *tallennepohjaista journalismia*, jonka toinen osa on muuta kuin digitaalisessa muodossa olevaa aineistoa hyödyntävä *dokumenttijournalismi*. (Ibid.) Bounegrün (2012, 18) mukaan tietokoneavusteinen journalismi oli ensimmäinen järjestelmällinen, systemaattinen tapa käyttää tietokoneita datan keräämiseen ja analysoimiseen journalististen tuotosten parantamiseksi. CARista kerrotaan tarkemmin luvussa 3.

Luennoitsija ja bloggaaja Tony Hirst on erotellut kolme mahdollista näkökulmaa datajournalismin määrittelyyn, jotka ovat vapaasti suomennettuina *lopputuote*, *ammattitaito* ja *prosessi* (Hirst 2013, viitattu 25.2.2016).

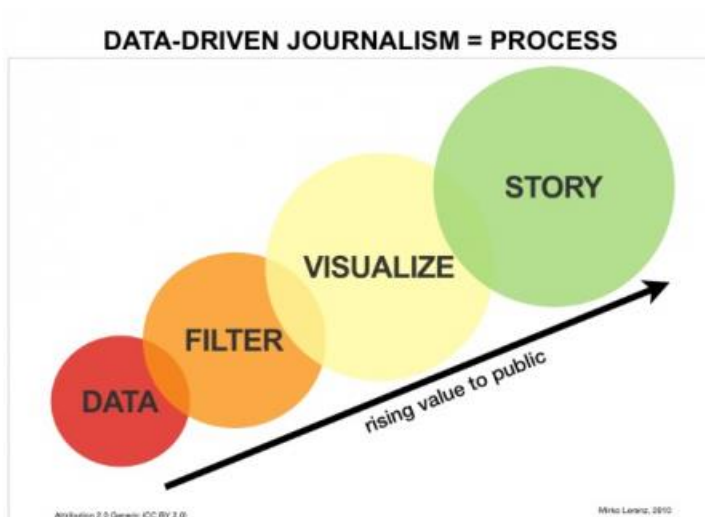
Lopputuotenäkökulmassa datajournalismi määrittyy sen lopullisen esitystavan mukaan. Lopputuote on visuaalinen esitys journalistisessa jutussa käytetystä datasta; se havainnollistaa jollakin tavalla niitä tietoja, joita toimittaja on jutussaan käsitellyt ja joiden varaan juttu perustuu. Hirst kuitenkin huomauttaa, että lopputuotteen tekeminen on usein jutunteon helpoin vaihe: suurin osa ajasta käytetään datan etsimiseen, datan saamiseen sellaiseen muotoon että sillä voisi ylipäätään tehdä jotain sekä sen esitystavan miettimiseen ja löytämiseen, jolla data mahdollisesti halutaan jutussa havainnollistaa.

Tiettyjen taitojen näkökulmassa taas datajournalismi ei ole niinkään sitä, mitä toimittajat tuottavat ja julkaisevat. Sen sijaan datajournalismia on se *ammattitaito*, joka journalisteilla tulee olla, jotta he voivat tuottaa ja julkaista näitä dataa havainnollistavia esityksiään eli datajournalistisia juttuja (lopputuotteita). Määrittely suorastaan huutaa tulkintaa. Ovatko samalla logiikalla ajatellen sanomalehden toimittajan kirjoitustaito ja kuullun ymmärtäminen myös journalismia? Onko pelkkä

kysymysten esittäminen haastateltavalle journalismia? Nämä kaikki – ja myös paljon muuta, esimerkiksi journalistin työhön liittyvien eettisten ohjeiden tunteminen – ovat journalistin ammattitaitoa, jota ilman journalistisia lopputuotteita ei syntyisi. Vaikka toimittaja kirjoittaisi ammattitaitonsa pohjalta juttunsa puhtaaksi, mutta sitä ei koskaan julkaistaisi, ei journalismia olisi journalismin perinteisen määritelmän mukaan syntynyt, koska sanomaa ei olisi lähetetty kenellekään. (ks. Kunelius 2009, 17; Hemánus 1995, 27–28.) Toisaalta Hirstin (2013, viitattu 25.2.2016) luonnehdinta, jonka mukaan datajournalismi on datalle sitä, mitä radiojournalismi on äänelle ja videojournalismi liikkuvalla kuvalla, ohjaa ajattelemaan data- ja mitä tahansa muuta journalismia eräänlaisena suodattimena, jossa journalisti ammattitaitoineen on osa laajempaa prosessia.

Hirstin prosessinäkökulmassa datajournalismi ymmärretään prosessina, johon liittyy tai joka sisältää jollakin tavalla dataa, mutta ei välttämättä rakennu pelkän dataa havainnollistavan lopputuotteen ympärille. Datajournalistinen prosessi saattaa esimerkiksi paljastaa jonkin ”johtolangan”, jonka pohjalta journalistinen juttu kehittyy. Näin ollen data on yksinkertaisesti tietolähde muiden joukossa. Datan pohjalta syntynyt juttu voi yhtä hyvin sisältää datajournalistisen lopputuotteen, kuten dataa havainnollistavan visuaalisen esityksen, tai ihan vain pelkkää tekstiä.

Useat datajournalismin määritelmät korostavat juuri sen prosessinomaista luonnetta. Esimerkiksi saksalainen toimittaja Mirko Lorenz on sanonut, että datajournalismi voidaan mieltää prosessiksi, jossa raakadatasta jalostetaan merkityksellistä tietoa. Hänen mukaansa menetelmä kasvattaa journalismin arvoa yleisön näkökulmasta, kun monimutkaiset asiat ja faktat saadaan tiivistettyä selkeään muotoon, joka on helppo ymmärtää ja muistaa. (Lorenz 2010, viitattu 29.2.2016).



KUVA 3. Mirko Lorenzin useasti Suomessakin siteerattu (mm. Harju & Sirkkunen 2015; Toikkanen 2014) datajournalismin prosessia havainnollistava malli (Lorenz 2010, viitattu 29.2.2016)

On huomionarvoista, että Lorenzin mallissa datan visualisoinnin rooli on aivan keskeinen: datajournalistista juttua ei synny ilman dataa havainnollistavaa esitystä. Yleisön näkökulmasta datan visualisointi myös kasvattaa koko jutun journalistista arvoa.

Journalismin professori emeritus Philip Meyer Pohjois-Carolinan Chapel Hillin yliopistosta korostaa hänkin tiedon prosessointia määritellessään datajournalismia. Meyerin mukaan toimittajat olivat aikaisemmin niukan informaation aikakaudella kuin "metsästäjä-keräilijöitä". Nykyään tietoa on yltäkyläisesti, joten sen prosessointi ja käsittely on tärkeämpää. (Meyer 2012, 6) Meyerin mukaan datajournalismissa prosessointia tapahtuu sekä *analyysin* että *esitystavan* tasolla. Ensinnä journalistien tulee analyysin keinoin saada nykyajan loputtomasta datavirrasta jotain mielekästä tietoa irti. Kaikkein tärkeimmät ja olennaisimmat tiedot on sitten vielä esitettävä yleisölle jollakin järkevällä tavalla. Lisäksi tähän prosessiin kuuluu käytettyjen työmenetelmien avaaminen ja jutun julkaiseminen siten, että kuka tahansa voi arvioida siinä esitettyjen tietojen todenmukaisuutta toistamalla ko. prosessin. Näin datajournalismi on Meyerin mukaan verrattavissa tieteelliseen metodiin. (Ibid.)

Yleensä datajournalististen juttujen yhteydessä korostetaan data-analyysin läpinäkyvyyttä. 'Datajutun' yhteydessä voidaan esimerkiksi julkaista tietolaatikko, jossa selvitetään analysoinnin metodologiaa, työvaiheita ja mahdollisia rajoitteita. Usein data-analyysin tulokset ovat perusteltuja

juuri siksi, että tietoteknisiä toimintoja voidaan tieteellisen tutkimuksen tavoin toistaa ja tarkistaa. (Kuutti 2015, 125–126.)

Edellä mainittuun luonnehditaan liittyen huomautettakoon, että journalismi on perinteisesti ymmärretty nimenomaan työprosessina (ks. Hemánus 1995, 28; American Press Institute 2016, viitattu 18.2.2016). Siispä datajournalismin määrittely erityisenä prosessina, jossa joitain tietoja käsitellään jollakin tavalla, ei vielä sellaisenaan eroa olennaisesti journalismin määritelmästä.

Näyttää olevan hankala löytää ytimekästä määritelmää, joka tiivistäen kertoisi, mitä datajournalismi on suhteessa journalismiin. Yhteistä lähes kaikille tässäkin tutkielmassa pyöritellyille datajournalismin määritelmille tuntuisi olevan, että niissä datalla tarkoitetaan ehdottomasti raakaa kontekstoimatonta tietoa sähköisessä, koneluettavassa muodossa.

On myös merkille pantavaa, että terminä datajournalismi on alkanut yleistyä vasta 2010-luvulle tultaessa. Voisi siis olettaa, että datajournalismista on alettu yleisemmin puhua yhtä aikaa internetin, sosiaalisen median, (avoimen) datan saatavuuden ja erilaisten datan käsittelyä helpottavien työkalujen yleistymisen sekä tietokoneiden kehittymisen myötä (ks. Lorenz 2010, viitattu 1.3.2016; Bradshaw 2012, 2; Tebest 2012, viitattu 25.11.2014; Radcliffe 2016, viitattu 1.3.2016).

Mutta onko tietotekniikan ja datan käyttäminen – siis jotenkin muuten kuin nykyaikaisena kirjoituskoneena ja pintapuolisena tietolähteenä – osana journalistisia työmenetelmiä edes uusi ilmiö? Esimerkiksi data-aineistoja käytettiin jo ensimmäisissä kausittaisissa julkaisuissa 1600-luvulla ja sanomalehdissä 1800-luvulla. Tiedon visualisoinnilla on satoja ellei tuhansia vuosia pitkä historia aina antiikin ensimmäisistä kartoista lähtien. (Bounegru 2012, 19; Klein 2015, viitattu 1.3.2016; Radcliffe 2016, viitattu 1.3.2016; Friendly 2006, 3, viitattu 1.3.2016). Tietokonetta puolestaan käytettiin tiettävästi ensimmäistä kertaa journalistisen data-analyysin tekemiseen vuonna 1952 Yhdysvalloissa (Henn 2012, viitattu 1.3.2016).

3 DATA, JOURNALISMI JA TIETOKONEET – HISTORIALLINEN NÄKÖKULMA

Datajournalismin teoreettinen määrittely suhteessa journalismiin itsessään osoittautui haastavaksi. Varsinaista eroa näiden kahden termin välillä ei välttämättä edes ole – kuten tässä luvussa tuodaan esille, dataa on käytetty journalismissa tavalla tai toisella jo kauan. Tietotekniikan ottamat edistysaskeleet sen sijaan ovat kiistaton tosiasia. Ehkäpä datajournalismin käsite avautuu parhaiten juuri kehittyneiden työvälineiden, menetelmien ja kasvaneiden tietoresurssien myötä.

Jotta datajournalismille voisi antaa mitään mielekästä merkitystä, on sitä lähestyttävä muiden, huomattavasti tätä muotisanaa vanhempien käsitteiden ja journalististen työtapojen kautta. Pieni aikamatka lienee siis paikallaan.

Katsaus aloitetaan lähes journalismin alkuhämärästä, josta siirrytään kohti tietokoneiden esiinmarssia yhteiskuntaan ja sitä myöten toimituksiin ja osaksi journalistisia käytäntöjä. Tietokoneavusteisen ja tutkivan journalismin perinteiden kautta saavutaan 2010-luvulle ja datajournalismiin. Mikä on muuttunut, vai onko lopulta mikään?

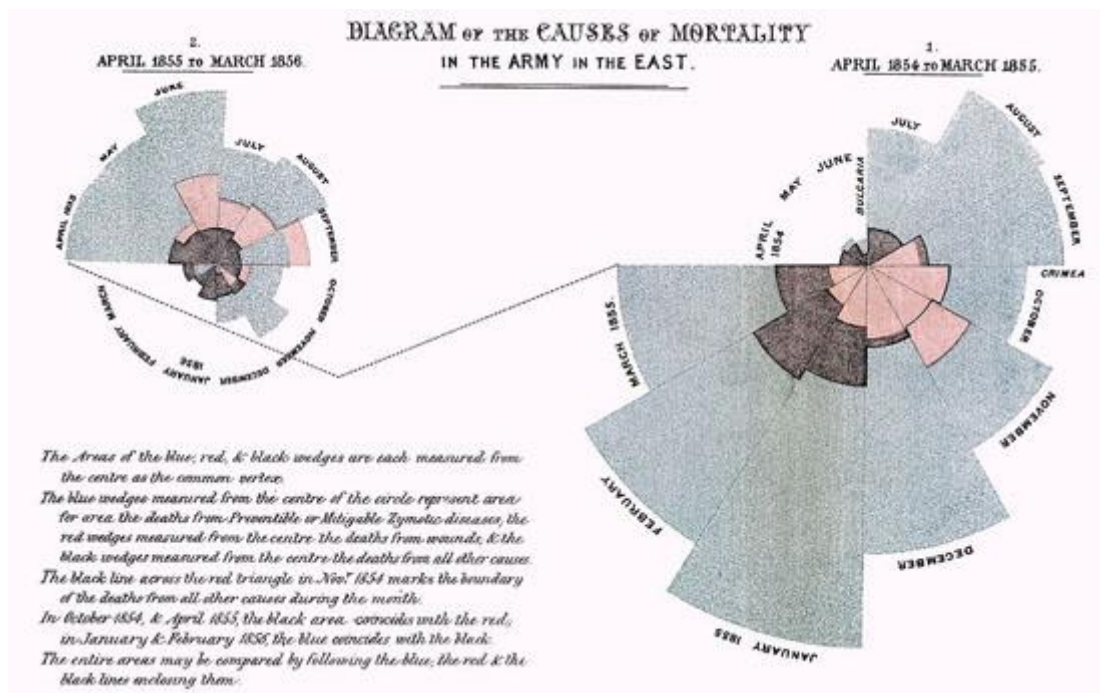
3.1 Ei mitään uutta auringon alla?

Journalismin alkuräjähdyksenä pidetään myöhäiskeskiajalla ja uuden ajan alussa kehittyneiden kauppayhteyksien synnyttämää tiedon tarvetta. Eurooppalaiset kauppiat kirjoittivat tuolloin käsin kuulumisia toisilleen, ja näissä kirjeissä oli tapana kertoa myös yleisemmin tuoreimmista tapahtumista. Tällaisten uutiskirjeiden historiallisina esi-isinä voidaan pitää antiikin Rooman virallista 'päivälehteä' Acta Diurnaa ja muinaisen Kiinan Han-dynastian 'virallisten residenssien raportteja'. Uutiskirjeiden pohjalta on kuitenkin kehittynyt se idea uutisesta, jonka varaan ensimmäiset sanomalehdet perustettiin. (Howard 2014, viitattu 1.3.2016; Kunelius 2009, 58.)

1500-luvulla uutisia alettiin kerätä järjestelmällisemmin yhteen. Esimerkiksi Venetsiassa myytiin tuolloin käsin kopioituja, lyhyitä uutisia sisältäviä lehtisiä, jotka kulkeutuivat aina Lontooseen asti. (Kunelius 2009, 58) Nämä uutiskirjeet ja lehtiset, joita kirjoittivat usein kauppiaiden palkkaamat

kirjeenvaihtajat, sisälsivät paljon muun muassa liiketoimintaan liittyvää 'dataa'. (Klein 2015, viitattu 1.3.2016.)

Varhaisissa sanomalehdissä julkaistiin myös runsaasti dataan pohjautuvia artikkeleita. Tavallisia data-aineistoihin perustuvia juttuaiheita olivat kiinteistöverojen panttioikeudet, kuolleisuus- ja kuolinsyytilastot, kulutushyödykkeiden hinnat sekä satamissa vierailevien laivojen lastit. (Ibid.)



KUVA 4. Florence Nightingalen visualisointi brittien armeijan sotilaiden kuolinsyistä Krimin sodan aikana julkaistiin vuonna 1858. Punainen osa edustaa haavoittuneina kuolleita, musta muihin syihin kuolleita ja sininen tauteihin kuolleita (Bounegru 2012, 19–20)

Yksi esimerkki varhaisesta datan visualisoinnista on brittiläisen sairaanhoitajan Florence Nightingalen infografiikka 1800-luvun puolivälistä. Hänen Ison-Britannian parlamentille antamassaan raportissa eriteltiin tietoja brittien armeijan sotilaiden kuolinsyistä Krimin sodan aikana. Visualisoinnin keinoin havainnollistetusta datasta selvisi, että valtaosa sodassa taistelevista brittiläisistä kuoli tauteihin eikä taisteluissa venäläisten kanssa. (Radcliffe 2016, viitattu 1.3.2016; Bounegru 2012, 19.)

Toinen hyvä esimerkki varhaisesta datajournalismista samalta aikakaudelta on The New York Tribune -lehden tutkimus Yhdysvaltain senaattorien ja kansanedustajien matkakulukorvauksista.

Vuonna 1848 julkaistussa jutussa kyseenalaistettiin se käytäntö, jonka perusteella senaattoreille ja kansanedustajille maksettiin kulukorvauksia näiden tekemistä matkoista pääkaupungin ja kotipaikkakuntansa välillä. Lain mukaan korvauksia maksettiin tuohon aikaan edustajan “tavallisimman matkustusreitillä perusteella”. Tribune-lehden päätoimittaja Horace Greeley sai käsiinsä tiedot senaattoreille ja kansanedustajille maksetuista matkakorvauksista. Hän äimistyi korvausten suuresta määrästä, ja alkoi tutkia asiaa tarkemmin. (Radcliffe 2016, viitattu 1.3.2016; Klein 2015, viitattu 1.3.2016.)

Greeley päätti hankkia postilaitokselta tiedot postin käyttämistä jakelureiteistä, ja laski niiden perusteella lyhimmän mahdollisen reitin kunkin edustajan kotipaikkakunnalta pääkaupunkiin. Sitten hän laski käytössä olleella laskutavalla edustajille uudet kilometrikorvaukset lyhimpien mahdollisten reittien mukaan, ja vertasi näitä summia edustajille todellisuudessa maksettuihin matkakorvauksiin. Tutkimus paljasti, että useimmille edustajille maksettiin liikaa matkakorvauksia. He olivat ilmoittaneet käyttävänsä pääkaupunkiin kulkemiseen huomattavasti pidempiä reittejä, vaikka tarjolla olisi ollut ainakin teoriassa lyhyempiä vaihtoehtoja. (Ibid.)

Greeleyn tähtäimessä eivät olleet niinkään 'rahanahneet poliitikot', vaan se käytäntö, jonka perusteella matkakuluja korvattiin kyseenalaistamatta niiden paikkansapitävyyttä. Hän ehdotti, olisiko matkat joka tapauksessa järkevämpää korvata lyhimmän mahdollisimman reitin mukaan, käyttipä edustaja sitä tai ei. Osittain Greeleyn jutun ja Tribunen muiden aihetta käsitelleiden kirjoitusten ansiosta edustajien matkakuluista säättävää lakia lopulta muutettiin. Sivujuonteena kerrottakoon, että ylimääräisiä kilometrikorvauksia saaneiden edustajien joukossa oli eräs Abraham Lincoln (677 dollaria liikaa korvauksia – nykyrahassa yli 18 700 dollaria). (Ibid.)

Edellä mainittuja tapauksia vielä parikymmentä vuotta vanhempi esimerkki journalistisesta datan käyttämisestä löytyy brittiläisen The Guardian -lehden – silloiselta nimeltään Manchester Guardian – ihka ensimmäisestä numerosta vuodelta 1821. Juttu käsitteli Manchesterin koulujen oppilasmääriä ja koulujen kustannuksia, ja siinä käytettiin lähteenä hallinnosta vuodettua data-aineistoa, joka myös julkaistiin lehdessä. Luvut paljastivat, että oppilaiden todellinen määrä oli paljon suurempi, mitä koulutus uudistusta ajaneet tahot olivat väittäneet. (Bounegru 2012, 19; Klein 2015, viitattu 2.3.2016.)

Mainitsemisen arvoinen on myös skotlantilainen insinööri William Playfair, jonka jo vuonna 1786 julkaistu teos Atlas sisälsi 44 dataa havainnollistavaa visualisointia. Playfair joko jalosti tai keksi valtaosan yhä tänä päivänä käytössä olevista tilastollisista infograafeista. (Rogers 2014, 6.)

Näiden esimerkkien valossa on perusteltua väittää, että dataa on käytetty osana journalistisia juttuja jo kauan ennen ensimmäisten tietokoneiden keksimistä saati niiden yleistymistä yhteiskunnassa ja osaksi journalistista työprosessia. Kuten Radcliffe (2016, viitattu 2.3.2016) huomauttaa, ovat sekä nämä varhaiset datajournalistiset jutut että datan visualisoinnit vielä 2010-luvulla aiheiltaan ja ulkoasultaan hämmästyttävän ajankohtaisia.

with end of Edge-hill. In its progress it knocked down several workmen, one of whom was so violently affected in the back of the head, that for relief he had recourse to bleeding; and a girl in a garden had her arm suddenly lifted up by its effort, and left it lacerated for some time after. The workmen of Mr. Duff were struck on the arm whilst on the line, but was merely stunned. A lady near Edgware, who was sewing at the time, felt in the fingers in contact with the needle a sensation resembling that of a slight electric spark. The electric fluid entered the house of Mrs. Clay, in Edge Vale, where its progress was not less alarming to the inmates, than destructive to the premises; and we have never heard of a more surprising escape, than that of the several individuals dispersed in a house, of which almost every room bears testimony of the ravages of the unaccountable element. It appears probable, from an examination of the apartments which the fluid has made, and the direction in which the bricks, timber, &c. have been found, that, attracted by the iron railing in front of the house, it entered the wall on one side the door, where it has shattered the bricks, torn to pieces the wood and brick-work between the door-pillar and the arch-way of the doorway, lifted the boards on the top; shook the fastenings to pieces, bearing part of the frame, and leaving a black soot on the joint-work; thence it passed up through the north door, splitting the bricks and the stone at the bottom of the chimney; the glass of which was shattered to pieces, and the whole frame disintegrated and forced into the house. Over the window it forced, in its way to the roof, a large hole, above which the soot appears exactly as if flames had issued from it. Its course appears next to have been towards the chimney: the eaves were shattered to pieces; the ridge stones displaced; many bricks and much cement torn from the wall; and the lead in many places forced up. It probably reached the rooms below by the chimney. In the lower room the eaves, plaster and paper are in several places broken, and the lead, as if scorched by its way, scorched the gilding of the chimney glass, and perched the top ornaments, but did not disturb the polished fire-iron, just below. Six squares of glass were driven out in this room. In the room above, in one corner, stood a bundle of rods, to which it made its way, perhaps from the chimney, between the lattice and the wall, so it forced off the plaster, and shot a quantity of it against a chest of drawers, where it struck the feet of a statue, evidently with

| DAY SCHOOLS.—Establishments. | Boys | Girls | Total | Ann. Exp. | Remarks. |
|------------------------------------|------|-------|-------|-----------|---|
| Grammar School | 155 | 155 | 310 | 1000 | |
| Blue Coat ditto | 80 | 80 | 160 | 2000 | Taught, clothed and boarded. |
| Green Coat ditto | 50 | 50 | 100 | 300 | Taught and clothed. |
| Collegiate Church ditto | 50 | 50 | 100 | 30 | And offertory money: do. do. |
| Strangers ditto | 10 | 10 | 20 | 100 | (Supposed)—Taught and clothed. |
| St. Mary's ditto | 12 | 12 | 24 | 40 | French arising from Sacramental Offerings. |
| St. John's ditto | 9 | 9 | 18 | 40 | (Supposed)—Expenses raised by voluntary Subscription. |
| St. Paul's ditto | 30 | 30 | 60 | 250 | Taught, clothed and boarded, by voluntary Subscription. |
| Ladies' Jubilee | 30 | 30 | 60 | 250 | (Supposed)—Taught and partly clothed. This School is supported by the benevolence of a single individual. |
| Black King-street | 21 | 21 | 42 | 40 | Voluntary Subscription, and Collection at Churches. |
| NATIONAL SCHOOLS, Granby-row | 191 | 119 | 310 | 600 | |
| Boltun-street, Suffolk | 300 | 170 | 470 | 1000 | |
| | 831 | 581 | 1412 | £5410 | |
| Dissenters. | | | | | |
| LANCASTRIAN SCHOOLS, Marshall-st. | 492 | 225 | 717 | 400 | Voluntary Subscription. |
| UNITARIAN, Mosley-street | 198 | 35 | 233 | 50 | ditto ditto |
| CATHOLIC | 198 | 121 | 319 | 104 | ditto ditto |
| SUNDAY SCHOOLS. | | | | | |
| Establishments. | | | | | |
| Collegiate Church, St. Paul's Hill | 201 | 205 | 406 | | |
| St. Ann's, Back King-street | 50 | 56 | 106 | | |
| St. Mary's, Back South Parade | 120 | 110 | 230 | | |
| St. Paul's, Green-street | 120 | 103 | 223 | | |
| Tower-street | 68 | 71 | 139 | | |
| Jersey-street | 314 | 281 | 595 | | |
| St. George's, St. George's | 141 | 112 | 253 | | |
| St. John's, St. John's-street | 118 | 163 | 281 | | |
| St. James's, St. James's-street | 102 | 158 | 260 | | |
| St. Michael's, Miller-street | 234 | 252 | 486 | £1023 | |
| St. Peter's, Jackson's-row | 120 | 120 | 240 | | |
| Alport Town | 95 | 95 | 190 | | |
| St. Clement's and St. Luke's | 353 | 1071 | 1424 | | |
| St. Stephen's, Blossom-street | 182 | 297 | 479 | | |
| Oldfield-road | 129 | 204 | 333 | | |
| Trinity, King's Head Yard | 220 | 260 | 480 | | |
| Holme, Duke-street | 182 | 169 | 351 | | |
| All-Saints, Oxford-road | 196 | 191 | 387 | 30 | |
| Arbwick | 60 | 110 | 170 | 25 | |
| | 3434 | 4212 | 7646 | £2078 | |

11, June 9, at 10, at Guildhall, London. A Mr. Toulson and Bannell, Thompson-street.

1841 Jonathan, of Sunderland, in the county of Durham, grocer, d. c.; May 12, 19, June 8 12, at Guildhall, London. Att. Messrs. G. Hobbs and T. T. Thompson-street.

PAYN Thomas, and John Daniel Pryn, of Calton-street, in the city of London, watchmakers d. c.; May 5, 12, June 9, at 1, at Guildhall, London. Att. Mr. Hudson, King's-bench-street.

SMITH John, now or late of Pattingham, in 4 dozens, in the county of York, grocer, in a draper, d. c.; May 11, 12, June 9, at 11, at Dor and Park Tavern, King's-square-Hall, Mr. Walsley, Hall.

TATE John, of Liverpool, in the county of L'caster, provision-merchant, d. c.; May 17, June 9, at 1, at the George Inn, Liverpool. Mr. Denton, Liverpool.

WARR Joseph, late of Basiley, in the county of Oxford, (but now a prisoner in the King's Bench Prison), lawyer, d. c.; May 5, 13, June 9, at 12, at Guildhall, London. Att. Messrs. F. and Munday, Holborn.

WHARFON Robert and Henry Wharton, of L'caster, in the county of Lancaster, joiners; house-occupiers, late occupants in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Att. Mr. Hudson, Liverpool.

WILMOTT Daniel, of Pinner-street, Botolph Claydon, in the county of Norfolk, master-marsh, d. c.; May 3, at 10, May 12, at June 9, at 10, at Guildhall, London. Messrs. Paterson and Pule, Old Broad-street.

DIVIDENDS.

May 22. Messrs. A. Ashmore, London, merchants.

22. J. Davie, Shrewsbury, six-penny.

22. W. Bewley, Manchester, tailor.

22. W. and A. Copp, Exeter, draper.

22. W. Broom, Exeter, draper and tailor.

22. J. Dobb, Staple-bury, Kent, draper & tailor.

22. J. Lee, London, watchmaker.

24. T. Cassidy, Liverpool, feather-merchant.

22. J. Williams, London, draper.

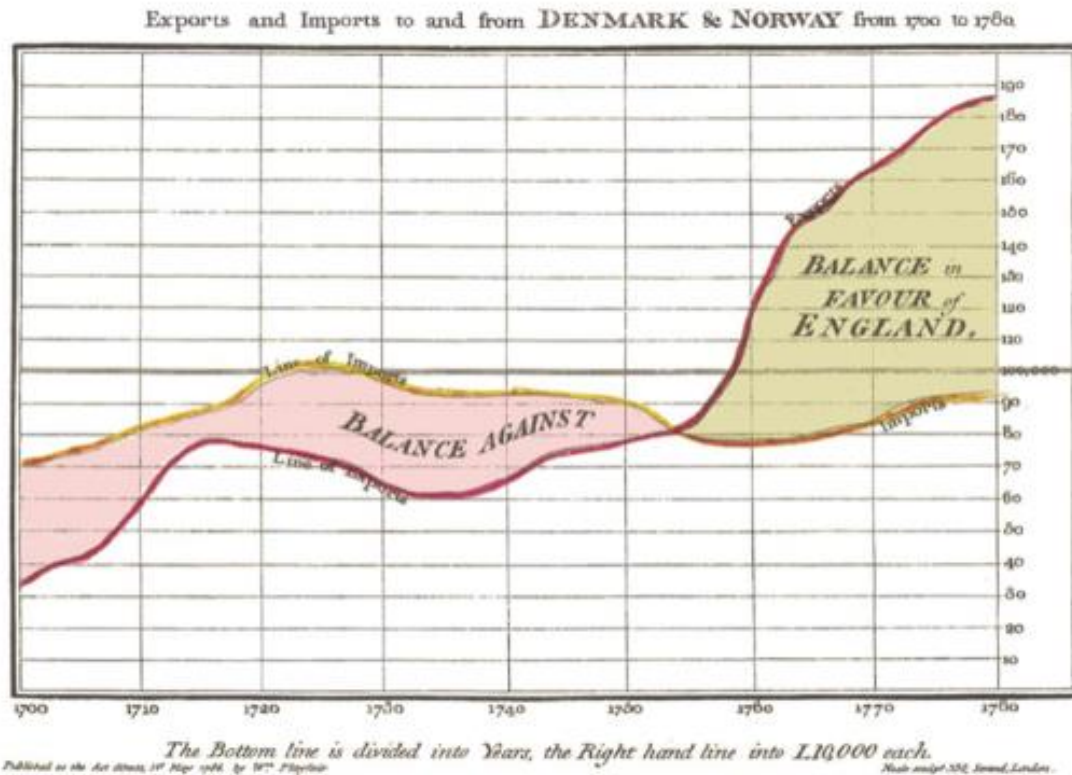
22. M. B. Schöninger, London, indigo-merchant.

June 5. Ryder & Nunnally, London, sugar-refiners.

SUBSCRIPTIONS OF PARTNERS.

John Harrison and Brothers, Manchester, oil-merchant. — Widow Welch and Sons, Manchester common carrier. — Geo. Ramsden and Co. 22

KUVA 5. Manchester Guardian -lehden juttu Manchesterin koulujen oppilasmääristä ja koulujen kustannuksista: datajournalismia vuodelta 1821 (Bounegru 2012, 20)



KUVA 6. William Playfairin vuonna 1786 julkaistu Atlas sisälsi 44 infograafia mukaanlukien tämän, joka kuvaa Englannin kaupankäyntiä Tanskan ja Norjan kanssa (Rogers 2014, 6)

3.2 Kohti tietokoneavusteista journalismia

Yhtenä iltana Yhdysvalloissa vuonna 1952 saavutettiin ainakin kolme merkkipaalua journalismin historiassa. Tuolloin televisio- ja radioyhtiö CBS ja kanavan uutisankkuri Walter Cronkite hyödynsivät UNIVAC-nimistä tietokonetta presidentinvaaleja seuranneessa lähetyksessään. Se oli tiettävästi ensimmäisiä kertoja maailmassa, kun tietokonetta käytettiin apuna tulosten ennustamisessa vaalilähetyksessä. Se oli myös ensimmäinen televisioitu vaalilähetyks, joka kattoi koko Yhdysvallat rannikolta rannikolle. Lisäksi se oli ensimmäinen presidentinvaalilähetyks, jonka Cronkite teki CBS:lle. (Henn 2012, viitattu 26.11.2014.)

UNIVAC ennusti republikaanien ehdokkaan Dwight D. Eisenhowerin ottavan murskavoiton demokraattien ehdokkaasta Adlai Stevensonista. Tietokone perusti arvionsa vaali-iltana siihen mennessä laskettuun noin kolmeen miljoonaan annettuun ääneen, mikä oli vain murto-osa lopullisesta äänimäärästä. Näiden alustavien tulosten perusteella tietokone ennusti Eisenhowerin saavan

sata ääntä jokaista Stevensonin ääntä vastaan. UNIVACin ohjelmoijat eivät olleet ottaneet tällaisia todennäköisyyksiä edes huomioon; Eisenhowerin povaaminen vaalien ylivoimaiseksi voittajaksi soti kaikkia ennako-odotuksia ja kyselytutkimuksia vastaan. Siksi tietokonetta vaali-iltana operoineet ohjelmoijat eivät tohtineet välittää tätä ensimmäistä ennustetta eteenpäin itse lähetykseen: ennuste vaikutti yksinkertaisesti liian naurettavalta. UNIVAC oli kuitenkin oikeassa, ja Eisenhowerista tuli kuin tulikin presidentti. (Henn 2012, viitattu 26.11.2014.)

Televisioyhtiöissä oltiin vielä seuraavina vuosina epävarmoja tietokoneiden käyttämisestä vaalien tulosten ennustamiseen. Vuoden 1954 kongressivaaleissa esimerkiksi televisioyhtiö NBC päätti, että ”hyvä toimittaja on aina parempi kuin mikään tilastokone”. Kuitenkin jo seuraavissa presidentinvaaleissa 1956 kaikki Yhdysvaltain kolme suurta televisiokanavaa käyttivät tietokoneita apunaan. (Ibid.)

8.30 P.M.

IT'S AWFULLY EARLY, BUT I'LL GO OUT ON A LIMB.

UNIVAC PREDICTS--with 3,398,745 votes in--

| | STEVENSON | EISENHOWER |
|-----------|------------|------------|
| STATES | 5 | 43 |
| ELECTORAL | 93 | 438 |
| POPULAR | 18,986,436 | 32,915,049 |

THE CHANCES ARE NOW 00 to 1 IN FAVOR OF THE ELECTION OF EISENHOWER.

KUVA 7. UNIVACin printtaamassa tulosteessa luki Eisenhowerin voiton todennäköisyyden kohdalla 00 eikä 100, koska ohjelmoijat eivät uskoneet tarvitsevana yli kaksinumeroisia todennäköisyyslukuja (Henn 2012, viitattu 2.3.2016)

Cox (2000, 8, viitattu 3.3.2016) viittaa Reavyyn, jonka mukaan tietokoneet ilmestyivät uutismedioiden toimituksiin vähitellen kolmessa vaiheessa, jotka ainakin osittain myös limittyivät toisiinsa. Vaiheet olivat *liiketoiminta*, *tuotanto* ja *informaatio*. Liiketoiminnallisen vaiheen alussa tietokoneita käytettiin pääasiassa kirjanpitoon ja levikkiin liittyvissä töissä. 1960-luvun puoliväliin

tultaessa toimitukset käyttivät tietokoneita jo monenlaisiin tehtäviin aina tavaraluetteloinnista sähköisten muotoilemisen. Vuosikymmenen vaihteessa tietokoneet olivat jo laajemmin käytössä toimitusten uutistuotannossa, mikä johti vähitellen informaatiovaiheeseen.

Merkittävä askel kohti tietokoneavusteista journalismia otettiin Yhdysvalloissa vuonna 1967. Detroitissa puhkesi tuolloin valtavat mellakat, jotka johtivat sekä siviilien että poliisin väkivaltaisuuksiin, ryöstelyyn ja lopulta armeijan väliintuloon. Kymmeniä ihmisiä kuoli, satoja loukkaantui ja tuhansia pidätettiin. Yli kaksituhatta rakennusta tuhoutui. (Younge 2011, viitattu 26.11.2014.)

Detroit Free Press -lehden toimittaja Philip Meyer ja psykologian professori Nathan Caplan sekä tutkija John Robinson Michiganin yliopistosta ryhtyivät selvittämään mellakoiden taustoja. Dataa kerättiin kyselytutkimuksilla ja sitä analysoitiin IBM 360 -tietokoneen avulla. Analyysi paljasti, että vastoin yleisiä käsityksiä henkilöiden taloudellisella asemalla tai koulutustaustalla ei ollut vaikutusta siihen, osallistuivatko nämä henkilöt mellakointiin vai eivät. Tutkimuksen mukaan suurimmat syyt levottomuuksien puhkeamiseen olivat poliisien väkivalta, huonot asumisolot ja työttömyys. Meyerin Pulitzer-palkittu juttu avasi kokonaan uuden aikakauden, jossa tietokoneita alettiin hyödyntää yhä enemmän journalismin tekemisessä. (Younge 2011, viitattu 26.11.2014; Cox 2000, 7, viitattu 3.3.2016.)

Vuosi Meyerin jutun jälkeen The Miami Herald -lehden toimittaja Clarence Jones palkkasi lakitieteen opiskelijoita syöttämään oikeusasiakirjoja tietokoneeseen. Jones ja lehden järjestelmäpäällikkö kävivät tietokoneen COBOL-ohjelmointikielen avulla läpi yhteensä 13 000 reikäkorttia dataa. Tutkimukset paljastivat väärinkäytöksiä Daden piirikunnan oikeuslaitoksen toiminnassa. Juttuprosessia on sanottu ensimmäiseksi kerraksi, kun julkisia rekistereitä on analysoitu tietokoneella journalistisessa tarkoituksessa. (Cox 2000, 7, viitattu 3.3.2016.)

Vuonna 1973 Philip Meyer – joka kuusi vuotta aiemmin oli tehnyt urauurtavaa työtä Detroitin mellakoiden tutkimisessa – julkaisi ensimmäisen version kuuluisasta kirjastaan Precision Journalism (suom. *tarkkuusjournalismi*). Kirjan uudistetussa versiossa vuodelta 1991 Meyer kuvailee journalismin muuttuneen 1970-luvun alussa “tieteellisemmäksi”, koska suuret tietomäärät alkoivat tulla paremmin toimittajien saataville tietokoneiden yleistymisen myötä. (Cox 2000, 8, viitattu 3.3.2016.)

Meyerin tarkkuusjournalismissa sosiologian ja käyttäytymistieteiden tutkimusmenetelmiä hyödynnetään osana journalistista työprosessia. Hänen visionsa oli, että tarkkuusjournalismia harjoittaisivat sekä journalismia että yhteiskuntatieteitä opiskelleet ammattilaiset perinteisissä mediataloissa. Tarkkuusjournalismin käsite syntyi osittain vastareaktionä 'uudelle journalismille', jossa yhdysvaltalaiset kirjoittajat kuten Jimmy Breslin, Gay Talese ja Tom Wolfe alkoivat käyttää kaunokirjallisia keinoja osana journalismia. (Meyer 1979, 9; Bounegru 2012, 19.)

He ovat lahjakkaita kirjoittajia - - ja kertovat, miltä tapahtumat näyttävät suodatettuina heidän omien uskomustensa, asenteidensa ja kokemustensa pohjalta. Henkilökohtainen journalismi ei sinänsä ole uusi ilmiö - - sen sijaan kaunokirjallisten keinojen käyttäminen [journalismissa] on uutta. Tapahtumien rakentaminen ja sisäinen monologi – jossa ei pelkästään kuvailla, mitä journalisti tekee ja sanoo, mutta mitä hän ajattelee – edustavat uutta tapaa kertoa eräänlainen [journalistinen] totuus. (Meyer 1979, 9.)

Kaunokirjallisuuden sijasta Meyer painotti datan keräämisen ja analysoimisen tieteellisiä menetelmiä journalismin pyrkimyksissä kohti objektiivisuutta ja totuutta. Hän katsoi tarkkuusjournalismin olevan vastaus journalismin yleisimpiin ongelmiin: riippuvuuteen lehdistötiedotteista ja sinisilmäisyyteen viranomaislähteitä kohtaan. Tällaiset journalismin 'heikkoudet' kumpusivat Meyerin mukaan informaatiotieteellisen osaamisen ja tieteellisen metodin puuttumisesta alan sisältä. (Bounegru 2012, 19.)

Meyer perusteli näkemystään journalismin tieteellistymisestä – tai siitä, miksi journalismin tulisi pyrkiä olemaan enemmän tieteen kaltaista – myös sillä, että tiede ja journalismi jakavat monia yhteisiä piirteitä. (Myer 1991, viitattu 5.3.2016). Näitä ovat *skeptisyys, avoimuus, kehitellyn teorian toiminnallistaminen, totuuden epävarman luonteen ymmärtäminen ja yksinkertaisimman selityksen etsiminen*. Sekä tieteellinen metodi että journalismi ovat Meyerin mukaan molemmat kiinnostuneita todellisuuden testaamisesta, mutta yleensä kahdella erolla. Ensinnäkin journalismilla on kovempi kiire testata todellisuutta, ja toiseksi journalistit ovat testaamisessaan yleensä passiivia. Sen sijaan, että he tekisivät itse omat havaintonsa, päätelmänsä ja kokeensa, he tyytyvät kuulemaan eri osapuolia erilaisine näkökulmineen ja pyrkimyksineen. Tämän menetelmän heikkous on, että journalistilla ei välttämättä ole tarpeeksi asiantuntemusta arvioimaan eri lähteitä tarpeeksi kriittisesti. Lisäksi tällainen metodi elää sen olettaman kanssa, että kaikki mielipiteet ja väitteet ovat yhdenvertaisia suhteessa todellisuuteen. (Ibid.)

Datan käsittelyn rooli oli Meyerin tieteellistä menetelmää imitoivassa journalismissa keskeinen. Meyerin (1991, viitattu 5.3.2016) mukaan tarkkuusjournalismissa toimittajan tuli pystyä

keräämään ja tallentamaan dataa (mieluiten tietokoneelle), noutamaan tallennettua dataa uudestaan joko omista tietokannoista tai muualta, analysoimaan dataa, tarvittaessa supistamaan suurta määrää dataa pienemmäksi sekä lopulta viestimään dataa eteenpäin ymmärrettävässä muodossa.

Yhdysvalloissa erityisesti tutkivat toimittajat alkoivat 1970-luvulla hyödyntää tietokoneita ja lähestyä tieteellistä metodia juttuprojekteissaan. Esimerkiksi New York Timesissa tuohon aikaan työskennellyt David Burnham analysoi New Yorkin poliisilaitoksen arkistomateriaaleja tietokoneella vuonna 1972. Hän hankki väestöluvut, rikosilmoitukset ja pidätystilastot New Yorkin eri poliisipiireistä. Tietokoneanalyysin pohjalta selvisi, että kaupungissa tehdyt rikosilmoitukset eivät vastanneet poliisin tekemiä pidätyksiä eri poliisipiireissä. (Cox 2000, 9, viitattu 5.3.2016.)

Seuraavana vuonna Philadelphia Inquirer -lehden toimittajat Don Bartlett ja James Steele tekivät Philip Meyerin avustamana juttusarjan Unequal Justice (suom. Epäoikeudenmukainen oikeus). He syöttivät Philadelphian alioikeuden paperimuodossa olleet tilastot tietokoneeseen ja muuttivat ne koneluettavaan muotoon analysoidakseen, minkälaisia tuomioita oikeus jakoi ja mitä suuntauksia tilastoista oli havaittavissa. (Ibid.)

New York Timesin Burnham jatkoi samana vuonna 1973 urauurtavaa työtään tietokoneavusteisen journalismin kehittämisessä. Hän halusi selvittää, oliko newyorkilaisen valkoisen keskiluokan pelko joutua mustien tekemien rikosten uhreiksi aiheellinen. Analysoimalla tietokoneen avulla oikeuden antamia tuomioita ja poliisin pidätystilastoja selvisi, että New Yorkissa mustaihoinen henkilö ainakin murhattiin kahdeksan kertaa todennäköisemmin kuin valkoihoinen. (Ibid.)

Mikroprosessoreilla varustetut pienemmät tietokoneet alkoivat yleistyä yhteiskunnassa ja sitä myöten myös uutistoimituksissa 1980-luvulla. Suomessakin herättiin, ja tietokoneet huomioitiin muun muassa työehtosopimuksissa: lehdistön TESiin neuvoteltiin vuonna 1980 näyttöpäättekorvaus, joka oli voimassa aina vuoteen 1995 asti (Oras, sähköpostiviesti 4.9.2015). Korvauksen ideana lienee ollut se, että tietokoneiden katsottiin lisäävän toimittajien työpanosta, jolloin tästä kasvaneesta työn arvosta haluttiin saada asianmukainen korvaus (Liski, keskustelu 3.9.2015). Jotkut suomalaiset toimittajat ovat saattaneet nähdä asian myös niin, että korvauksella kompensoitiin näyttöpäätteellä työskentelemisestä aiheutuneet 'haitat'. Suomen Journalistiliiton työehtoasiames Tytti Oras kertoo:

Työehtosopimuksella sovittiin muun muassa, että ”atk-perusteisten” lehdenvalmistusjärjestelmien käyttöönotosta tulee neuvotella paikallisesti. Atk-laitteiden käyttämistä myös vastustettiin ja tarina kertoo, että jossain jopa pelättiin näyttöpäätteen ”vievän luovuuden”. Sittemmin toki havaittiin tietotekniikan olevan varsin kätevä apu toimitustyössä. (Oras, sähköpostiviesti 4.9.2015.)

Aluksi yksittäiset toimittajat hankkivat omat koneensa, mutta myöhemmin lehdet alkoivat varustaa niillä toimituksiaan ja kustansivat työntekijöidensä tietokoneet. Ensi alkuun toimittajat kokivat tietokoneiden lähinnä korvaavan kirjoituskoneet tekstinkäsittelyssä. Yksi varhaisimmista yleisistä käyttötarkoituksista oli kuitenkin myös päästä käsiksi verkossa oleviin tietokantoihin ja hyödyntää niitä journalistisissa töissä. (Cox 2000, 10, viitattu 5.3.2016.)

Yksi 1980-luvun tietokoneavusteisen journalismin pioneereista oli yhdysvaltalainen Elliot Jaspin. Ollessaan vuosikymmenen puolivälissä Providence Journalin toimittajana hän pani merkille, että kuuden kuukauden aikana kolme lasta oli kuollut samalla tavalla: astuessaan ulos koulubussista ja jäätyään heti sen samaisen bussin alle, josta olivat juuri tulleet ulos. Jaspin alkoi tutkia tapauksia. Hän penkoi koulubussien kuljettajista koostuvaa tietokantaa ja löysi sieltä samoja henkilöitä, jotka esiintyivät myös liikennerikkomustilastoissa. Lisäksi hän vertasi kusmien tietokantaa oikeuden rekistereihin ja löysi bussikuskeja, jotka olivat tuomittuja huumekauppiaita. Jaspinin jutun myötä Rhode Islandin osavaltio uudisti koulubussien kuskeja koskevaa toimintaansa. (Ibid.)

Toinen juttu, jossa Jaspin käytti tietokonetta 1980-luvulla, julkaistiin sekkin Providence Journalissa. Hän kävi tietokoneen avulla läpi yhteensä 35 000 vähävaraisille tarkoitettua asuntolainaa Rhode Islandin osavaltiossa. Analyysi paljasti, että kaikkein alhaisimmat korot ja suurimmat lainat oli myönnetty osavaltion korkea-arvoisten virkamiesten lapsille. (Cox 2000, 11, viitattu 5.3.2016.) Juttu johti kaikkiaan 25 syytöseen ja lainoja myöntäneen instituution toiminnan täydelliseen uudistamiseen. (DeFleur 1997, 5.)

Klassinen esimerkki tietokoneavusteisesta tutkimisesta journalismista on Atlanta Journal and Constitution -lehdessä vuonna 1989 julkaisu juttusarja 'The Color of Money'. Toimittajat Bill Dedman ja Dwight Morris vertasivat väestödataa Georgian osavaltion metropolialueella kuuden vuoden aikana annettuihin asunto- ja remonttilainoihin, joita oli kaikkiaan 109 000 kappaletta. Tietokoneanalyysi paljasti, että valkoihoiset saivat Atlantan pankeilta ja muilta rahoituslaitoksilta viisi kertaa enemmän lainoja kuin samaan tuloluokkaan kuuluvat tummaihoiset. Paljastusten

myötä Atlantan yhdeksän suurinta pankkia ilmoittivat asettavansa haettavaksi 65 miljoonan dollarin edestä lainoja erikseen kaupungin tummaihoiselle väestölle. Juttusarja palkittiin tutkivan journalismin Pulitzerilla. (Gordon 1999, viitattu 10.3.2016; DeFleur 1997, 89–90.)

Samana vuonna Yhdysvalloissa perustettiin Missouri Institute for Computer-Assisted Reporting (suom. Missourin tietokoneavusteisen journalismin instituutti). Perustavina tahoina olivat tietokoneavusteisen journalismin pioneeri Elliot Japsin ja Missourin yliopiston journalismikoulu. (Houston 1999, viitattu 10.3.2016.) Instituutin tehtäväksi määriteltiin kouluttaa toimittajia etsimään, vapauttamaan ja analysoimaan elektronista informaatiota. Nimi vaihtui myöhemmin National Institute for Computer-Assisted Reportingiksi (NICAR). (Cox 2000, 11, viitattu 7.3.2016.)

Edellä mainitut esimerkit eivät ole suinkaan ainoita, joissa tietokoneita on hyödynnetty osana journalistisia työprosesseja 1970- ja 1980-lukujen Yhdysvalloissa, mutta ne antavat tarpeeksi tarkkan kuvan niistä lähinnä tutkivan journalismin suuntauksista, joissa tietokoneet toivat selvää journalistista lisäarvoa ja olivat jopa välttämättömiä juttujen valmistumisen kannalta. Margaret DeFleurin kirjasta *Computer-Assisted Investigative Reporting* (1997) löytyy kattava erittely muista tietokoneavusteisen journalismin esimerkeistä erityisesti 1980-luvulta. 1990-luvulla tietokoneavusteisen journalismin suuntaus oli Yhdysvalloissa levinnyt jo niin laajalle, että kattavan esityksen kokoaminen tämän aikakauden jutuista on käytännössä mahdotonta (Cox 2000, 11, viitattu 7.3.2016; DeFleur 1997, 85).

Parasie & Dagiral (2012, 9, viitattu 7.3.2016) viittaavat Garrisoniin (1998), jonka mukaan tietokoneavusteinen journalismi ei ollut enää 1990-luvun puolivälistä lähtien ainoastaan suurimpien lehtien työkalu, vaan käytäntö oli levinnyt myös keskikokoisiin ja pieniin uutisorganisaatioihin. Tietotekniikan kehittyminen ja koneluettavassa muodossa olevan datan saatavuus johtivat tietokantojen yhä laajempaan ja monimuotoisempaan journalistiseen käyttöön. Cox (2000, 11, viitattu 7.3.2016) viittaa Houstoniin (1996), joka erittelee kolme "tietokoneavusteisen journalismin perustyökalua", jotka olivat käytössä 1990-luvulla. Nämä olivat taulukkolaskentaohjelma, tietokantojen hallintaohjelma ja verkkolähteet. Taulukkolaskentaohjelmilla analysoitiin numeerista dataa, ja tietokantojen hallintaohjelmat olivat hyödyllisiä elektronisten lähteiden järjestelmissä. Tuon ajan verkkolähteitä olivat Houstonin (ibid.) mukaan sähköposti, keskusteluryhmät ja -foorumit sekä tietokantakirjastot. Myös tilastolliset ja kartoittamiseen käytetyt tietokoneohjelmat olivat yleisiä työkaluja.

Parasie & Dagiral (2012, 9, viitattu 7.3.2016) viittaavat Reisneriin (1995) ja Garrisoniin (1998), joiden mukaan toimittajat eivät myöskään enää käyttäneet tietokantoja ainoastaan informaation keräämiseen, tallentamiseen ja järjestämiseen. Niitä saatettiin julkaista lehdissä tai verkossa lukijoiden tutkittavaksi, ja toisinaan lukijoita pyydettiin myös täydentämään toimitusten omia tietokantoja.

Voidaan sanoa, että tietokoneiden, tietotekniikan ja datan hyödyntäminen journalismissa oli ennen tämän teknologian yleistymistä muutamien tutkivalla otteella työhönsä suhtautuvien uudisraivaajien harteilla. Vielä valtavirtaistuttuaankin 1980-luvulta eteenpäin tietotekniikan mahdollisuudet datan käsittelyssä olivat vahvasti sidotut tutkivan journalismin perinteeseen. DeFleur (1997, 91) kuitenkin huomauttaa, että tietokoneavusteisia työmenetelmiä käytettiin 1990-luvulla paitsi palkittujen tutkivien juttujen tekemiseen, myös rutiininomaisimpiin tai kevyempiin töihin. Tällaisia olivat muun muassa St. Petersburg Timesin väestödataa hyödyntäneet jutut: toisessa oli selvitetty että Floridassa asui 431 eskimoa, ja toinen oli sinkuille tarkoitettu juttu piirikunnista, joissa asui paljon yksin eläviä miehiä tai naisia. Kansas City Star -lehti puolestaan ylläpiti tietokantoja ajokorteista, rekisteröidyistä äänestäjistä ja piirikuntien kiinteistötiedoista. Tätä dataa analysoitiin laajoissa projekteissa, mutta sitä käytettiin myös nopeammissa töissä, kuten huumekauppiaille kämppiä vuokranneista vuokranantajista tai hälytysajoneuvoja ilman ajokorttia kuljettaneista palomiehistä kertovissa jutuissa. (Ibid.)

Tietokoneavusteisten työmenetelmien levinneisyyttä tutkittiin Yhdysvalloissa vuonna 1992. Kyselytutkimukseen vastasi 192 päätoimittajaa erikokoisista sanomalehdistä ympäri maan. Hieman yli puolet vastaajista kertoi, että heidän lehdissään oli aloitettu julkishallinnollisen datan analysoiminen tietokoneilla osana juttuprosesseja. Keskimäärin näissä lehdissä oli toteutettu seitsemän julkiseen dataan pohjautuvaa juttua. Kuitenkin joka kolmas julkishallinnon dataa analysoiva lehti oli tehnyt vähintään 11 tietokoneavusteista juttua vuodessa. Neljännos vastaajista myös uskoi, että toimitusten kyky analysoida julkishallinnon rekistereiden dataa tulee olemaan erittäin tärkeässä roolissa tulevaisuudessa. (DeFleur 1997, 92.)

Lopuksi on mainittava, että tämä katsaus tietokoneavusteisen journalismin historiaan on ollut hirvittävän keskittynyt Yhdysvaltoihin. Osaltaan lähdekirjallisuuden pohjalta voidaan varmaankin tehdä johtopäätös, että Yhdysvallat on ollut tietokoneavusteisen ja tutkivan journalismin perinteen edelläkävijä maailmassa. On kuitenkin aiheellista miettiä ja ainakin kysyä – lähestymättä aihetta

kuitenkaan tässä sen tarkemmin – onko näitä työmenetelmiä kehitetty muualla maailmassa, ja jos on, niin mihin aikaan ja millä tavalla?

Maininta Philip Meyerin Precision Journalism -kirjan toisessa laitoksessa osoittaa, että tietokoneavusteisen journalismin tai tarkkuusjournalismin menetelmiä kohtaan oli ainakin *kiinnostusta* Yhdysvaltain ulkopuolella jo 1970-luvulla.

Syksyllä 1978 joukkoviestintätutkimuksen kansainvälisen järjestön tapaamisessa Varsovassa pidettiin kurssi tarkkuusjournalismista. Eräs professori Neuvostoliitosta totesi, että se [tarkkuusjournalismi] oli kuin tuli – hyödyllistä, mutta mahdollisesti vaarallista. Kaksi puolalaista professoria väitteli, oliko [tarkkuusjournalismi] hyvä vai huono asia marxismin kannalta. Yksi professori Intiasta valitti, että se [tarkkuusjournalismi] kuormittaisi kolmannen maailman toimitusten budjetteja liikaa. Ja kanadalainen sanoi, ettei se [tarkkuusjournalismi] löisi itseään läpi Quebecissä ennen kuin aiheesta julkaistaisiin kirja ranskaksi. He kaikki ottivat asian vakavasti, ja se oli hyvä. (Meyer 1979, alkusanat kirjan toista laitosta varten.)

3.3 CARista datajournalismiin

On vaikea sanoa, milloin tietokoneavusteinen journalismi tarkalleen ottaen muuttui datajournalismiksi, saati mikä näiden kahden ero lopulta on. Bounegrin (2012, 21) mukaan datajournalismin on sanottu eroavan tietokoneavusteisesta journalismista siten, että se on nimensä mukaisesti kiinnostuneempi itse datasta. Siinä missä tietokoneavusteinen journalismi on työmenetelmä, jolla kerätään ja analysoidaan dataa parantaen näin yleensä tutkivan journalismin laatua, on datajournalismissa enemmän kyse siitä, kuinka data istuu osaksi koko juttuprosessia ja työnkulkua. Tässä mielessä datajournalismissa kiinnitetään jopa enemmän huomiota pelkkään dataan itsessään kuin sen ominaisuuksiin juttuaiheiden lähteenä tai juttuja jollakin tavalla parantavana elementtinä. Ajattelutapa näkyy Bounegrin mukaan siinä, että lehdet ovat alkaneet julkaista datasettejä juttujensa yhteydessä tai yksinään. (Ibid.)

Lorenzin (2010, viitattu 9.3.2016) mukaan tietokoneavusteinen journalismi on lähinnä tutkivan journalismin tekniikka, jolla etsitään dataa suurista tietokannoista tietokoneen avulla. Datajournalismi on puolestaan hänen mukaansa prosessi, joka vaikuttaa koko journalismin työnkulkuun perustavalla tavalla. Siinä data on analyysin, visualisoinnin ja ennen kaikkea tarinankerronnan lähtökohta. (Ibid.)

Coddington (2014, viitattu 10.3.2016) on tosissaan pyrkinyt erottamaan tietokoneavusteisen journalismin, datajournalismin ja *laskennallisen journalismin* toisistaan. Hänen mukaansa nämä kaikki työkäytännöt ovat niputettavissa yhden sateenvarjokäsitteen alle, jota hän kutsuu *määrälliseksi journalismiksi*. Kaikilla kolmella on kuitenkin Coddingtonin mukaan perustavaa laatua olevia eroja, ja hän luonnehtiikin niitä kaukaisiksi sukulaisiksi. (Ibid.)

Määrällisen journalismin kehitys käynnistyi Coddingtonin (ibid.) mukaan Philip Meyerin tarkkuusjournalismin työmenetelmistä, joita olivat yhteiskuntatieteistä tutut empiiriset menetelmät, kuten mielipidekyselyt ja muut katsaukset, sisällönanalyysit ja tilastolliset analyysit, joilla pyrittiin saamaan parempia vastauksia journalistisiin kysymyksiin. Tarkkuusjournalismista siirryttiin vähitellen tietokoneavusteiseen journalismiin (CAR), joka oli läheisesti kytköksissä yhdysvaltalaiseen tutkivan journalismin perinteeseen. Kytkös oli Coddingtonin mukaan joskus jopa taakka: toimituksissa CARia pidettiin yleisesti ainoastaan pitkien tutkivien projektien työkaluna. CARin puolestapuhujat saattoivat itse myötävaikuttaa tähän käsitykseen kutsumalla työmenetelmää toisinaan "uudeksi tutkivan journalismin muodoksi". Tietokoneavusteisesta journalismista on sitten siirrytty 2000-luvulla datajournalismiin (englanninkielisestä termistä on ollut jonkin verran kiistaa: joukkoviestinnän tutkijat suosivat käytettävän termiä 'data-driven journalism', kun taas ammattitoimittajien keskuudessa 'data journalism' on yleisemmin käytössä). (Ibid.)

Coddington (ibid.) toteaa, että datajournalismin määritelmät ovat olleet perinteisesti varsin laajoja aina sellaisista luonnehdinnoista lähtien, joiden mukaan datajournalismi on käytännössä mitä tahansa journalistista toimintaa, joka käsittelee dataa jollakin tavalla. Datajournalismia on myös määritelty sen kautta, kuinka se on lähentänyt eri aloja ja käytäntöjä toisiinsa: sitä on sanottu tilastollista analyysia, tietojenkäsittelytieteitä, visualisointia, verkkosivujen suunnittelua ja journalistista raportointia keskenään sekoittavaksi journalismin muodoksi. Datajournalismi on ollut myös läheisesti tekemisissä avoimen datan ja avoimen lähdekoodin työkalujen käyttöönottamisessa ja levittämisessä, vaikkakaan avoin data ei ole välttämättä tai ainoastaan datajournalismin kenttään kuuluva käsite. (Parasie & Dagiral 2013, Coddingtonin 2014, viitattu 15.3.2016, mukaan.)

Datajournalismi on Coddingtonin mukaan myös huomattavasti löyhemmin kytköksissä tutkivaan journalismiin kuin CAR. Lisäksi datajournalismi eroaa CARista siinä, että se painottaa datan

visualisointia visuaalisen suunnittelun ja journalististen arvojen välisen suhteen kautta. Edelleen datajournalismissa lukijoiden nähdään toimivan usein yhteistyössä toimittajien kanssa journalististen totuuksien ja moraalisten väittämien muodostamisessa. (Coddington 2014, viitattu 15.3.2016.)

Laskennallinen journalismi on puolestaan katsottu olevan yläkäsite, joka kattaa sekä tietokoneavusteisen journalismin että datajournalismin: “[Laskennallinen journalismi on] algoritmien, datan ja yhteiskuntatieteellisen osaamisen käyttämistä journalismin vahtikoira-funktion toteuttamiseksi” (Hamilton & Turner 2009, Coddingtonin 2014, viitattu 10.3.2016, mukaan). Tämä on kuitenkin Coddingtonin mielestä liian laaja määritelmä, jossa ei tehdä minkäänlaista eroa näiden kolmen työkäytännön (tietokoneavusteinen journalismi, datajournalismi ja laskennallinen journalismi) välille. Hän tarjoaakin omaa ehdotustaan laskennallisen journalismin määritelmäksi:

[Se on] teknologisesti suuntautuneen journalismin lajityyppi, joka keskittyy tietojenkäsittelyyn ja laskennallisen ajattelun soveltamiseen informaation esittämiseksi, keräämiseksi ja merkityksellistämiseksi datan ja yhteiskuntatieteellisten metodien yleisemmän journalistisen käytön sijaan (Coddington 2014, viitattu 15.3.2016).

Tällainen journalismi painottaa tiettyjen työkalujen käyttämisen asemasta tiettyä tapaa *ajatella laskennallisesti* ja tämän ajattelutavan pohjalta syntyviä prosesseja. Laskennallisen ajattelun käsite on peräisin 1900-luvun puolivälin tietojenkäsittelytieteistä, ja se rakentuu informaation *abstrahoinnin* ja *automaation* ympärille. (Ibid.) Abstrahoinnilla tarkoitetaan *abstraktion* suorittamista, eli tiedon yksinkertaistamista tai käsitteellistämistä: kykyä erotella ja pelkistää informaatiota tai ongelmia niiden välittömän asiayhteyden ohi. (Valpola 2000; ibid).

Coddington (2014, viitattu 15.3.2016) viittaa Wingiin (2008), jonka mukaan abstrahointi on laskennallisen ajattelun kulmakivi ja älyllinen, tiedostettu prosessi eikä varsinaisesti jokin tietokoneen suorittama työ tai tehtävä. Tietojenkäsittely on siten abstrahoidun informaation automatisointia. Coddington (2014, viitattu 15.3.2016) viittaa edelleen Flew'iin (2012), jonka mukaan tämä automatisointi tapahtuu usein algoritmien muodossa, joita pidetään toisinaan laskennallisen ajattelun kolmantena elementtinä abstrahoinnin ja automaation rinnalla.. Algoritmit voivat priorisoida, luokitella ja suodattaa informaatiota, ja niitä pystytään hyödyntämään journalismissa monella tavalla aina jakelusta uutisten kirjoittamiseen (Anderson 2013a; Carlson 2014, Coddingtonin 2014, viitattu 15.3.2016, mukaan).

Myös laskennallinen journalismi on datajournalismin tavoin nähty olevan CARin jälkeläinen. Coddington (2014, viitattu 15.3.2016) viittaa Gynnildiin (2014), joka on luonnehtinut Philip Meyerin olleen aikaansa edellä laskennallisen ajattelun soveltamisessa journalismiin. On myös todettu, että toimittajat ovat harjoittaneet laskennallista journalismia ja ajatelleet laskennallisesti vuosikymmenten ajan kuitenkin lokeroimatta työtapojaan mitenkään erityisesti 'laskennallisiksi' (Coddington 2014, viitattu 15.3.2016).

Coddingtonin mukaan laskennallisella journalismilla on yhtäläisyyksiä sekä CARin että datajournalismin kanssa (ibid.). Se menee kuitenkin tietokoneavusteista journalismia pidemmälle keskittyessään erityisesti tietojenkäsittelyn prosessointikykyyn informaation ryhmittelyssä, automatisoinnissa ja käsitteellistämässä. Ja vaikka sen yksi merkittävimmistä yhteisistä piirteistä datajournalismin kanssa on laskennallisten työkalujen käyttäminen ja yhteistyöhön perustuvat työprosessit datan analysoimiseksi ja esittämiseksi, ei kaikki datajournalismi kuitenkaan ole laskennallista. Laskennallinen journalismi toimii pääsääntöisesti informaation käsitteellistämisen kautta tuottaakseen algoritmisesti laskennallisia malleja. Datajournalismi toimii puolestaan lähinnä datasettien analysoimisen kautta tuottaakseen datapainotteista journalismia. (Stray 2011; Stavelin 2014, Coddingtonin 2014, viitattu 15.3.2016, mukaan.)

Coddingtonin käsitys näiden kolmen termin eroista on siis seuraava: ensinnäkin ne kaikki ovat erilaisia *määrällisesti* suuntautuneen journalismin muotoja. Tietokoneavusteisella journalismilla on juuret yhteiskuntatieteellisissä menetelmissä ja julkisen vallankäytön valvomiseen suuntautuneessa perinteisessä tutkivassa journalismissa; datajournalismille on luonteenomaista useita toimijoita osallistava avoimuus ja poikkiammatillinen yhteistyö; ja laskennallinen journalismi on keskittynyt informaation käsitteellistämisen ja automatisoinnin prosessien soveltamiseen. (Ibid.)

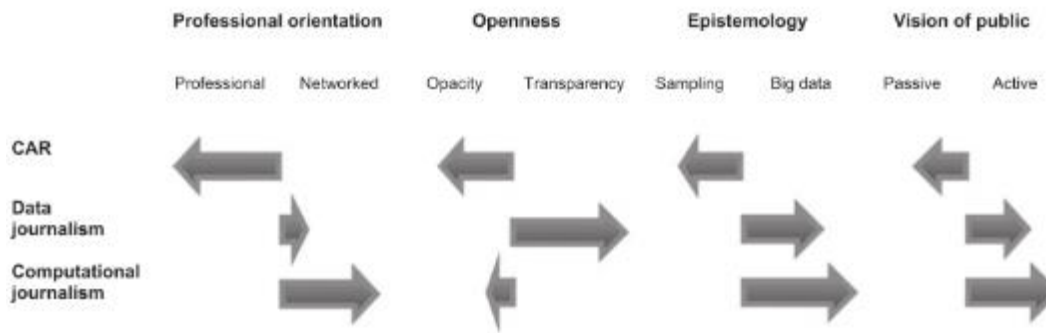
Selventääkseen johtopäätöksiään Coddington on kehittänyt luokittelujärjestelmän kuvaamaan tietokoneavusteisen journalismin, datajournalismin ja laskennallisen journalismin eroja. Tässä typologiassa luokittelu tapahtuu neljässä ulottuvuudessa, joissa kukin termi määrittellään suhteessa kahteen ääripäähän:

1. *journalistinen ammattitaito* – *verkostoitunut informaatio*, eli onko toiminta journalismin perinteisessä mielessä ammattimaista (ammattijournalistit) vai avoimempaa, verkostoituneempaa ja perinteisessä mielessä amatöörimäisempää;

2. *sulkeutuneisuus – läpinäkyvyys*, eli kuinka paljon työprosessia ja -menetelmiä avataan yleisölle;

3. *kohdistettu näytteenotto – 'big data'*, eli perustuuko journalistinen työ rajattuun otantaan jostain datasta vai kattavampaan datasettiin; ja

4. *passiivinen yleisö – aktiivinen yleisö*, eli nähdäänkö journalismin yleisö perinteisessä mielessä passiivisena vastaanottajana vai osallistuvana ja vuorovaikuttavana yleisönä.



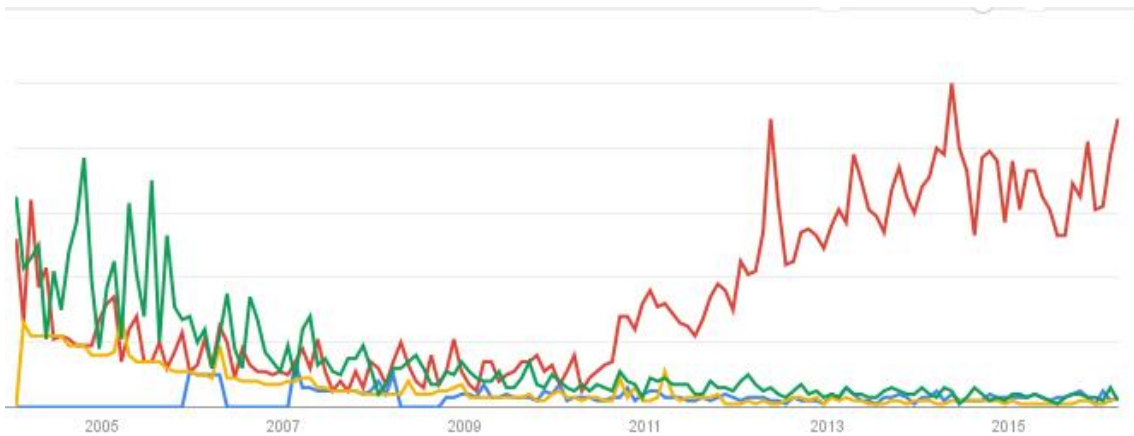
KUVA 8. Mark Coddingtonin luokittelujärjestelmä tietokoneavusteisen journalismin, datajournalismin ja laskennallisen journalismin eroista neljässä ulottuvuudessa (Coddington 2014, viitattu 10.3.2016)

Datajournalismi tuntuu joka tapauksessa olevan näistä termeistä nykyään kaikkein laajimmassa käytössä (ks. kuva 9). Tästä huolimatta vuonna 1989 Yhdysvalloissa perustettu Kansallinen tietokoneavusteisen journalismin instituutti (NICAR) on yhä olemassa ja toimii edelleen samalla nimellä kuin lähes 30 vuotta sitten. NICARin sivuilla tosin huomautetaan, että tietokoneavusteista journalismia käytettiin terminä viimeisten kahden vuosikymmenen ajan kuvaamaan niitä samoja asioita, joita nykyään kutsutaan datajournalismiksi. (Investigative Reporters and Editors 2016, viitattu 8.3.2016.)

Coddingtonin (2014, viitattu 15.3.2016) mukaan tietokoneavusteinen journalismi on terminä enää harvoin käytössä siitä yksinkertaisesta syystä, että tietokoneiden käyttämistä journalismin 'apuna' ei ole enää mielekäästä erikseen mainita. Sellaiset menetelmät, jotka aikoinaan luokiteltiin osaksi

tietokoneavusteista journalismia kuten tiedonhankinta verkosta, online-arkistojen käyttäminen ja sähköpostin välityksellä tehdyt haastattelut, ovat nykyään arkipäivää. Philip Meyer totesi saman asian jo vuonna 1999: ”Tietokoneavusteinen journalismi on kiusallinen muistutus, että olemme astumassa [uudelle vuosituhannele] ainoana ammattikuntana, jossa koetaan tarvetta alleviivata tietokoneiden käyttämistä. Kaikki muut pitävät tietokoneita itsestäänselvyytenä.” (Meyer 1999, viitattu 15.3.2016.)

Coddington (2014, viitattu 15.3.2016) arvioi datajournalismin ottaneen terminä valta-aseman suhteessa tietokoneavusteiseen journalismiin 2000-luvun ensimmäisen vuosikymmenen lopulla. Esimerkiksi Poynter-mediainstituutin sivuilla vuonna 2005 julkaistu dataa ja journalismia käsittelevä juttu oli vielä otsikoitu näin: ”A Guide to Computer Assisted Reporting (suom. Opas tietokoneavusteiseen journalismiin)” (Stith 2005, viitattu 15.3.2016).



KUVA 9. Google-hakujen suhteelliset määrät vuodesta 2004 hakutermeillä 'data journalism' (punainen), 'computer assisted reporting' (vihreä), 'computer assisted journalism' (keltainen) ja 'computational journalism' (sininen) (Google Trends, viitattu 15.3.2016)

Nykypäivän datajournalismin yhtenä ensimmäisistä sovelluksista pidetään EveryBlock-nimistä palvelua, jonka perusti yhdysvaltalainen web-kehittäjä Adrian Holovaty vuonna 2007 (Bounegru 2012, 12; Holovaty 2016, viitattu 16.3.2016). EveryBlock oli hyperlokaaliin paikallisuutisointiin keskittynyt palvelu, jonka kautta lukijat pystyivät etsimään juuri omiin asuinalueisiinsa liittyvää tietoa. Holovaty myös loi vuonna 2005 ChicagoCrime.org-verkkosivuston, joka on Bradshaw'n (2014, viitattu 17.3.2016) mukaan osa datajournalismin 'kaanonia' ja tietävästi ensimmäisiä kertoja, kun ulkopuolista dataa yhdistettiin Google Maps -karttasovelluksen kanssa (Holovaty

2016, viitattu 17.3.2016). Palvelussa Chicagon poliisin rikosdataa oli plotattu (piiretty) vuorovaikutteiselle kartalle, josta käyttäjät pystyivät tutkimaan, missä päin kaupunkia mitään rikoksia tapahtui. Tämä erosi perinteisestä tietokoneavusteisesta journalismista Bradshaw'n mukaan kahdella tavalla: lukijat saivat itse valita, mitä rikostietoja tutkivat ja mistä päin kartalta sen sijaan, että tämän valinnan olisi tehnyt toimittaja. Toiseksi kartan data päivittyi koko ajan automaattisesti; se ei "istunut" toimittajan kovalevyllä. (Bradshaw 2014, viitattu 17.3.2016.) Lisäksi Holovaty teki jo vuonna 2002 haettaviin tietokantoihin pohjautuvia projekteja Atlanta Journal Constitution -lehdessä. (Taylor 2009, viitattu 16.3.2016).

Holovatyn vuonna 2006 julkaisemassa esseessä 'A fundamental way newspaper sites need to change' hän perustelee, miksi journalistien ja lehtien tulisi alkaa julkaista jäsenettyä dataa koneluettavassa muodossa perinteisten tekstijuttujen rinnalla:

Useissa tapauksissa [perinteiset] lehtijutut eivät enää riitä. Suuri osa toimittajien päivittäin keräämästä tiedosta on strukturoitua informaatiota: sitä voidaan järjestellä ja analysoida automaattisesti tietokoneella. Ja silti tämä informaatio tislataan isoksi tekstimöykkyksi – lehtijutuksi – jota ei ole mitenkään mahdollista käyttää uudelleen. (Holovaty 2006, viitattu 16.3.2016.)

Holovaty ei tarkoita "uudelleen käyttämisellä" esimerkiksi uutisten lukemista puhelimella tai kämmentietokoneella – joka ilmeisesti vielä tuolloin oli käytössä oleva termi – vaan informaation kierrättämistä ja yhteen keräämistä sekä niin sanotusti idioottivarmaa tallentamista koneluettavassa muodossa. (Ibid.)

Sanotaan vaikka, että sanomalehti on kirjoittanut jutun tulipalosta. Jutun lukeminen puhelimella on hienoa ja mainiota – eläköön teknologia! Mutta todellisuudessa haluaisin pystyä tutkimaan yksitellen kyseisen tulipalon raakoja faktoja - - ja verrata näitä yksityiskohtia – kuten päivämäärää, kellonaikaa, paikkaa, uhreja, mikä paloasema oli kyseessä, kuinka pitkä matka tuolta paloasemalta oli itse paikalle, palomiesten nimiä, työkokemusta ja kauanko heillä kesti saapua paikalle – aikaisempien tulipalojen yksityiskohtiin. Ja seuraaviin, milloin ne sitten ikinä tapahtuvatkaan. (ibid.)

Holovatyn teksti inspiroi muun muassa palkittua toimittaja-koodaria Matthew Waitea perustamaan politiikan faktantarkistukseen keskittyvän Politifact-sivuston, josta tuli sittemmin ensimmäinen koskaan Pulitzerilla palkittu verkkosivusto (Holovaty 2016, viitattu 16.3.2016). Bradshaw on ehkä ihan aiheellisesti sanonut Holovatyn olleen "2000-luvun Philip Meyer" (Bradshaw 2016, viitattu 19.3.2016).

Politifactin perustaja Matthew Waite on todennut (Taylor 2009, viitattu 16.3.2016), että hänen oma kehityksensä 'tietokoneavusteisesta journalistista' toimittajaksi, joka osaa koodata, on ollut luonnollinen polku sellaiselle henkilölle, joka on jatkanut CARin parissa työskentelyä ja siihen liittyvien uusien taitojen ja työkalujen opiskelua. Holovaty (ibid.) näkee niin ikään, että CARiin erikoistuneet toimittajat ovat tehneet aikaisemmin hyvin samantyyppistä työtä, mitä modernit 'koodaavat toimittajat' nykyään tekevät. Sen sijaan uusi ilmiö on hänen mukaansa palkata koodareita tai ohjelmistokehittäjiä suoraan toimituksiin. (Ibid.)

Uutisorganisaatiot lähinnä Yhdysvalloissa ja Isonsa-Britanniassa ovat 2000-luvun ensimmäisen vuosikymmenen puolivälistä lähtien alkaneet palkata ohjelmoijia – tai 'toimittaja-koodareita', joskaan tämän tittelin suhteen ei olla alalla yksimielisiä (Pilhofer 2010, Parasie & Dagiral 2012, viitattu 16.3.2016, mukaan). Perinteiset sanomalehdet The New York Times ja The Guardian etunenässä sekä itsenäiset uudet toimijat kuten ProPublica ovat perustaneet toimituksiinsa työryhmiä suunnittelemaan niin sanottuja uusisovelluksia. Tällaiset uudet sisällöt ovat koostuneet verkonomaisista esitystavoista, kuten vuorovaikutteisista kartoista tai muusta datan visuaalisesta havainnollistamisesta. Näiden 'sovellusten' taustalla on laaja kirjo erilaista tietoteknistä osaamista datan keräämisestä, käsittelystä, analysoimisesta ja visualisoimisesta. (Parasie & Dagiral 2012, 3, viitattu 16.3.2016.)

Suomessa tällaista kehityskulkua edustaa esimerkiksi toimittaja ja tietotekniikan diplomi-insinööri Teemo Tebest, joka työskentelee Yleisradion PlusDeskissä (Tebest 2016, viitattu 17.3.2016). 'Plus' on Ylen syventävämpää ja taustoittavampaa journalismia julkaiseva osasto, jossa usein kokeillaan erilaisia verkonomaisia ilmaisukeinoja (YLE Plus 2016, viitattu 17.3.2016). Voitaneen tosin kyseenalaistaa, kuinka 'syventäviä' tai 'taustoittavia' sellaiset niin ikään PlusDeskin tekemät jutut kuin: "Millainen teini olet? Tee nuorten trendisanatesti" tai "Testaa, mikä liikunta sopii sinulle parhaiten – paini, sauvakävely vai jokin ihan muu?" lopulta ovat. (YLE Plus 2015, viitattu 17.3.2016; YLE Plus 2016, viitattu 17.3.2016.) Tällainen arviointi tosin vaatisi sen selvittämistä, mitä sanoilla 'syventävä' ja 'taustoittava' tarkoitetaan tai voidaan tarkoittaa, eikä moiseen urakkaan taida olla tämän tutkielman puitteissa viisasta ryhtyä.

Yksi koodareiden ja journalistien lähentyneen suhteen ilmentymä on vuonna 2009 perustettu Hacks/Hackers-organisaatio. Yhdysvalloissa syntynyt ruohonjuuritason ryhmä on kasvanut kansainväliseksi toimittajien ja koodareiden verkostoksi. Tavoite on ollut yhdistää kaksi perinteisesti erilaista ajatusmaailmaa miettimään yhdessä uutisten ja informaation tulevaisuutta.

'Hack' on vanha englantilainen lempi- tai haukkumanimi toimittajille, ja 'hacker' on yleisnimitys koodaajille, ohjelmistosuunnittelijoille ja muille tietotekniikkaa erityisen hyvin ymmärtäville henkilöille. (Vehkoo 2011, 91; Hacks/Hackers 2010, viitattu 17.3.2016.)

Dataa ja lukijoiden aktiivisempaa yhteiskunnallista osallistumista yhdistettiin Isossa-Britanniassa kuitenkin jo ennen Hologatya, koko nykyistä datajournalismin käsitettä ja ennen kuin toimittajat vielä puhuivat avoimesta datasta. (Bradshaw 2014, viitattu 17.3.2006). MySociety-niminen yhteiskunnallinen yritys – jolla on juuret vuonna 1996 perustetussa UK Citizens Online -ryhmittymässä – julkaisi ensimmäiset kansalaisaktiivisuutta tukevat sovelluksensa vuonna 2004 (mySociety 2016, viitattu 17.3.2016). Näitä olivat kansalaisten ja poliitikoiden keskinäistä viestintää helpottavat palvelut (mm. WriteToThem ja ReportEmptyHomes). Niitä seurasivat sovellukset, joiden avulla kansalaiset pystyivät tarkkailemaan, kuinka heidän edustajansa äänestivät (TheyWorkForYou) ja minkälaisia moraalisia ja yhteiskunnallisia arvoja edustajilla oli (Democracy Club). Bradshaw'n mukaan nämä olivat sellaisia rooleja ja funktioita, joita journalismi on perinteisesti ainakin väittänyt pyrkivänsä täyttämään. MySocietylla on myös Bradshaw'n mukaan ollut suuri vaikutus siihen ympäristöön, jossa datajournalistit työskentelevät ja kuinka he julkaisevat juttujaan. Sovellukset synnyttivät vuosia myöhemmin lukuisia kopioita ja vastaavia palveluja ympäri maailman, jotka ammensivat inspiaraatiota mySocietyn avauksista. (Bradshaw 2014, viitattu 17.3.2016.)

Vuoden 2010 Wikileaks-tietovuodot Afganistanin ja Irakin sodista pakottivat monet mediatilat opettelemaan kokonaan uusia taitoja valtavien datamassojen käsittelemiseksi: esimerkiksi Afganistanin sotaa koskevat lokit kattoivat yli 90 000 riviä dataa (Bradshaw 2014, viitattu 17.3.2016; Rogers 2011, viitattu 17.3.2016). Tämä tarkoitti käytännössä siirtymistä taulukkolaskentaohjelmien käyttämisestä SQL (Structured Query Language) -tietokantoihin. Nämä uudet tietotekniset vaatimukset olivat myös omiaan saattamaan koodareita ja journalistejä toimituksissa yhteisen pöydän ääreen. (Bradshaw 2014, viitattu 17.3.2016.)

Bounegrün (2012, 17) mukaan Euroopan journalismikeskus (EJC) järjesti samana vuonna yhden maailman ensimmäisistä kansainvälisistä datajournalismin konferensseista Amsterdamissa. Tuohon aikaan aiheesta ei käyty vielä kovin paljon syvällistä keskustelua, ja vain muutaman organisaation tiedettiin yleisemmin tekevän 'datajournalismia'. Viimeistään Wikileaksin tietovuotojen ja näiden tietojen journalistisen käsittelyn sekä laajan kansainvälisen huomion

jälkeen datajournalismi alkoi olla laajempaan tietoisuuteen levinnyt termi – tai jopa eräänlainen muotisana. (Ibid.; Lorenz 2010, viitattu 17.3.2016.)

Vuonna 2011 datajournalismin voitiin sanoa löytäneen tiensä ainakin yhdysvaltalaisiin toimituksiin, joissa käytettiin “ - - päivittäin paikallis-, osavaltion- tai liittovaltionhallinnon verkkosivuja datan etsimiseksi juttuja varten” (Aitamurto, Sirkkunen & Lehtonen 2011, viitattu 17.3.2016). Lisäksi Euroopan journalismikeskuksen samana vuonna teettämän kyselytutkimuksen mukaan suurin osa kyselyyn vastanneista toimittajista työskenteli jo tavalla tai toisella 'dataprojektien' parissa. Toisaalta suurimpana esteenä dataan pohjautuvien projektien aloittamiselle nähtiin, ettei datalla osattu välttämättä oikein tehdä mitään. Kiireellisintä koulutusta koettiin tarvittavan tarinoiden tai juttujen kehittämisessä datan pohjalta, mutta osaamista kaivattiin myös datan visualisoimiseen ja analysoimiseen. (Ibid.)

Joka tapauksessa datajournalismilla oli tuolloin jo keskeinen rooli monen uutisorganisaation verkkostrategiassa (ibid.). Datajournalismin nähtiin olevan osa organisaatioiden siirtymistä uutis- ja informaatio sivustoista kohti eräänlaisia *sovellusaloja*. Uutena ajatuksena oli, että mediat tarjoaisivat yleisölle vuorovaikutteisempaa sisältöä ja keinoja sisältöjen itsenäiseen jatkojalostamiseen. Datajournalismin nähtiin tarjoavan tähän mahdollisuuksia. (Ibid.)

Houstonin (2015, 9-10) mukaan datajournalismi on vuodesta 2001 lähtien alkanut levitä ympäri maailman. Yhdysvaltain lisäksi dataa analysoidaan nykyään osana uutistuotantoa niin Euroopassa, Latinalaisessa Amerikassa, Aasiassa, Afrikassa ja Australiassa. Datan journalistisen käyttämisen ja erityisesti datan visualisoinnin ympärille on syntynyt kansainvälisiä kilpailuja. Esimerkiksi Global Editors Networkin kilpailun (Data Journalism Awards) finaalissa oli vuonna 2015 yhteensä 78 osallistujaa maailman eri puolilta (Global Editors Network 2016, viitattu 19.3.2016). Kehityksen taustavaikuttajana on Houstonin (2015, 10) mukaan ollut Yhdysvaltain tutkivien toimittajien yhdistys ja tietokoneavusteisen journalismin instituutti, jotka ovat järjestäneet CAR-koulutusta jo lähes kolmen vuosikymmenen ajan. Myös Global Investigative Journalism Network (Maailmanlaajuinen tutkivan journalismin verkosto) ja sen joka toinen vuosi järjestämät, vahvasti CARiin painottuneet konferenssit ovat vaikuttaneet datajournalismin yleistymiseen. (Ibid.)

4 MIKSI JOURNALISMIN TULISI OLLA KIINNOSTUNUT DATASTA?

Luvun otsikko tiivistää sen ydinkysymyksen, johon tässä tutkielmassa pyritään kirjallisuuden ja asiantuntijoiden avulla (myöskin tiivistetysti) vastaamaan. Osa kysymyksen vastauksista on saattanut vilahtaa jo tutkielman aikaisemmillä sivuilla, mutta tässä luvussa ne pyritään kokoamaan yhteen.

4.1 Dataa, dataa kaikkialla

Ihmiskunta tuottaa valtavasti digitaalista tietoa: IBM:n mukaan 2,5 *kvintiljoonaa* tavua joka päivä. Se on luku, jossa ykkösen perään pitää laittaa 30 nollaa. (IBM 2016, viitattu 20.3.2016.) Digitaalisen informaation määrä vaikuttaisi myös kasvavan kiihtyvällä vauhdilla. Kaikesta maailman datasta 90 prosenttia on syntynyt viimeisten kahden vuoden aikana. (Ibid.) Jos tietoa oli ennen vanhaan niukasti saatavilla, on sitä nykyään liikaa. Mutta koska data esiintyy yhä useammin koneluettavassa muodossa, ähkystä on mahdollista selvitä.

Tietoa on sanottu uudeksi öljyksi. Vaikka moni toimittaja saattaa vielä menettää työpaikkansa teknologian kehityksen vuoksi, tullaan tietualan osaajia tarvitsemaan yhä enemmän. Tämä on yksi syy, miksi journalistien tulisi kiinnostua datasta ja sen mahdollisuuksista: ihmiset, jotka ovat hyviä käsittelemään suuria tietomääriä, ovat lähitulevaisuudessa – ja nyt – hyviä toimittajia. (Vehkoo 2011, 95, 110.) Niin mediayritykset kuin muidenkin alojen toimijat ja järjestöt etsivät jo nyt ihmisiä, jotka pystyvät sukeltamaan datamassojen sekaan ja tuomaan sieltä takaisin jotain järkevää, maailman ja asioiden tilaa selittävää tai muuten arvokasta tietoa. (Lorenz 2012, 4.)



KUVA 10. Infografiikka auttaa hahmottamaan, mitä netin suosituimmissa sovelluksissa tapahtuu yhden minuutin aikana (James 2015, viitattu 20.3.2016)

Vaikka toimittajat ovat käyttäneet tietokoneita jo kauan eivätkä periaatteet datan journalistisessa hyödyntämisessä ole juurikaan muuttuneet, yhteiskunta on muuttunut ja teknologia on kehittynyt. Kuten Bradshaw (2012, 2) toteaa, elämme digitaalisessa maailmassa, jossa lähes mikä tahansa asia voidaan esittää – ja usein esitetään – numeroin; siis tietokoneen bitteinä, ykkösinä ja nollina. Digitaalisen informaation valtavan määrän, sen helpon saatavuuden ja kehittyneen tietotekniikan hyödyntäminen ovat kenties ne ainoat 'uudet' asiat, joihin viitataan puhuttaessa datajournalismista. (Ibid.)

Tiedon tallennustila on nykyään valtava kehittyneempien kovalevyjen ja pilvipalveluiden myötä. Uudet kannettavat tietokoneet, tabletit ja älypuhelimet jättävät vain muutaman vuoden takaiset laitteet laskentateholtaan varjoonsa. (Houston 2015, 4.) Käytettävissä olevia ilmaisia työkaluja

datan hankkimiseen verkosta, sen analysoimiseen ja visualisoimiseen on enemmän kuin koskaan. Samoin julkista, koneluettavaa dataa on saatavissa ennennäkemättömät määrät avoimen datan portaaleissa, ja myös julkishallinnolle osoitettujen tietopyyntöjen tekeminen on helpottunut. (Radcliffe 2016, viitattu 20.3.2016.) Datan visualisoimisesta paremman ymmärryksen ja analyysin saamiseksi on tullut toimituksissa arkipäivää (Houston 2015, 4). Jutuissa käytetty data voidaan myös tarjota lukijoiden tutkittavaksi sellaisenaan sen sijaan, että sitä eriteltäisiin vain perinteisen tekstin tai staattisen uutisgrafiikan muodossa. Uusien työkalujen myötä datan analysointia ja läpikäyntiä voidaan tehdä lukijoille helpommaksi vuorovaikutteisilla sovelluksilla verkossa. (Howard 2014, viitattu 20.3.2016.)

4.2 Parempaa journalismia

Perinteisen median monopoliasema joukkoviestinnässä on jo kauan ollut ohi. Tänä päivänä internetin täyttävät uutisten lisäksi lukemattomat sanomat eri muodoissa: kenellä tahansa on mahdollisuus julkaista mitä tahansa. Maailman, maan, kaupungin tai vaikka tietyn kadun tapahtumista ei enää kerro välttämättä ensimmäisenä maakuntalehti tai paikallisradio. Se, joka sattuu olemaan älypuhelin taskussaan paikalla kun jotain tapahtuu, saa potentiaalisen skuupin. Uutiset ja tapahtumat virtaavat maailman puolelta toiselle lähes reaaliajassa; niitä jaetaan, pisteytetään ja kommentoidaan – tai ne jätetään kokonaan huomioimatta. Tällaisessa pirstaloituneen informaation yhteiskunnassa tapahtumien, ilmiöiden ja asioiden syvällisempi ymmärtäminen on yhä arvokkaampaa. (Lorenz 2012, 3.)

Käyttämällä dataa journalistisen työn painopiste muuttuu. Toimittajien ei enää tarvitse kilpailla koko muun maailman kanssa nopeudessa ja yrittää kertoa tapahtumista ensimmäisenä. Sen sijaan toimittajat pystyvät kertomaan, mitä nämä tapahtumat oikeasti *tarkoittavat*: kuinka esimerkiksi hankalasti hahmottuva uhka kuten työttömyys vaikuttaa ihmisiin iän, sukupuolen tai koulutuksen perusteella. Datan käyttäminen voi muuttaa vaikeaselkoisen, käsitteellisen asian kaikille ymmärrettäväksi. (Ibid.)

Yhdysvaltalainen professori ja entinen tutkiva journalisti Brant Houston havaitsi vuosituhannen vaihteessa, mihin maailma on informaation suhteen menossa, ja mitä haasteita se asettaa luotettavalle journalismille:

Toimittajat tarvitsevat hienostuneempia taitoja 2000-luvulla. Pikakirjoitus ei enää riitä. Pinnallisista, huonosti taustoitetuista kiistoista raportoiva journalismi ei enää riitä. Tilastojen ja numeroiden huolimaton käyttö ei enää riitä. Eikä kieltäytyminen tarjoamasta koulutusta [tietokoneavusteisesta journalismista] myöskään enää riitä. Ilman tällaista koulutusta journalistit eivät pysty erottumaan muista tiedonvälittäjistä tai juoru- ja huhupuheiden levittäjistä. (Houston 1999, viitattu 24.3.2016.)

Houston (2015, 6), huomauttaa, että tietokoneavusteisella journalismilla ei alun alkaenkaan viitattu juttujen kirjoittamiseen tietokoneella ja netissä surffaamiseen. Sen sijaan CAR tarkoitti – ja tarkoittaa yhä – tietokantojen hankkimista ja datan analysointia, jotta päivittäiseen journalismiin saadaan luotua syvyyttä ja muodostettua yhteyksiä käsiteltävien asioiden välille. Se tarkoittaa menetelmiä, joiden avulla saadaan kaivettua esiin vihjeitä, jotka puolestaan käynnistävät monitahoisempia juttuprojekteja laajemmasta näkökulmasta ja lisäävät toimittajan ymmärrystä käsittelemistään asioista. Esimerkiksi oikeusjärjestelmästä juttua kirjoittava toimittaja, joka tuntee systeemin suuret kuviot analysoituaan 150 000 tuomion tiedot, on useita askelia edellä verrattuna toimittajaan, joka näkee vain kourallisen tapauksia viikossa. (Ibid.)

Datan hyödyntäminen asettaa journalistit samalle viivalle poliitikkojen, byrokraattien, yritysmaailman edustajien ja etujärjestöjen äänitorvien kanssa. Aikaisemmin he ovat kaikki nauttineet etulyöntiasemasta suhteessa toimittajiin tietokantojen ja digitaalisen informaation hyödyntämisessä, koska heillä on ollut siihen enemmän rahaa ja muita resursseja. Viranomaiset ovat jo kauan syöttäneet rutiininomaisesti informaatiota tietokoneisiin ja analysoineet sitä. Sekä suuret että pienet yritykset käyttävät taulukkolaskentaohjelmia ja tietokantoja päivittäin. Etujärjestöt käsittelevät niin ikään toistuvasti dataa tavoitteidensa edistämiseksi. (Houston 2015, 7.) Toimittajilla on ollut tapana uskoa tällaisia virallisia lukuja: ne otetaan vakavasti, vaikka luvut olisivat täysin tekaistuja. 'Datanlukutaito' auttaa toimittajia terävöittämään kriittistä suhtautumistaan numeroihin ja saamaan jalansijaa taistellessaan organisaatioiden pr-osastojen kanssa. (Kayser-Bril 2012, 8.)

Tokio Shimbun -lehden toimittaja Isao Matsunami (2012, 8) on todennut, että datan merkitys tuli japanilaiselle medialle – joka on yleisesti ollut jäljessä digitaalisessa journalismissa – selväksi Fukushima ydinvoimalaonnettomuuden jälkeen.

Olimme hukassa, kun hallituksella ja asiantuntijoilla ei ollut antaa uskottavaa dataa [onnettomuuden] tuhoista. Kun viranomaiset salasivat radioaktiivisen materiaalin leviämistä ennustavan datan yleisöltä, [toimittajat] emme olleet valmistautuneet tutkimaan sitä [dataa] vaikka se olisi vuodettu meille. Vapaaehtoiset alkoivat itse kerätä radioaktiivisuusdataa

omilla laitteillaan, mutta meillä ei ollut osaamista tilastotieteestä, interpoloinnista, datan visualisoinnista ja niin edelleen. Journalisteilla on oltava pääsy raakaan dataan, ja heidän on opetettava olemaan luottamatta viranomaisten tulkintoihin siitä. (Matsunami 2012, 8.)

Lorenzin (2012, 4) mukaan toimittajat, jotka hallitsevat datan käsittelyn, huomaavat, että koviin faktoihin ja syvällisempään ymmärrykseen pohjautuvien juttujen tekeminen on helpotus: vähemmän arvailua, vähemmän kommenttien metsästämistä. Sen sijaan toimittaja pystyy muodostamaan käsittelemästään asiasta vahvan itsenäisen kannan, jota data-aineisto tukee. Tällainen työtapa voi vaikuttaa koko journalismin rooliin yhteiskunnassa. (Ibid.)

Suomalainen tutkiva toimittaja Jarno Liski kertoo, että hänen työssään data on usein avainasemassa:

Käytän ihan jatkuvasti jotain dataa - - se on ihan kaiken jutunteon osana se data. Jos on jokin asia, josta mulla on käsitys, että tästä on oltava jonkun viranomaisen jossain järjestelmässä faktatietoa, tarkat numerotiedot, niin pyrin ne sitten pyytämään itselleni ja katsomaan miten asia on. Mulle se on enemmän sellasta 'computer assisted reportingia'. Jos virkamies sanoo, että jonkun palvelun käyttö on vähentynyt voimakkaasti ja nyt me päätettiin sitä lopettaa se palvelu, niin mulle tyypillinen työtapa on sitä pyytää data, et johonkin se varmaan kirjataan sen palvelun käyttö. Mä pyydän siitä muutaman vuoden tiedot ja katson täsmälleen et moniko ihminen on sitä palvelua käyttänyt viis vuotta sitten ja moniko on käyttänyt sitä viime vuonna ja niin edes päin. Et se on niinku sen tyylistä se mun datan käyttö ihan 99-prosenttisesti. (Liski, haastattelu 17.2.2015)

Hyvä toimittaja haluaa aina nähdä alkuperäisen tietolähteen, koska joka kerta, kun muut ihmiset valitsevat ja lajittelevat tietoa toimittajan puolesta, se voi vääristyä joko tahallisesti tai tahattomasti, eikä vääristymää ole välttämättä helppo huomata. Datajournalismi voi estää tällaista tapahtumasta. (Houston 2015, 7.)

Bradshaw (sähköpostiviesti 9.11.2015) toteaa, että ihmisten sanomisten ja tekemisten referoimisen sijasta journalismin tulisi olla kiinnostunut siitä, puhuvatko nämä ihmiset totta ja kuinka heidän tekemisensä vaikuttavat muihin ihmisiin. Toisin sanoen: toimittajien pitäisi olla kiinnostuneita faktoista ja täten myös datasta. Ei pidä kuitenkaan olettaa, että data kertoisi *luonnostaan* totuuden: siihen on suhtauduttava kuin mihin tahansa muuhun tietolähteeseen. (Ibid.)

Kuten vuosikymmeniä takaperin, tietokannat ovat edelleen ihmisten kokoamia. Siten niissä on aina läsnä inhimillisen erehdyksen mahdollisuus, minkä vuoksi ne sisältävät puutteita ja virheitä,

jotka on otettava huomioon. Pelkkä tietokanta ei myöskään ole vielä journalistinen juttu (Houston 2015, 4; vrt. Holovaty 2009, viitattu 21.3.2016), vaan enemmänkin informaatiota kasvava pelto, jonka sato on korjattava huolellisesti ja varoen. Tietokantojen antaman informaation lisäksi tulee tehdä omia havaintoja ja haastatteluja, ja näitä tietoja on verrattava keskenään – ihan niin kuin monipuolisiin tietolähteisiin perustuvassa journalismissa on tapana. (Houston 2015, 4.)

Journalisteja – ei omia etujaan ajavia puolestapuhujia – tarvitaan tänä päivänä ehkä enemmän kuin koskaan toimittamaan yleisölle perusteellisesti tutkittuja näkökulmia informaatiosta ja datasta, sekä kertomaan niiden avulla mukaansatempaavia tarinoita. Ilman edes alkeellista tietämystä data-analyysin hyödyistä ja haittapuolista tämän päivän toimittajan on hankala ymmärtää ja kertoa, kuinka maailma nykyään toimii. Yleisön palveleminen ja vallan vahtikoirana toimiminen on myös tällöin paljon vaikeampaa. (Houston 2015, 4, 7.)

Periaatteet datan journalistisessa käyttämisessä ovat pysyneet kutakuinkin samoina vuosikymmenet. Philip Meyer toteaa kirjassaan Precision Journalism, että “ - toimittajat olisivat harvemmin väärässä, jos [he] omaksuisivat käyttöönsä yhteiskuntatieteiden tutkimusmenetelmiä”. (Meyer 1979, 3.) Tavoite kertoa maailmasta mahdollisimman faktapohjaisesti ei ole kadonnut journalismin keskiöstä yhtään mihinkään; jos se on vuosien mittaan johonkin liikkunut, niin ehkä yhä lähemmäksi alan kovinta ydintä. Totuuden arvo nousee sitä mukaa, mitä vaikeampi totuutta on kasvavasta informaation tulvasta löytää. Kysyntää olisi paljon, mutta tarjontaa niukasti. (Meyer 1999, viitattu 22.3.2016.)

Laadukkaampi journalismi voisi myös auttaa perinteisiä medioita selviämään kurjistuvasta taloustilanteesta. Kun kerrottavien uutisten ja tapahtumien nopeudessa ja määrässä ei voida kilpailla koko muun internetin kanssa, tulisi panostaa siihen ainoaan kilpailuvalttiin, joka vielä on tallella: laatuun. Jotta perinteiset uutisorganisaatiot voisivat kukoistaa, niiden on tehtävä journalismistaan jälleen tarpeellista ja välttämätöntä. Tämä tarkoittaa, että toimittajien tulee kohdistaa energiansa sellaiseen työhön, joka ei ole helposti korvattavissa. (Vehkoo 2011, 13, 21.) Kuutin (2015, 7) mukaan kerrottavien asioiden tarkkuus ja paikkansapitävyys ovat olennaisimpia tekijöitä yleisön journalismia kohtaan tuntemaan luottamuksen rakentumisessa. Kuutti (2015, 29) viittaa lisäksi Raivioon (2011, 2012), jonka tutkimuksissa osoittautui, että ainoa kiistaton korkeatasoisen journalismin ominaisuus on kerrottavien tietojen paikkansapitävyys. Yksistään tietojen paikkansapitävyys ei kuitenkaan vielä riitä tekemään journalismista korkeatasoista: toimittajan on selvitettävä näiden tietojen asiayhteydet ja taustat, minkä lisäksi toimittajan

näkökulmat ja johtopäätökset tulisivat perustua hänen keräämiinsä faktoihin. (Kuutti 2015, 29–30.)

Dataan perustuva journalismi nähdään paikkansapitäväksi nimenomaan datalle tehtävän tietoteknisen käsittelyn seurauksena. Tällaisten analyysien todistusvoima on parhaimmillaan varsin kiistaton, mikäli data-aineisto on luotettava ja analyysi on tehty huolellisesti. Toimittajan data-analyysin pohjalta tekemiä tulkintoja ja johtopäätöksiä ei kyseenalaisteta yhtä kärkeästi kuin pitkälti toimittajan intuitioon nojaavassa perinteisessä journalismissa. (Kuutti 2015, 125.)

Käyttämällä dataa on mahdollista julkaista sellaisia juttuja, joita yleisö ei muualta saa. Datajournalismi auttaa saamaan omia skuppeja ja erottumaan kilpailijoista. Laskevien levikkien ja pienenevien mainostulojen aikana tämäntyyppinen sisältö on medioille kullarvoista. (Stith 2005, viitattu 22.3.2016.)

Myös Vermanen (2012, 7) toteaa, että datajournalismi tarjoaa mahdollisuuden täyttää kaksi tärkeää tehtävää uutisorganisaatioissa: uniikkien juttuaiheiden löytämisen ja vallan vahtikoirana toimimisen. Nämä ovat sanomalehdille merkittäviä tavoitteita eritoten taloudellisesti huonoina aikoina.

Yhdysvaltalainen tutkiva televisiotoimittaja Paul Adrian (1999, 13, viitattu 22.3.2016) kertoo maan televisiokanavien havahtuneen median murrokseen jo vuonna 1998. Katsojaluvut olivat alkaneet laskea, ja yleisö alkoi siirtyä yleislähetysten parista kaapelikanaville ja internetiin. Televisiokanavilla ei tiedetty, mitä muuttuvassa mediaympäristössä selviytymiseksi tulisi tehdä. Adrian näki vastauksen piilevän tietokoneavusteisessa journalismissa ja datassa:

Näiden taitojen avulla pystymme tekemään uniikkia ja silmiinpistävää sisältöä niin päivittäisiin uutisiin kuin laajempiin projekteihin. Jos televisiolla on huono maine, se johtuu siitä, että olemme pinnallisia. Olemme loistavia live-lähetyksissä ja nopeassa uutisoinnissa, mutta yksityiskohdat jäävät meiltä usein hämäräksi. Kun tulevaisuuden yleisöt etsivät luotettavan tiedon lähteitä, pinnallisen [median] maine tappaa meidät. Tarvitsemme nyt enemmän kuin mitään muuta merkittävää sisältöä, johon katsojat luottavat. Tietokoneavusteinen journalismi auttaa kehittämään tällaista sisältöä ja luotettavuutta. (ibid.)

4.3 Kuinka se tehdään oikein – esimerkkinä Miami Herald & hurrikaani Andrew

Yksi tuhoisimmista Yhdysvaltoihin iskeneistä hurrikaaneista rantautui Floridaan 24. elokuuta 1992. Andrew'ksi ristitty hurrikaani aiheutti laajoja tuhoja: kymmeniä ihmisiä kuoli, ja aineelliset vahingot nousivat 25 miljardiin dollariin vuoden 1992 valuutassa laskettuna. Hurrikaanin tuulet puhalsivat kovimmillaan 78 metriä sekunnissa. (Wikipedia 2014, viitattu 24.3.2016.)

Miami Herald -lehden toimittaja Steve Doig ryhtyi selvittämään hurrikaanin aiheuttamia tuhoja. Hän hankki käsiinsä kaksi erillistä datasettiä: hurrikaanin aiheuttamat rakennusvauriot ja tuulenvoimakkuudet paikkatietoineen. (Liski, haastattelu 17.2.2015; O'Murchu 2012, 10.) Doig yhdisti nämä tiedot päällekkäin kartalle, mistä hän teki huomion, että suurimmat rakennusvauriot eivät olleetkaan siellä, missä oli ollut suurimmat tuulenvoimakkuudet (Liski, haastattelu 17.2.2015). Havainto ei vastannut yleistä olettamusta, että suurimmat tuulenvoimakkuudet olisivat aiheuttaneet myös suurimmat vauriot rakennuksiin.

Usein spekuloin, että sellainen suomessa tai maailmallakin nykyaikainen datajournalistinen juttu jäisi hyvin herkästi tähän - - että julkaistaisiin vain nettiin, et "katso kartalta missä oli kovimmat tuulenvoimakkuudet ja missä oli kovimmat rakennusvauriot". Mutta ne [Herald] sitten lähti selvittään, että mistä tää johtui ja jalkautti toimittajat ihan perinteiseen tutkivaan journalismiin - -. (Liski, haastattelu 17.2.2015.)

Heraldin ja Doigin tutkimukset paljastivat, että paikallisessa rakennusvalvonnassa oli korruptiota. Rakennusfirmat olivat pystyneet rakentamaan taloja, jotka eivät olleet tarpeeksi kestäviä. Tämä oli onnistunut osaksi suoraan vaalirahoituksen ansiosta, osaksi taas siitä syystä, että presidentti Ronald Reaganin toimikaudella oli vähennetty virkamiehiä ja supistettu julkista sektoria. Rakennusvalvojilla oli yksinkertaisesti liikaa työmaita valvottavanaan, ja tämän seuraukset nähtiin sitten, kun hurrikaani iski. (Ibid.) Lehden tekemät paljastukset palkittiin julkisen palvelun Pulitzerilla vuonna 1993. (O'Murchu 2012, 10; The Pulitzer Prizes 2016, viitattu 24.3.2016.)

Siinä [Heraldin jutussa] oli hyvää se, että data-analyysin lopputulokseen suhtauduttiin lähinnä sen juttuprosessin alkupisteenä, että data kerto miten on, mutta sitten toimittajat lähti selvittään että miksi tää on näin ja minkä takia tää on merkityksellistä ja mitä tähän ehkä liittyy. Se millä se pulitzer voitettiin, niin se kaikki oli sellasta että ei se mistään datasta löydy, mutta datasta saatiin se näyttö, että tiedettiin että jotain täs on pielessä ja lähdetiin selvittään että mitä siinä on pielessä. (Liski, haastattelu 17.2.2015.)

Vaikka Heraldin juttu on vuodelta 1992, se on yhä tänä päivänä – lähes 25 vuotta ilmestymisensä jälkeen – mahtava esimerkki datan käyttämisestä journalismissa (Liski, haastattelu 17.2.2015; O'Murchu 2012, 10).

Dataa saatetaan usein käyttää sen itsensä vuoksi, jotta saadaan vaikkapa tehtyä näyttävän näköistä grafiikkaa, joka ei kuitenkaan välttämättä kerro mitään kenellekään (Vehkoo 211, 111). O'Murchu painottaa nimenomaan 'journalistista' näkökulmaa puhuttaessa datajournalismista.. Sen sijaan, että dataa analysoitaisiin tai visualisoitaisiin huvikseen tai koska visualisoinnit näyttävät hienoilta, tulisi sitä ajatella työkaluna, jolla päästään lähemmäksi totuutta yhteiskunnassa tapahtuvista asioista. (O'Murchu 2012, 9.)

Myös Yleisradion 'atk-toimittaja' Teemo Tebest sanoo, että datan käyttämisellä olisi aina oltava harkitut perustelut.

- - Millon datajournalismi on ylipäänsä onnistunutta - - tai että meillä on jotain aineistoja, me hyödynnetään se verkossa jollain tavalla. Niin ehkä se pointti on nimenomaan siinä, että se mitä me ikinä tehdäänkään sillä datalla tai mikä se esitystapa on, niin sen täytyy jollain tavalla olla sellainen, että se antaa mulle lisäinformaatiota siihen aiheeseen. Mikä se sit ikinä onkaan se tapa miten se tehdään. Mut semmonen niinku vaan et meil on dataa ja visualisoidaan se, mut se ei anna mulle mitään, se ei havainnollista mulle mitään mitä mä en muuten jo tietäis. Niin se on aina vähän turhaa. (Tebest, haastattelu 18.2.2015.)

Tebestin (ibid.) mukaan datajournalismi ymmärretään nykyään usein lopputuotteen kautta, esimerkiksi verkossa klikattavana karttana tai muunlaisena esitystapana. Hän kuitenkin toivoisi, että datajournalismi olisi enemmän aineistojen analysoimista osana toimitustyötä, jonka lopputulos olisi sitten tämän työn ilmentymä muodossa tai toisessa.

Datajournalismiin saisi enemmän tulla nimenomaan sitä tutkivaa puolta mukaan. Et ei vaan kerrottais että mitä on, vaan että miksi. Että päästäs sille tasolle, joka nimenomaan olis sitä tutkivaa. Et ei kerrottais, että tässä on nyt tällainen data-aineisto ja sä voit selata sitä näin. Vaan kerrottais, et mitä tämä tarkoittaa. Et päästäs vastaamaan niihin kysymyksiin että miksi joku data-aineisto näyttää joltain, eikä vaan kerrota että tässä nyt ovat ne top-5 köyhintä kuntaa, vaan kerrottas että miksi ne on top-5 köyhintä kuntaa. (Tebest, haastattelu 18.2.2015.)

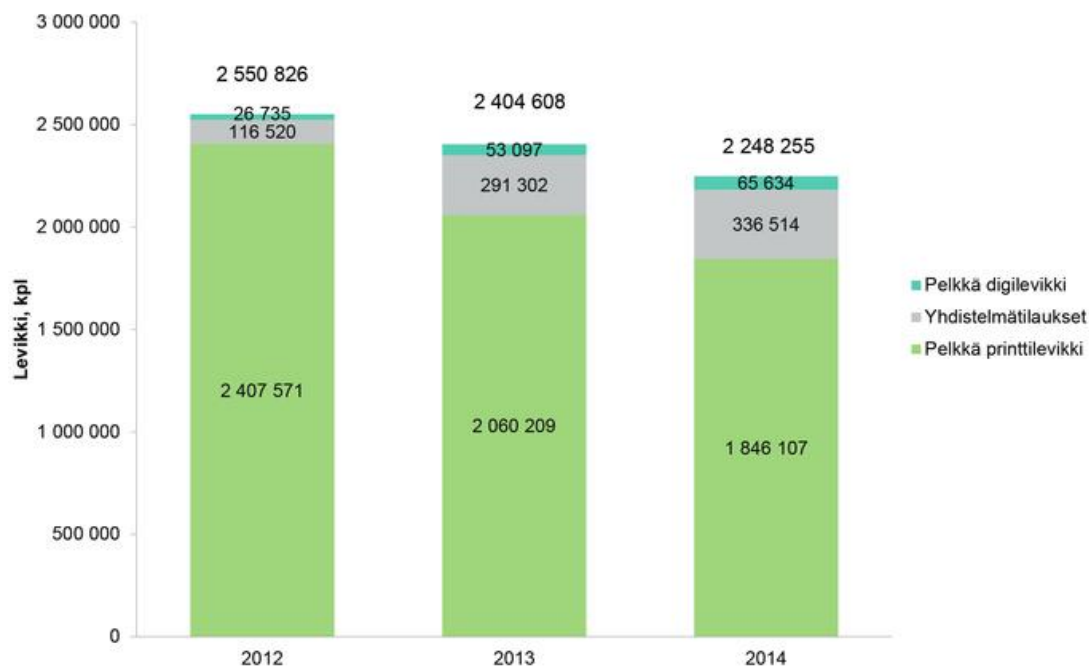
5 TYÖKÄYTÄNNÖT SUOMALAISESSA MAAKUNTALEHDISTÖSSÄ

Osana tätä tutkielmaa on haluttu kartoittaa, kuinka laajassa käytössä dataan pohjautuvat työmenetelmät ovat suomalaisessa maakuntalehdistössä. Maakuntalehdellä tarkoitetaan tässä yhteydessä vähintään kolmesti viikossa ilmestyvää sanomalehteä, joka keskittyy pääasiassa tietyn maakunnan tai alueen asioista uutisoimiseen (Kuutti 2012, 168). Iltapäivälehdet, puoluelehdet ja muut erikoissanomalehdet on siis jätetty tämän tarkastelun ulkopuolelle.

Maakuntalehdistö valikoitui tutkimuksen kohteeksi, koska sillä on ollut perinteisesti vahva asema Suomessa. Maakuntien kakkos- ja kolmoslehdet ovat surkastuneet pois, ja jäljelle jääneet alueiden päälehdet ovat olleet taloudellisesti hyvässä kunnossa. Niillä on ollut monopoliasema sekä uutis- että ilmoitusmarkkinoilla, ja niitä ovat pönkittäneet Suomessa muiden Pohjoismaiden tapaan vankka lukemisen ja lehden tilaamisen perinne. (Vehkoo 2011, 32–33.)

Tästä huolimatta painettujen sanomalehtien levikki on Suomessa tasaisessa laskussa. Maan kaikkien sanomalehtien printtilevikki tippui 23 prosenttia vuosina 2012–2014. (Sanomalehtien Liitto 2016, viitattu 25.3.2016). Sanomalehtien Liittoon kuuluvien sanomalehtien lukumäärä on myös pudonnut: kun vuonna 2000 lehtiä oli 207 kappaletta, vuonna 2014 niitä oli 182. Sanomalehtien digilevikki kuitenkin yli kaksinkertaistui vuosien 2012 ja 2014 välillä. Toisaalta digitaalisten tilausten osuus oli vuonna 2014 yhä varsin vaatimaton sanomalehtien kokonaislevikistä: noin 2,9 prosenttia, siinä missä printti kattoi yksinään yhä yli 80 prosenttia levikistä. (Ibid.)

Lehdissä myös käydään yt-neuvotteluja ahkerasti. 17:ssä sanomalehtiä kustantavassa yrityksessä käynnistettiin yhteistoimintaneuvottelut pelkästään vuonna 2015. (Suomen Ammattiliittojen Keskusjärjestö 2016, viitattu 25.3.2016.) Vehkoo (2011, 18) viittaa Nikuseen (2011), jonka mukaan taloudellinen lama ja mainostulojen sekä levikkien lasku on antanut suomalaisille sanomalehdille syyn toteuttaa pitkään hautuneita suunnitelmia. Taustalla on ollut muun muassa varautuminen Yhdysvaltain tilanteeseen, missä useita lehtiä on lopetettu talousvaikeuksien takia. Lama on myös ollut tilaisuus hankkiutua eroon vanhoista työntekijöistä, joiden kouluttaminen uuden teknologian osaajiksi ei ole houkuttanut. (Vehkoo 2011, 23.)



KUVA 11. Sanomalehtien levikkikehitys Suomessa vuosina 2012–2014 (Sanomalehtien Liitto 2016, viitattu 25.3.2016)

Vehkoon (2011, 29) mukaan suomalaisissa toimituksissa journalismin muutoksiin on vastattu varovaisella puolustuspelillä. Kuluja on karsittu leikkaamalla toimituksista, ja niistä on samalla tullut entistä päällikkövetoisempia, jolloin yksittäisten toimittajien itsenäisyys on kaventunut. Kutistuneissa toimituksissa työn määrä kasvaa, eikä analyyttiselle tutkivalle journalismille riitä kiireen takia aikaa. Erilaisia ratkaisuja ja uusia ideoita ei niin ikään ole aikaa pysähtyä miettimään. Vehkoo toteaaakin, että: “Sanomalehti on kuin raskas valtamerilaiva, jonka kääntäminen on hidasta puuhaa. Useimpina päivinä se vain puuskuttaa eteenpäin, sivuilleen katsomatta.” (Vehkoo 2011, 30–31.)

Vehkoo ei kuitenkaan ole täysin menettänyt toivoaan. Hänen mukaansa suomalaisilla mediataloilla on mahdollisuus vaalia printtiä ja kehittää digitaalisia sisältöjä: niillä on tähän sekä aikaa että rahaa. Vain innostusta, ideoita ja rohkeutta puuttuu. Ja niitäkin olisi, jos toimittajat vain uskallettaisiin “vapauttaa avokonttorikarsinoistaan”. (Vehkoo 2011, 31.)

Sanomalehtien heikentynyttä taloustilannetta ja laskevia levikkejä sekä kutistuvien toimitusten kulttuuria vasten on kiinnostava kysyä, missä määrin data ja siihen liittyvät journalistiset työmenetelmät ovat osa maakuntalehtien arkea. Nähdäänkö datan käyttäminen esimerkiksi keinona tehdä parempia sisältöjä ja saada uusia lukijoita? Kuinka tärkeä rooli datalla on osana lehtien journalismia ylipäätään?

5.1 Kysely maakuntalehtiin datan käyttämisestä

Datajournalismin roolia maakuntalehdistössä lähdettiin selvittämään lehtien omasta näkökulmasta, eli tarkoitus oli kerätä tietoa lehtien työntekijöiltä kyselylomakkeella (liite 1). Tavoitteena oli kerätä suhteellisen luotettavaa tietoa siitä, kuinka paljon ja minkälaista datajournalismia maakuntalehdissä tehdään, sekä miksi dataa käytetään tai vastaavasti ei käytetä. Lomake sisälsi sekä avoimia että suljettuja kysymyksiä, ja se toteutettiin Google Forms -työkalulla.

Tutkimuksen kohteeksi valikoitui yhteensä 19 maakuntalehteä ympäri Suomen. Pyrkimyksenä oli lähettää kysely maakuntien niin sanottuihin ykköslehtiin siten, että jokaisesta maakunnasta olisi mukana yksi lehti. Poikkeuksen muodostivat Lappi, Etelä-Savo ja Kymenlaakso, joissa jokaisessa ilmestyy kaksi maakuntalehteä. Pohjois-Savon maakuntalehti Savon Sanomat jäi tutkimuksen ulkopuolelle, koska tutkija unohti kyseisen lehden olemassaolon kyselyä laatiessaan. Ahvenanmaan Ålandstidningen koki saman kohtalon. Helsingin Sanomat ei puolestaan ollut mukana tutkimuksessa, koska datajournalismilla tiedetään olevan lehdessä vankahko jalansija (ks. Nieminen 2015; Toikkanen 2014), ja tämä tutkimus keskittyi nimenomaan niihin maakuntalehtiin, joissa ei ainakaan tiettävästi ole esimerkiksi omaa datadeskiä ja datatoimittajaa. Helsingin Sanomat voidaan lisäksi katsoa olevan enemmän valtakunnallinen lehti kuin tiukasti tietyn maakunnan asioihin keskittynyt julkaisu.

Ennen kyselylomakkeen lähettämistä soitettiin erikseen jokaisen lehden päätoimittajalle tai vastaavalle päällikköportaan työntekijälle. Tarkoitus oli selvittää, kenelle kyselylomake kannattaa kussakin lehdessä lähettää, jotta siihen saisi parhaan mahdollisen vastauksen. Oletus oli, että yksi toimituksen työtävät ja rutiinit perusteellisesti tunteva henkilö riittäisi vastaamaan kustakin

lehdessä. Kysely ei siis ollut suunnattu kaikille lehtien toimittajille vaan kustakin lehdestä yhdelle – useimmissa tapauksissa jollekin päällikölle – henkilölle, joka osaisi kertoa yleisellä tasolla toimituksensa työmenetelmistä.

Linkki kyselyyn lähetettiin joulukuussa 2015 sähköpostilla seuraaviin maakuntalehtiin: Aamulehti, Etelä-Saimaa, Etelä-Suomen Sanomat, Hämeen Sanomat, Ilkka, Itä-Savo, Kainuun Sanomat, Kaleva, Karjalainen, Keskipohjanmaa, Keski-suomalainen, Kouvolan Sanomat, Kymen Sanomat, Länsi-Savo, Lapin Kansa, Pohjalainen, Pohjolan Sanomat, Satakunnan Kansan ja Turun sanomat.

Vaikka kyselyn otantaan kuului 19 maakuntalehteä, lähetettiin kysely vain 15 henkilölle, koska joissain tapauksissa yksi ja sama henkilö vastasi useamman samaan konserniin kuuluvan lehden puolesta.

5.2 Kyselyn tulokset

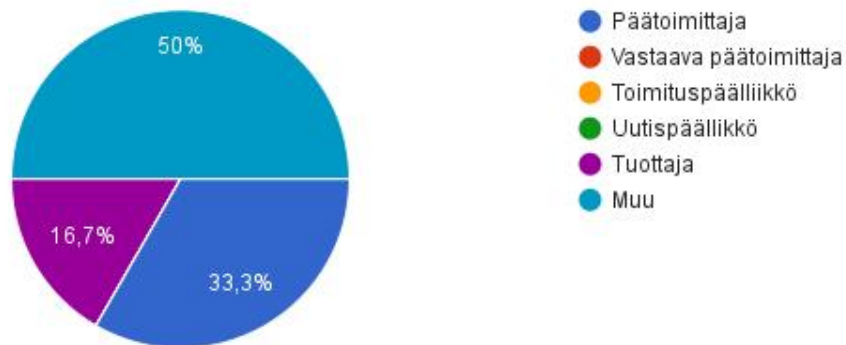
Kyselyyn vastasi lopulta kuusi henkilöä, jotka edustivat kaikkiaan yhdeksää maakuntalehteä. Vastaukset saatiin seuraavista lehdistä: Ilkka & Pohjalainen; Kouvolan Sanomat, Kymen Sanomat & Etelä-Saimaa; Keski-suomalainen; Karjalainen; Kaleva ja Turun Sanomat.

Kun kysely lähetettiin yhteensä 15 henkilölle, joista kuusi vastasi, oli kyselyn vastausprosentti 40. Näiden henkilöiden edustamat yhdeksän lehteä kattoivat lähes puolet (47 prosenttia) kaikista kyselyn 19 lehdestä.

Vastaajissa oli kaksi päätoimittajaa, yksi tuottaja, yksi verkkopäällikkö, yksi erikoistoimittaja ja yksi verkkotuottajan sijainen.

Ilkan ja Pohjalaisen vastaukset oli niputettu yhteen, samoin oli Kouvolan Sanomien, Kymen Sanomien ja Etelä-Saimaan.

Titteli / toimi (6 vastausta)



KUVA 12. Kyselytutkimuksen vastaajien prosentuaalinen jakauma heidän työnimikkeensä perusteella. Puolet vastaajista kuului luokkaan 'muu' (Google Forms 2016, viitattu 26.3.2016)

Kyselyssä pyydettiin kertomaan, tehdäänkö vastaajien edustamissa lehdissä datajournalismia vai ei. Termi oli määritelty kyselyssä seuraavasti:

Datajournalismi on työprosessi, jonka aikana hankitaan, muokataan ja analysoidaan dataa. Tämän prosessin lopputuloksena julkaistaan datajournalistinen juttu ja mielellään myös jutussa käytetty data. Kuten jo datajournalismin nimestä käy ilmi, on tässä prosessissa olennaisesti mukana myös perinteisempi käsitys journalismin tekemisestä. Datajournalismissa yhdistyvätkin journalismin työtavat datan käsittelyyn, ohjelmointiin ja visualisointiin.

Lisäksi tässä opinnäytetyössä datajournalismiksi on laskettu myös 'datajournalistiset työtavat' osana juttujen tekemistä, vaikkei juttua/juttuprosessia varsinaisesti datajournalismiksi muuten kutsuttaisikaan, eikä jutusta esimerkiksi tehtäisi verkkolehteen mitään uutissovellusta tai muuta visuaalista, verkonomaista toteutusta.

Tällaisia 'datajournalistisia työtapoja' ovat muun muassa tietopyyntöjen tekeminen julkishallinnolle, datan hankkiminen jollain muulla tavalla esimerkiksi toimituksen omista tietokannoista tai vaikkapa 'raaputtamalla' dataa verkosta, datan käsitteleminen ja sen käyttäminen jutun ideana, pohjana, alkupisteenä tai jollakin tavalla osana juttua.

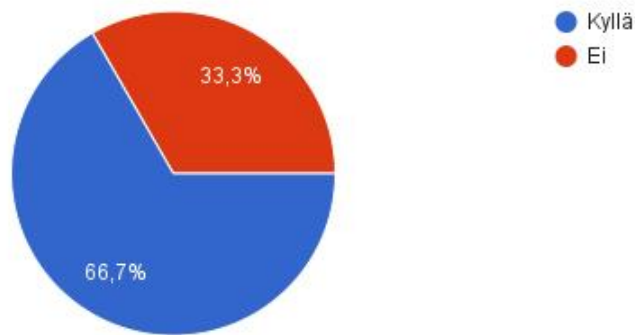
On huomionarvoista, että monet (data-)journalismin ammattilaiset eivät kuitenkaan pidä esimerkiksi jotakin tilastoaineistoa referoivaa juttua vielä datajournalismina. Usein tällaisia mediaa kiinnostavia tilastoja ovat erilaiset barometrit, yritysten osavuosikatsaukset tai tilastokeskuksen julkaisemat aineistot.

Alalla vallitseva käsitys siis on, että jostakin tällaisesta 'datasta' uutisoiminen referoivaan tyyliin, jossa datalla/datalle ei varsinaisesti ole tehty mitään, ei vielä ole datajournalismia. Sen sijaan vaikkapa em. Tilastoaineistojen hyödyntäminen jollakin tavalla ja niissä

esiintyvän datan analysoiminen, muokkaaminen tai jokin muu käyttäminen osana juttuprosessia on datajournalismia. (liite 1.)

Tehdäänkö mediassanne yllä olevaan kuvailuun sopivaa datajournalismia?

(6 vastausta)



KUVA 13. Kyselyn vastaajien prosentuaalinen jakautuminen sen mukaan, tehdäänkö heidän toimituksissaan datajournalismia vai ei (Google Forms 2016, viitattu 26.3.2016)

Vastaajista neljä ilmoitti, että heidän toimituksissaan (Kaleva; Karjalainen; Keski-suomalainen; Kouvolan Sanomat, Kymen Sanomat & Etelä-Saimaa) tehdään datajournalismia. Kaksi (Ilkka & Pohjalainen; Turun Sanomat) kertoi, että datajournalismia ei tehdä lainkaan.

Seuraavaksi käydään läpi vastauksia niistä lehdistä, joissa datalla kerrottiin olevan jonkinlainen rooli toimitustyössä. Vastaajia pyydettiin kuvailemaan omin sanoin, minkälaista datajournalismia heidän edustamiensa lehtien toimituksissa tehdään ja antamaan jutuista esimerkkejä, mikäli se oli mahdollista. Lisäksi heiltä kysyttiin, *miksi* dataa käytetään lehden journalismissa ja olivatko vastaajat tyytyväisiä tällaisen journalismin määrään sekä laatuun toimituksissaan.

Keskisuomalaisesta vastattiin, että mahdollisuutta datajournalismin tekemiseen harkitaan lehdessä tapauskohtaisesti: ”Aineistoina voi olla esimerkiksi kaupungin keräämä aineisto tai tilastokeskuksen tuottama data. Esimerkkeinä verotietojen hyödyntäminen, interaktiivisten karttojen laatiminen tai esimerkiksi kaupungin julkaisemien ostolaskujen tutkiminen.”

Kalevan vastauksessa sanottiin, että heillä aihepiirejä datajournalismin tekemiseen ei ole rajattu mitenkään.

Verkossa [on] paljon juttuja, jossa yksi tai useampi data on tuotu/visualisoitu kartalle. Paikallisuus ja alueellisuus [ovat] keskiössä, kuten muussakin sisällössä. Printissä karttajuttujen lisäksi käytetään infografiikkaa ja esim. Puumaisia kaavioita datan visualisoinnissa.

Kaleva oli vastaajista ainoa, josta annettiin konkreettisia esimerkkejä lehden tekemistä datajournalistisista jutuista. Esimerkkejä oli kaikkiaan neljä, ja ne oli otsikoitu näin: Karttasovellus kertoo, kuinka huoltosuhde romahtaa suurkunnissakin; Rikkailla entistä paremmat tulot – katso kartasta missä rikkaat asuvat; Näissä paikoissa parkkisakkoja napsahtaa eniten Oulussa – katso sakkorysät kartalta; Yritystukipotti kasvoi roimasti Pohjois-Pohjanmaalla – katso, ketkä saivat eniten tukia.

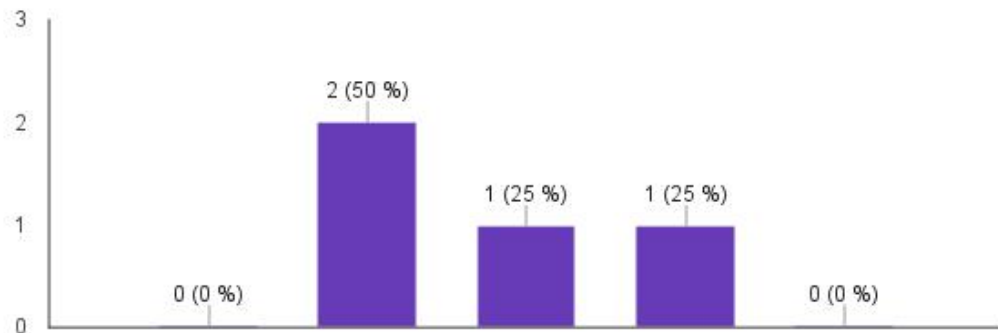
Sanomalehti Karjalaisesta kerrottiin datajournalismin tekemisestä näin: ”[Teemme datajournalismia] monista sellaisista aihepiireistä, joista meidän teknisillä (=vaatimattomilla) keinoillamme dataa on saatavissa. Toimintaa rajoittaa tietoteknisten keinojen ja osaamisen vaje.”

Kouvolan Sanomien, Kymen Sanomien ja Etelä-Saimaan yhteinen verkkopäällikkö kertoi, että näissä lehdissä tehdään ”Lähinnä vain ns. perinteisen mallin juttuja, eli käydään läpi tilastoja ja perustetaan juttu niihin. Ei [tehdä] juuri lainkaan datajournalismia ns. modernissa muodossa eli infograafit jne.” Kaikki lehdet, jotka ilmoittivat tekevänsä datajournalismia, sanoivat tekevänsä sitä sekä printtiin että verkkoon.

Kysyttäessä kuinka tärkeä osa datajournalismi on toimitusten työtä asteikolla 1-5, jossa 1 tarkoitti 'ei lainkaan merkittävä' ja 5 tarkoitti 'erittäin merkittävä', vastaukset vaihtelivat. Sanomalehti Karjalainen ja Kouvolan Sanomat, Kymen Sanomat sekä Etelä-Saimaa arvioivat datajournalismin roolin toimituksissaan vastaavan tällä asteikolla arvoa 2. Kaleva antoi arvoksi 3 ja Keskisuomalainen 4.

Arvioikaa, kuinka tärkeä osa datajournalismi on toimituksenne työtä?

(4 vastausta)

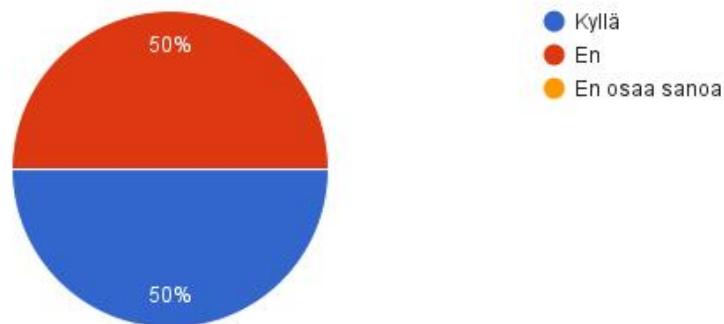


KUVA 14. Kyselyn vastausten prosentuaalinen jakautuminen sen mukaan, kuinka tärkeäksi datajournalismi toimituksissa koettiin asteikolla 1-5. Pystysuora y-akseli edustaa vastaajien määrää ja vaakasuora x-akseli asteikon pisteitä 1-5

Yhdessäkään lehdessä ei ollut datajournalismiin erikoistuneita toimittajia tai erillistä datatoimitusta. Kalevassa, Kouvolan Sanomissa, Kymen Sanomissa ja Etelä-Saimaassa datajournalismia tehdään vastausten mukaan vähintään kerran viikossa. Keski-suomalaisessa ja Karjalaisessa dataa käytetään journalismin tuottamiseen vähintään kerran kuukaudessa. Yksikään vastaajista ei ollut tyytyväinen lehtensä tekemän datajournalismin määrään. Keski-suomalaisessa ja Kalevassa oltiin kuitenkin tyytyväisiä tehdyn datajournalismin laatuun. Karjalaisessa, Kouvolan Sanomissa, Kymen Sanomissa ja Etelä-Saimaassa laatuun ei oltu tyytyväisiä.

Oletteko tyytyväinen toimitukseenne tekemän datajournalismin laatuun?

(4 vastausta)



KUVA 15. Kyselyn vastausten jakautuminen sen mukaan, oltiinko lehdistä tyytyväisiä niissä tehdyn datajournalismin laatuun. Kalevassa ja Keski-suomalaisessa laatuun oltiin tyytyväisiä; Karjalaisessa, Kouvolan Sanomissa, Kymen Sanomissa ja Etelä-Saimaassa ei

Karjalaisesta kerrottiin, että "datajournalismin kehittäminen on jäänyt akuutimpien töiden varjoon". Kalevassa oltiin laatuun tyytyväisiä, mutta sieltäkin todettiin, että "toiveissa on toki saada aikaiseksi nykyistä kunnianhimoisempia toteutuksia".

Kouvolan Sanomissa, Kymen Sanomissa ja Etelä-Saimaassa "[datajournalismia] pitäisi pystyä tekemään modernilla tavalla, mutta resurssipulan takia joudumme menemään helpoimman kautta".

Oletteko tyytyväinen toimitukseenne tekemän datajournalismin määrään?

(4 vastausta)



KUVA 16. Yhdessäkään lehdessä ei oltu tyytyväisiä tehdyn datajournalismin määrään

Vastaajia pyydettiin myös kertomaan, miksi heidän edustamissaan lehdissä tehdään datajournalismia.

Keskisuomalaisessa syyt datan käyttämiseen olivat “Oman uutistyön tekeminen, aineiston hyödyntäminen parhaalla mahdollisella tavalla, lukijan palveleminen ja esimerkiksi päätöksenteon tekeminen läpinäkyväksi.”

Karjalaisesta vastattiin, että “Datajournalismin keinoin on mahdollista yhdistellä tietoja siten, että syntyy uuta [sic] tietoa, mikä on tutkivan journalismin keskeinen tehtävä.”

Kalevassa datajournalismi nähtiin yhtenä työkaluna siinä missä muutkin: “Teemme journalismia erilaisilla työkaluja ja -tapoja hyödyntäen [sic]. Datajournalismi on yksi työkalu muiden joukossa. Tiettyjen asioiden jäsentäminen on tehokkainta datajournalistisin menetelmin.”

Datan käyttämistä perusteltiin samansuuntaisesti myös Kouvolan Sanomissa, Kymen Sanomissa ja Etelä-Saimaassa: “Tilastot jne. ovat materiaalin ja jutun lähde siinä missä mikä tahansa muukin.”

Kyselyyn vastanneista maakuntalehdistä datajournalismia ei tehty Ilkassa, Pohjalaisessa ja Turun Sanomissa. Näitä vastaajia pyydettiin perustelemaan, mistä tämä mahdollisesti johtui.

Ilkassa ja Pohjalaisessa nähtiin kaksi syytä, miksi lehti ei ole ottanut datajournalistisia työmenetelmiä käyttöönsä:

1. Resurssipula. Työn alla on sopivien työkalujen etsiminen ja tuunaaminen meille sopiviksi. Aikapula kuitenkin rajoittaa kehitystyötä.
2. Luulen, että osalle talon väkeä sana datajournalismi ei ole aivan selkeä, eikä varsinkaan tiedetä, miten sitä tehdään. Eli kun työkalut ovat olemassa, pitää vielä kouluttaa toimitus ymmärtämään, miten niitä hyödynnetään ja mitä hyötyä niistä on toimittajalle itselleenkin.

Turun Sanomien perustelut kuuluivat näin: “Määritellyssä muodossa välttämättä ei [tehdä datajournalismia]. Tilastoja, barometreja ja dataa käsitellään ja analysoidaan kyllä monissa jutuissa talousuutisista politiikkaan ja kulttuurista urheiluun.”

Ilkassa ja Pohjalaisessa oli kuitenkin aikeita ruveta tekemään datajournalismia tai käyttämään kyseisiä menetelmiä toimitustyössä: “Ehdottomasti. Olen itse puhunut datajournalismin puolesta jo lähes 10 vuotta sitten (silloin toisen työnantajan palveluksessa). Se on erittäin helppo ja käyttökelpoinen tapa tuoda lukijoille lisäarvoa.”

Turun Sanomista vastattiin, että lehdessä ei suunnitella datajournalismin tekemistä tulevaisuudessakaan.

5.3 Johtopäätökset

Kyselytutkimus oli helppo kohdistaa halutulle perusjoukolle (Suomen maakuntien 'ykköslehdet'), koska tämä joukko oli yleisesti tiedossa eikä se ollut kohtuuttoman suuri. Näin ollen perusjoukosta oli mahdollista saada täydellisen kattava otanta. Tämä mahdollisuus kuitenkin torpedoitui heti alkuunsa tutkijan huolimattomuuden vuoksi, minkä johdosta kyselyä ei lähetetty kaikkiin Suomen maakuntien ykköslehtiin. Helsingin Sanomat puolestaan jätettiin pois tarkoituksella.

Kyselyn osoittaminen oikeille henkilöille varmistettiin soittokierroksella. Tällä haluttiin taata, että kyselyyn saataisiin mahdollisimman paljon vastauksia. Jokaisesta lehdestä saatiin puhelimitse kiinni joku henkilö, joka osasi kertoa, kenelle toimituksen väestä kysely kannatti lähettää.

Vastausprosentti jäi kuitenkin alle puoleen (kuusi vastaajaa 15:stä, eli 40 prosenttia). Vaikka otos oli lopulta pienehkö, oli se ainakin edustava: ketään ei etukäteen suosittu yli muiden, vaan kaikilla oli yhtä suuret mahdollisuudet vastata kyselyyn. (ks. Tampereen yliopisto 2003, viitattu 28.3.2016).

Vastanneista neljä ilmoitti, että heidän lehdissään tehdään datajournalismia. Kuitenkin vain yksi (Kaleva) pystyi antamaan tekemästään datajournalismista konkreettisia esimerkkejä. Kalevan vastauksessa todettiin, että “[meillä on] verkossa paljon juttuja, jossa yksi tai useampi data on tuotu/visualisoitu kartalle.”

Kalevan antamasta neljästä datajournalistisesta esimerkistä jokainen oli verkkoon tehty 'uutissovellus' tai 'lopputuote' (ks. mm. Hirst 2013, viitattu 28.3.2016), kuten vuorovaikutteinen

kartta (esim. 'Karttasovellus kertoo, kuinka huoltosuhde romahtaa suurkunnissakin'). Näistä jutuista kahden perässä mainittiin, että aiheesta kerrotaan lisää myös printtilehdessä.

Kalevassa datajournalismi nähtiin ennen kaikkea yhtenä työkaluna muiden joukossa. Tämän työkalun käyttöä perusteltiin siten, että "Tiettyjen asioiden jäsentäminen on tehokkainta datajournalistisin menetelmin." Lehden antamien esimerkkien valossa nämä "tiedyt asiat" tuntuisivat olevan suurehkojenkin data-aineistojen (esim. 19 000 parkkisakkoa) visualisoimista tavalla tai toisella. Parkkisakkojen lisäksi esimerkeissä oli käytetty väestödataa, väestön tulotietoja ja Ylen MOT-ohjelman kokoamaa yritystukiaineistoa. Kalevan jutuissa ikään kuin esiteltiin dataa visualisoinnin keinoin. Jutuissa ei ollut yhdistelty eri data-aineistoja, vaan kussakin oli käytetty yhtä datasettiä. Eli esimerkiksi mainitussa parkkisakkojutussa (Manninen 2014, viitattu 28.3.2016) oli käytetty ainoastaan parkkisakkodataa, joka oli visualisoitu klikattavalle kartalle.

Kalevan antamissa esimerkeissä ei juurikaan pyritty selittämään käsiteltyjen juttuaiheiden taustoja, eikä niissä kerrota, *miksi* jokin asia on niin kuin data kertoo sen olevan. Muun muassa Ylen atk-toimittajan Teemo Tebestin (haastattelu 18.2.2015) peräämä tutkivampi ote datajournalismiin siis puuttuu.

Kalevan esimerkissä "Rikkailla entistä paremmat tulot – katso kartasta missä rikkaat asuvat" tosin sanotaan näin:

Tänään julkaistusta Kalevan Sunnuntai-liitteestä selviää, miksi rikkaista on tullut entistä rikkaampia ja miten rikkaat rahaa kuluttavat. Varman toimitusjohtaja Matti Vuoria kertoo haastattelussa suhteestaan rahaan, kulutustottumuksistaan ja tulevaisuuden suunnitelmistaan ensi maaliskuussa tapahtuvan eläköitymisensä jälkeen. (Alasuutari & Leinonen 2013, viitattu 28.3.2016.)

Printtiin on siis tässä tapauksessa tehty verkon 'uutissovellusta' tukeva ja ainakin osittain selittävä juttu. Kalevan esimerkkijutut oli julkaistu vuosina 2012–2014.

Kalevan lisäksi datajournalismia kerrottiin tehtävän Karjalaisessa, Keskisuomalaisessa, Kouvolan Sanomissa, Kymen Sanomissa ja Etelä-Saimaassa. Yksikään näistä lehdistä ei kuitenkaan antanut esimerkkejä jutuistaan, joissa datajournalistisia työtapoja oli jollain tavalla käytetty. Miksi?

Ehkä datajournalistisia juttuja oli tehty vain printtiin, jolloin linkkejä näihin juttuihin ei voitu antaa. Tai ehkä vastaajat eivät syystä tai toisesta olleet sisällyttäneet vastauksiinsa linkkejä

datajournalistisiin juttuihin, vaikka niitä olisi lehtien verkkoversioista löytenytkin. Tai sitten vastaajat eivät puhuneet totta. Kaikki joka tapauksessa ilmoittivat, että datajournalismia tehdään sekä verkkoon että printtiin.

Kaikki vastaajat, joiden lehdissä tehtiin heidän mukaansa datajournalismia, eivät myöskään olleet tyytyväisiä tehdyn datajournalismin laatuun. Karjalaisesta valiteltiin kiirettä, joka on estänyt datajournalismin kehittämisen. Kouvolan Sanomissa, Kymen Sanomissa ja Etelä-Saimaassa syy oli "resurssipula". Jopa Kalevassa, jossa ei sinänsä nähty vikaa tehdyn datajournalismin laadussa, toivottiin saavan aikaiseksi "kunnianhimoisempia toteutuksia".

Kaikki vastanneet olivat sitä mieltä, että datajournalismia pitäisi tehdä nykyistä enemmän. Sen perään kuulutettiin myös Ilkassa ja Pohjalaisessa, joissa datajournalismia haluttiin alkaa tehdä tulevaisuudessa. Turun Sanomat oli kyselyyn vastanneista ainoa, jossa datajournalismille ei nähty tarvetta nyt eikä jatkossakaan.

Perustelut datan käyttämisestä osana journalismia vaihtelivat. Toisille (Kaleva; Kouvolan Sanomat, Kymen Sanomat & Etelä-Saimaa) se oli yksi työkalu lisää journalismin keinovalikoimaan. Toisaalta tämän työkalun käyttöä ei osattu juurikaan perustella eikä kertoa, miksi datajournalismia tarvitaan lisää. Keskisuomalaisesta sanottiin, että datajournalismi auttaa oman uutistyön tekemisessä, aineiston hyödyntämisessä parhaalla mahdollisella tavalla, lukijan palvelemisessa ja päätöksenteon tekemisessä läpinäkyväksi. Karjalaisen mukaan "datajournalismin keinoin on mahdollista yhdistellä tietoja siten, että syntyy uutta tietoa, mikä on tutkivan journalismin keskeinen tehtävä."

Näissä lehdissä ainakin tuntuisi olevan jonkinlainen käsitys datajournalismista ja sen tarjoamista mahdollisuuksista. Keskisuomalaisesta kerrottiin esimerkkeinä lehden tekemästä datajournalismista "verotietojen hyödyntäminen, interaktiivisten karttojen laatiminen tai esimerkiksi kaupungin julkaisemien ostolaskujen tutkiminen." Karjalaisesta todettiin, että datajournalismia tehdään "Monista sellaisista aihepiireistä, joista meidän teknisillä (=vaatimattomilla) keinoillamme dataa on saatavissa. Toimintaa rajoittaa tietoteknisten keinojen ja osaamisen vaje." Kumpikin lehti kertoi tekevänsä datajournalismia tai käyttävänsä datajournalistisia työmenetelmiä kerran kuussa. Konkreettiset juttuesimerkit kuitenkin puuttuivat.

Kaikissa kyselyyn vastanneissa lehdissä tiedettiin sellaisen termin kuin 'datajournalismi' olevan olemassa: ehkä siitä oli kuultu puhuttavan, luettu aiheesta jokunen artikkeli tai käyty jopa datajournalismiin liittyvässä koulutuksessa tai seminaarissa.

Datajournalismi oli määritelty kyselylomakkeessa melko lavasti. Määritelmässä kuitenkin korostettiin data-aineistojen toimituksellista analysointia ja käsittelyä osana journalismia: esimerkiksi valmiita tilastoja pelkästään *kuvailevat* jutut erotettiin datajournalismista selvästi.

Niin tai näin, neljä vastaajaa – jotka edustivat yhteensä kuutta maakuntalehteä – ilmoittivat tekevänsä datajournalismia vähintään kerran kuussa, jotkut jopa useammin. Tätä vasten tuntuu oudolta, että konkreettisia juttuesimerkkejä tuli vain yhdeltä lehdeltä. Myös perustelut datan käyttämiseksi jäivät pinnallisiksi. Jos esimerkiksi todettiin, että datan tai datajournalismin avulla voidaan tehdä päätöksentekoa läpinäkyvämmäksi, ei kerrottu *kuinka* se onnistuu. Jos sanottiin, että lehdessä on muun muassa hyödynnetty verotietoja tai tutkittu kaupungin ostolaskuja, ei kerrottu millä tavalla niitä verotietoja on hyödynnetty tai millä tavalla kaupungin ostolaskuja on tutkittu. Ei ole tarkoitus väittää, ettei näin *olisi* toimittu, mutta linkkejä näihin juttuihin tai perusteellisempia selvityksiä kyseisistä työmenetelmistä jäi joka tapauksessa kaipaamaan.

Selvää oli, että kaikissa vastanneissa lehdissä paitsi Turun Sanomissa haluttiin tehdä enemmän datajournalismia kuin tällä haavaa tehdään. Varsinaisia syvällisempiä perusteluja datajournalismin lisäämiselle ei kuitenkaan ollut. Yhdestäkään vastanneesta lehdestä ei esimerkiksi kerrottu, että datajournalismi nähtäisiin suoraan keinona hankkia lisää lukijoita tai selviytyä niin sanotussa median murroksessa.

Ilkasta ja Pohjalaisesta myös kerrottiin, että kaikki toimittajat eivät tiedä, mitä datajournalismilla tarkoitetaan ja että se on yksi syy, miksi näissä lehdissä ei käytetä datajournalistisia työmenetelmiä: ” Osalle talon väkeä sana datajournalismi ei ole aivan selkeä, eikä varsinkaan tiedetä, miten sitä tehdään. Eli kun työkalut ovat olemassa, pitää vielä kouluttaa toimitus ymmärtämään, miten niitä hyödynnetään ja mitä hyötyä niistä on toimittajalle itselleenkin.”

6 LOPUKSI

Tässä tutkielmassa on yritetty löytää perusteluja, miksi toimitusten tulisi – Suomessakin – kiinnostua datan käyttämisestä journalismin tekemisessä. Ääneen ovat päässeet tutkijat ja toimittajat niin Suomesta kuin ulkomailta. Heidän näkemyksensä on tuotu esiin lähinnä aihetta käsittelevistä kirjallisista lähteistä, joita on tuettu tutkielmaa varten tehdyillä haastatteluilla.

Tutkielman yksi kantavista ajatuksista on, että vaikka datajournalismi on terminä kohtalaisen tuore tapaus, sen juuret ulottuvat pitkälle journalismin historiaan. Datan käyttäminen ja sen analysoiminen tietokoneella eivät ole alalla uusia ilmiöitä, vaikka siltä saattaa helposti tuntua viime vuosina datajournalismin ympärille puhjenneen keskustelun vuoksi. Yksi tämän tutkielman tavoitteista on hälventää käsitystä datajournalismista yksinomaan 2010-luvun keksintönä.

Sen sijaan, että datan käyttäminen nähtäisiin jotenkin erityisenä journalismin osa-alueena, olisi hedelmällisempää mieltää se aivan yhtä olennaiseksi osaksi journalismia kuin haastattelujen tekeminen puhelimitse tai muistiinpanojen tekeminen kynällä. Kaiketi kukaan ei puhu puhelinjournalismista eikä kynäjournalismista – olisiko siis viimeistään nyt aika lopettaa puhuminen datajournalismista?

Ajatus ei ole ihan uusi eikä myöskään allekirjoittaneen oma. Tarkkuusjournalismin pioneeri Philip Meyer tuskaili vastaavan ongelman kanssa vuonna 1999; tuolloin vain puhuttiin tietokoneavusteisesta journalismista (CAR). Meyer totesi, että journalistit olivat ainoa ammattikunta, jonka piti vielä 2000-luvun kynnykselläkin mainita erikseen käyttävänsä tietokoneita työnsä tekemiseen. Meyeristä tämä oli noloa. Toimittaja Rich Gordon ehdottikin tuolloin, että tietokoneavusteisen journalismin sijasta tulisi palata käyttämään vanhempaa termiä 'tarkkuusjournalismi'. Hänen mukaansa CARissa oli nimenomaan kyse tarkkuudesta: siirtymisestä lausuntoja referoivasta journalismista kohti asioiden kiistatonta tilaa eli totuutta. (Gordon 1999, 22, viitattu 28.3.2016.) Ehkä kaikkein ihanteellisinta kuitenkin olisi, jos journalismin tarkkuutta ja faktapohjaisuutta tai näihin pyrkiviä työmenetelmiä ei tarvitsisi erikseen mainita. Tällainen erottelu tarkoittaisi, että 'tarkan' journalismin lisäksi olisi olemassa journalismia, joka on epätarkkaa ja faktapohjaltaan heikompaan, ja jonka tekemiseen siis pätsivät jollakin tavalla löysemmät säännöt, työtavat ja eettiset ohjeet kuin 'korkeamman tason' journalismiin. Ihan kaiken journalismin tulisi olla mahdollisimman totuudenmukaista ja koviin faktoihin perustuvaa – ja

jokaisen toimittajan tulisi hallita mahdollisimman paljon tehokkaita työkaluja totuuden selvittämiseksi.

Meyerin ohella muutkin tietokoneavusteisen journalismin ammattilaiset Yhdysvalloista painottivat (viimeistään) viime vuosikymmenen lopussa, kuinka datan hankkimisen ja sen analysoimisen tekniikat tulisi ottaa yleiseen käyttöön toimituksissa. Esimerkiksi Stephen Doig (1999, viitattu 28.3.2016) totesi, että taulukkolaskentaohjelman hallitseminen tulisi olla toimittajille yhtä olennainen pätevyysvaatimus kuin kirjoitustaito.

Kotikutoista näkökulmaa tutkielmaan haettiin kyselytutkimuksella, jolla pyrittiin selvittämään datajournalististen työmenetelmien yleisyyttä Suomen maakuntalehdistössä. Tutkimuksen perusteella monesta maakuntalehdestä ainakin väitetään, että niissä tehdään datajournalismia vähintään kerran kuussa. Käsien kosketeltavien esimerkkijuttujen puute pistää kuitenkin epäilemään väitteen todenperäisyyttä. Syvällisempi perehtyminen asiaan voisikin olla jatkossa paikallaan. Kyselytutkimusta voisi esimerkiksi laajentaa koskemaan kaikkia maakuntalehtien toimittajia, siinä missä se osana tätä tutkielmaa lähetettiin vain yhdelle esimiesportaan työntekijälle per lehti. Lisäksi maakuntalehtien juttuja tulisi analysoida itsenäisesti siltä kantilta, minkälaisia lähteitä jutuissa tietyllä aikajänteellä – vaikka kuukauden ajan – käytetään.

Kyselytutkimuksen pohjalta on selvää, että datajournalismia haluttaisiin tehdä maakuntalehdissä nykyistä enemmän. Esteenä dataan pohjautuvien työmenetelmien käytölle nähtiin pula sekä ajasta että osaamisesta. Ja vaikka osaamista esimerkiksi Excelin käyttämisestä olisi, tulisi toimittajien vielä ymmärtää, miten tätä taitoa voisi hyödyntää paremman journalismin tekemiseksi.

Gordon (1999, viitattu 28.3.2016) toteaaakin, että journalismi ei muutu miksiäkään sillä, että toimittajille aletaan opettaa Exceliä tai Accessia. Toimittajien on itse ruvettava ajattelemaan kokonaan eri tavalla, lähestymään informaatiota analyttisemmin ja uskomaan, että datasta pystyy löytämään sellaisen totuuden, jota ei voisi koskaan saada selville ihmisiä haastatteleamalla. Ongelma vain on siinä, että monet hakeutuvat tälle alalle siksi, että he tykkäävät kirjoittaa ja jutella ihmisille – eivät pyöritellä numeroita ja etsiä vastauksia tietokannoista. (Ibid.)

Pallo onkin heitettävä osaltaan toimittajakoulutuksen suuntaan, jotta tällainen ajattelutavan muutos olisi journalistien ammattikunnassa mahdollinen. Ilman systemaattista koulutusta datajournalististen menetelmien opettelu jää opiskelijoiden oman kiinnostuksen ja omalla ajalla

aiheeseen perehtymisen varaan. Tässäkin olisi paikka omalle tutkimukselleen: mitä toimittajakoulutuksessa tänä päivänä Suomessa – ja miksei maailmalla – painotetaan? Kuinka paljon nimenomaan datan journalistista käyttämistä opetetaan?

Oma roolinsa on myös medioiden, ja ehkä juuri erityisesti sanomalehtien toimituksilla. Yhden datajournalistisista työmenetelmistä kiinnostuneen toimittajan voi olla vaikea viedä ajatustaan pidemmälle työyhteisössä, jossa on totuttu tekemään asioita niin sanotusti perinteisellä tavalla. Heikko taloustilanne tuo omat vaikeutensa kunnianhimoisemman journalismin tekemiseen ylipäätään: kun aikaa ja muita resursseja on vähemmän käytössä, tehdään juttuja pintapuolisemmalla otteella. Myös aiheiltaan kevyet jutut, joita on helppo ja nopea toteuttaa, lisääntyvät tällöin herkästi. Journalismin rooli vallan vahtikoirana heikentyy.

Tästä tutkielmasta voi olla osaltaan hyötyä saada toimittajia, alan opiskelijoita ja miksei myös opettajia kiinnostumaan datan käyttämisestä journalismissa. Näiden työmenetelmien pitkien perinteiden hahmottaminen auttaa ymmärtämään, että kyseessä ei ole nyt vain joku uusi datajournalismin nimellä kulkeva muoti-ilmiö, vaan aivan olennainen työkalu, jolla on mahdollista tehdä parempaa, kunnianhimoisempaa ja merkittävämpää journalismia helpommin kuin koskaan.

LÄHTEET

6Aika. 2016. 6Aika - Avoimia ja älykkäitä palveluja. Hakupäivä 15.2.2016 <http://6aika.fi/6aika-avoimia-ja-alykkaita-palveluja/>.

Adrian, P. 1999. Put Your Career in the Fast Lane with CAR. Hakupäivä 22.3.2016 <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf>.

Aitamurto, T., Sirkkunen, E. & Lehtonen, P. 2011. Trends In Data Journalism. Hakupäivä 17.3.2016 http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf.

Alasuutari, P. & Leinonen, A. 2013. Rikkailla entistä paremmat tulot - katso kartasta missä rikkaat asuvat. Hakupäivä 28.3.2016 <http://www.kaleva.fi/uutiset/kotimaa/rikkailla-entista-paremmat-tulot-katso-kartasta-missa-rikkaat-asuvat/643931/>.

American Press Institute. 2016. What is journalism? Hakupäivä 18.2.2016 <https://www.americanpressinstitute.org/journalism-essentials/what-is-journalism/>.

Bounegru, L. 2012. Data Journalism in Perspective. Teoksessa J. Gray, L. Bounegru & L. Chambers (toim.) The Data Journalism Handbook. How Journalists Can Use Data to Improve the News. California, USA: O'Reilly Media, 17-22.

Bradshaw, P. 2015. FAQ: 24 questions about data journalism. Hakupäivä 25.2.2016 <http://onlinejournalismblog.com/2015/09/12/faq-24-questions-about-data-journalism/#more-20796>.

Bradshaw, P. 2015. Toimittaja. Sähköpostihaastattelu 9.11.2015. Tekijän hallussa.

Bradshaw, P. 2014. Is there a 'canon' of data journalism? Comment call! Hakupäivä 17.3.2016 <http://onlinejournalismblog.com/2014/05/20/is-there-a-canon-of-data-journalism/>.

Bradshaw, P. 2012. What Is Data Journalism? Teoksessa J. Gray, L. Bounegru & L. Chambers (toim.) The Data Journalism Handbook. How Journalists Can Use Data to Improve the News. California, USA: O'Reilly Media, 2.

Coddington, M. 2014. Clarifying Journalism's Quantitative Turn. Hakupäivä 15.3.2016 <http://www.tandfonline.com/doi/pdf/10.1080/21670811.2014.976400>.

Cox, M. 2000. The Development of Computer-Assisted Reporting. Hakupäivä 3.3.2016 <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.631.6220&rep=rep1&type=pdf>.

Dagiral, E. & Parasie, S. 2012. Data-driven journalism and the public good - "Computer-assisted-reporters" and "programmer-journalists" in Chicago. Hakupäivä 7.3.2016 https://sypar.files.wordpress.com/2011/05/parasie-dagiral-2012_data-driven-journalism_nms_preprint.pdf.

DeBarros, A. 2010. Data Journalism and the Big Picture. <http://www.anthonydebarros.com/2010/11/26/data-journalism-the-big-picture/>.

Doig, S. K. 1999. It's Time to Drive CAR Concepts Home. Hakupäivä 28.3.2016 <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf>.

Euroopan komissio. 2011. Digitaalistrategia: viranomaistiedot laajempaan käyttöön. Lehdistöiedote. Hakupäivä 13.2.2016 http://europa.eu/rapid/press-release_IP-11-1524_fi.htm.

Friendly, M. 2006. A Brief History of Data Visualization. Hakupäivä 1.3.2016 <http://www.datavis.ca/papers/hbook.pdf>.

Global Editors Network. Winners of the DJA 2015. Hakupäivä 19.3.2016 <http://www.globaleditorsnetwork.org/programmes/data-journalism-awards/dja-2015/>.

Gordon, R. 1999. Reflections of a One-Eyed Man. Hakupäivä 10.3.2016 <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf>.

Haasio, A. & Savolainen, R.,. 2004. Tiedonhankintatutkimuksen perusteet. Helsinki: BTJ Kirjastopalvelu.

Harju, A. & Sirkkunen, E. 2015. Mitä datajournalismi on? Hakupäivä 23.2.2016 <https://blogs.uta.fi/datajournalismi/esimerkkisivu/>.

Hemánus, P. 1995. Opettaako journalistia Siperia vai tieto-oppi? Teoksessa E. Sana (toim.) Tieto-opista mediapeliin - journalismin tutkimuksen näkökulmia. Helsinki: WSOY, 27-46.

Hemánus, P. 1997. Journalismi, ikkuna todellisuuteen?: Jäähyväisluento ja bibliografia. Tampereen yliopiston tiedotusopin laitoksen julkaisuja 1997:C 23.

Hemanus, P. & Tervonen, I. 1980. Objektiiivinen joukkotiedotus. [Hki]: Otava.

Henn, S. 2012. The Night A Computer Predicted The Next President. Hakupäivä 26.11.2014 <http://www.npr.org/sections/alltechconsidered/2012/10/31/163951263/the-night-a-computer-predicted-the-next-president>.

Hirst, T. 2013. Several Takes on Defining Data Journalism. Hakupäivä 25.2.2016 <http://schoolofdata.org/2013/06/11/several-takes-on-defining-data-journalism/>.

Holovaty, A. 2009. The definitive, two-part answer to "is data journalism?". Hakupäivä 24.2.2016 <http://www.holovaty.com/writing/data-is-journalism/>.

Holovaty, A. 2006. A fundamental way newspaper sites need to change. Hakupäivä 16.3.2016 <http://www.holovaty.com/writing/fundamental-change/>.

Houston, B. 1999. Changes in Attitudes, Changes in Latitudes. Hakupäivä 10.3.2016 <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf>.

Houston, B. 2015. Computer-assisted reporting : a practical guide. 4. New York: Routledge, Taylor & Francis Group.

Howard, A. 2014. The Art and Science of Data-Driven Journalism. <http://towcenter.org/research/the-art-and-science-of-data-driven-journalism>.

Howard, A. 2010. The growing importance of data journalism. <http://radar.oreilly.com/2010/12/data-journalism.html>.

IBM. 2016. What is big data? Hakupäivä 20.3.2016 <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>.

Investigative Reporters and Editors. 2016. About NICAR. Hakupäivä 8.3.2016 <http://ire.org/nicar/about/>.

James, J. 2015. Data Never Sleeps 3.0. Hakupäivä 20.3.2016 <https://www.domo.com/blog/2015/08/data-never-sleeps-3-0/>.

Julkisen hallinnon tietohallinnon neuvottelukunta. 2014. JHS 189 Avoimen tietoaiteiston käyttöluupa. Hakupäivä 9.2.2016 <http://docs.jhs-suositukset.fi/jhs-suositukset/JHS189/JHS189.html>.

Julkisen sanan neuvosto. 2013. Journalistin ohjeet ja liite. Hakupäivä 17.2.2016 http://www.jsn.fi/journalistin_ohjeet/.

Kauhanen-Simanainen, A. & Suurhasko, M. 2015. Avoimesta datasta innovatiiviseen tiedon hyödyntämiseen: Avoimen tiedon ohjelman 2013-2015 loppuraportti. Valtiovarainministeriö. Valtiovarainministeriön julkaisuja 31/2015. Hakupäivä 13.2.2016 <http://vm.fi/documents/10623/1107406/Avoimen+tiedon+ohjelman+loppuraportti/8eaaee68-6f3b-4a48-8b57-c5866315bf13?version=1.0>.

Kayser-Bril, N. 2012. An Answer to Data-Driven PR. Teoksessa J. Gray, L. Bounegru & L. Chambers (toim.) The Data Journalism Handbook. How Journalists Can Use Data to Improve the News. California, USA: O'Reilly Media, 7-8.

Klein, S. 2015. Antebellum Data Journalism: Or, How Big Data Busted Abe Lincoln. Hakupäivä 1.3.2016 <https://www.propublica.org/nerds/item/antebellum-data-journalism-busted-abe-lincoln>.

Korkman, P. & Yrjönsuuri, M. 1998. Filosofian historian kehityslinjoja. Helsinki: Gaudeamus.

Kotimaisten kielten keskus. 2016. 'Data'. Hakupäivä 8.2.2016
<http://www.kielitoimistonsanakirja.fi/>.

Kotimaisten kielten keskus. 2016. 'Tieto'. Hakupäivä 8.2.2016
<http://www.kielitoimistonsanakirja.fi/>.

Kunelius, R. 2009. Viestinnän vallassa : johdatusta joukkoviestinnän kysymyksiin. Helsinki: WSOY.

Kuutti, H. & Uskali, T. 2015. Models and Streams of Data Journalism. The Journal of Media Innovations 2.1, 77-88. Hakupäivä 9.3.2016
<https://journals.uio.no/index.php/TJMI/article/viewFile/882/1160>.

Kuutti, H. 2015. Todenmukainen journalismi. [Jyväskylä]: Jyväskylän yliopisto, viestintätieteiden laitos.

Kuutti, H. 2012. Mediasanasto. Jyväskylä: MediaDoc.

Liski, J. 2015. Toimittaja. Puhelinhaastattelu 17.2.2015. Tekijän hallussa.

Lorenz, M. 2010. Status and Outlook for data-driven journalism. Hakupäivä 29.2.2016
http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf.

Lorenz, M. 2012. Why Journalists Should Use Data. Teoksessa J. Gray, L. Bounegru & L. Chambers (toim.) The Data Journalism Handbook. How Journalists Can Use Data to Improve the News. California, USA: O'Reilly Media, 3-5.

Manninen, J. 2014. Näissä paikoissa parkkisakkoja napsahtaa eniten Oulussa - katso sakkorysät kartalta. Hakupäivä 28.3.2016 <http://www.kaleva.fi/uutiset/kotimaa/naissa-paikoissa-parkkisakkoja-napsahtaa-eniten-oulussa-katso-sakkorysat-kartalta/677942/>.

Matsunami, I. 2012. Providing Independent Interpretations of Official Information. Teoksessa J. Gray, L. Bounegru & L. Chambers (toim.) The Data Journalism Handbook. How Journalists Can Use Data to Improve the News. California, USA: O'Reilly Media, 8.

Merriam-Webster. 2015. 'Data'. Hakupäivä 3.2.2016 <http://www.merriam-webster.com/dictionary/data>.

Meyer, P. 2012. Filtering the Flow of Data. Teoksessa J. Gray, L. Bounegru & L. Chambers (toim.) The Data Journalism Handbook. How Journalists Can Use Data to Improve the News. California, USA: O'Reilly Media, 6.

Meyer, P. 1999. The Future of CAR: Declare Victory and Get out! Hakupäivä 15.3.2016 <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf>;

Meyer, P. cop. 1979. Precision journalism : a reporter's introduction to social science methods. 2. ed. Bloomington: Indiana University Press.

Minkoff, M. 2010. Bringing data journalism into curricula. <http://michelleminkoff.com/2010/03/24/bringing-data-journalism-into-curricula/>.

mySociety. 2016. About. Hakupäivä 17.3.2016 <https://www.mysociety.org/about/>.

Nieminen, A. 2015. Uuden ajan journalismia - datajournalismin määritelmä, merkitys ja tila Suomessa. Tiedotusopin pro gradu -tutkielma. 2015. Tampereen yliopisto. Hakupäivä 23.2.2016 <http://tampub.uta.fi/bitstream/handle/10024/97685/GRADU-1435923036.pdf?sequence=1>.

Niiniluoto, I. 1996. Informaatio, tieto ja yhteiskunta : filosofinen käsiteanalyysi. 5. täyd. p. Helsinki: Edita.

O'Murchu, C. 2012. An Essential Part of the Journalists' Toolkit. Teoksessa J. Gray, L. Bounegru & L. Chambers (toim.) The Data Journalism Handbook. How Journalists Can Use Data to Improve the News. California, USA: O'Reilly Media, 9-10.

Open Knowledge Finland. 2016. Visiot ja arvot. Hakupäivä 11.2.2016
<http://fi.okfn.org/about/visiojaarvot/>.

Open Knowledge. 2015. About Open Data Index 2015: Civil Society Audit of the state of Open Data. Hakupäivä 12.2.2016 <http://index.okfn.org/about/>.

Open Knowledge. 2015. Global Open Data Index. Hakupäivä 12.2.2016
<http://index.okfn.org/place/>.

Open Knowledge. 2016. Open Definition. Hakupäivä 11.2.2016
<http://opendefinition.org/od/2.1/en/>.

Oras, T. 2015. Työehtoasiamies, Suomen Journalistiliitto. Sähköpostihaastattelu 4.9.2015. Tekijän hallussa.

Poikola, A., Kola, P. & Hintikka, K. A. 2010. Julkinen data (CC) : johdatus tietovarantojen avaamiseen. Helsinki: Liikenne ja viestintäministeriö.

Radcliffe, D. 2016. Understanding the Past, Present and Future of Data Journalism. Hakupäivä 1.3.2016
<http://mediashift.org/2016/02/understanding-the-past-present-and-future-of-data-journalism/>.

Rogers, S. 2013. Facts are Sacred. Englanti: Faber & Faber.

Rogers, S. 2011. Wikileaks data journalism: how we handled the data. Hakupäivä 17.3.2016
<http://www.theguardian.com/news/datablog/2011/jan/31/wikileaks-data-journalism>.

Sanastokeskus TSK ry. 2016. 'Data'. Hakupäivä 9.2.2016 <http://www.tsk.fi/cgi-bin/netmot.exe?UI=figr&height=152&qfind=data>.

Sanomalehtien Liitto. 2016. Levikki. Hakupäivä 25.3.2016
<http://www.sanomalehdet.fi/sanomalehtitieto/levikki>.

Stith, P. 2005. A Guide to Computer Assisted Reporting. <http://www.poynter.org/2005/a-guide-to-computer-assisted-reporting/69334/>.

Suomen Ammattiliittojen Keskusjärjestö. 2016. Irtisanomisia, lomautuksia ja yt-neuvotteluja 2015. Hakupäivä 25.3.2016 <https://www.sak.fi/aineistot/tilastot/yt-ja-irtisanomistilastot/irtisanomisia-lomautuksia-ja-yt-neuvotteluja-2015-2015-10-16>.

Tampereen yliopisto. 2003. Otos ja otantamenetelmät. Hakupäivä 28.3.2016 <http://www.fsd.uta.fi/menetelmaopetus/otos/otantamenetelmat.html>.

Taylor, M. 2009. How Computer-Assisted Reporters Evolved into Programmer/Journalists. <http://mediashift.org/2009/08/how-computer-assisted-reporters-evolved-into-programmerjournalists219/>.

Tebest, T. 2016. Teemo Tebest. Hakupäivä 17.3.2016 <http://teelmo.info/>.

Tebest, T. 2015. Toimittaja. Puhelinhaastattelu 18.2.2015. Tekijän hallussa.

Tebest, T. 2012. Näkökulma: Datajournalismi ennen, nyt ja tulevaisuudessa. Hakupäivä 25.11.2014 <http://datajournalismi.blogspot.fi/2012/06/nakokulma-datajournalismi-ennen-nyt-ja.html>.

The Pulitzer Prizes. 2016. The 1993 Pulitzer Prize Winner in Public Service. Hakupäivä 24.3.2016 <http://www.pulitzer.org/winners/6763>.

Thibodeaux, T. 2011. 5 tips for getting started in data journalism. Hakupäivä 24.2.2016 <http://www.poynter.org/2011/5-tips-for-getting-started-in-data-journalism/147734/>.

Toikkanen, I. 2014. Pinnallisista klikkikartoista yhteiskunnalliseen merkittävyyteen - Datajournalismin työprosessi Helsingin Sanomissa ja Ylessä. Hakupäivä 29.2.2016 <https://jyx.jyu.fi/dspace/bitstream/handle/123456789/43189/URN%3aNBN%3afi%3ajyu-201404081484.pdf?sequence=1>.

Valpola, V. & Koukkunen, K. 2000. Suuri sivistyssanakirja. Helsinki: WSOY.

Valtion tieto- ja viestintäteknikkakeskus Valtori. 2016. Avoindata.fi kehityksen vuosi 2015. Hakupäivä 13.2.2016 <https://www.avoindata.fi/fi/content/avoindatafin-kehityksen-vuosi-2015>.

Valtioneuvosto. 2011. Pääministeri Jyrki Kataisen hallituksen ohjelma. Hakupäivä 13.2.2016 <http://valtioneuvosto.fi/documents/10184/147449/Kataisen+hallituksen+ohjelma/81f1c20f-e353-47a8-8b8f-52ead83e5f1a>.

Valtioneuvosto. 2015. Ratkaisujen Suomi. Juha Sipilän hallituksen strateginen ohjelma. Hakupäivä 13.2.2016 http://valtioneuvosto.fi/documents/10184/1427398/Ratkaisujen+Suomi_FI_YHDISTETTY_netti.pdf/801f523e-5dfb-45a4-8b4b-5b5491d6cc82.

Vehkoo, J. 2011. Painokoneet seis! : kertomuksia uuden journalismin ajasta. Helsinki: Teos.

Vermanen, J. 2012. Updating Your Skills Set. Teoksessa J. Gray, L. Bounegru & L. Chambers (toim.) The Data Journalism Handbook. How Journalists Can Use Data to Improve the News. California, USA: O'Reilly Media, 7.

Web Foundation. 2015. Open Data Barometer. Hakupäivä 12.2.2016 <http://barometer.opendataresearch.org/>.

Wikipedia. 2014. Hurrikaani Andrew. Hakupäivä 24.3.2016 https://fi.wikipedia.org/wiki/Hurrikaani_Andrew.

Yleisradio. 2016. YLE Plus. Hakupäivä 17.3.2016 <http://yle.fi/uutiset/plus/>.

Younge, G. 2011. The Detroit riots of 1967 hold some lessons for the UK. Hakupäivä 26.11.2014 <http://www.theguardian.com/uk/2011/sep/05/detroit-riots-1967-lessons-uk>.

Tämä kyselytutkimus on osa Oulun ammattikorkeakoulun viestinnän tutkimusohjelmassa lukuvuotena 2015-2016 tehtävää opinnäytetyötä. Kyselyyn voi halutessaan vastata anonyymisti, mutta lehden nimi tulisi kuitenkin mainita. Kyselyn tekijä on viidennen vuoden journalismin opiskelija Eetu Halonen (040 83 73 220 / roxutee@gmail.com).

Vastaajan tiedot

Kuvaus (valinnainen)

Vastaajan nimi

Lyhyt vastausteksti

Titteli / toimi

- Päätoimittaja
- Vastaava päätoimittaja
- Toimituspäällikkö
- Uutispäällikkö
- Tuottaja
- Muu...

Edustamasi lehti

Mahdollisimman tarkasti, esim. pelkkä 'Alma Media' ei ole tarpeeksi tarkka vastaus. Merkitkää kaikki edustamanne lehdet erikseen.

Datajournalismi

Termi on määritelty tässä opinnäytetyössä muun muassa seuraavasti:

"Datajournalismi on työprosessi, jonka aikana hankitaan, muokataan ja analysoidaan dataa. Tämän prosessin lopputuloksena julkaistaan datajournalistinen juttu ja mielellään myös jutussa käytetty data. Kuten jo datajournalismin nimestä käy ilmi, on tässä prosessissa olennaisesti mukana myös perinteisempi käsitys journalistin tekemisestä. Datajournalismissa yhdistyvätkin journalistin työtavat datan käsittelyyn, ohjelmointiin ja visualisointiin."

Lisäksi tässä opinnäytetyössä datajournalismiksi on laskettu myös 'datajournalistiset työtavat' osana juttujen tekemistä, vaikkei juttua/juttuprosessia varsinaisesti datajournalismiksi muuten kutsuttaisikaan, eikä jutusta esimerkiksi tehtäisi verkkolehden mitään uutissovellusta tai muuta visuaalista, verkonomaista toteutusta.

Tällaisia 'datajournalistisia työtapoja' ovat muun muassa tietopyyntöjen tekeminen julkishallinnolle, datan hankkiminen jollain muulla tavalla esimerkiksi toimituksen omista tietokannoista tai vaikkapa 'raaputtamalla' dataa verkosta, datan käsitteleminen ja sen käyttäminen jutun ideana, pohjana, alkupisteenä tai jollakin tavalla osana juttua.

On huomionarvoista, että monet (data-)journalistit eivät kuitenkaan pidä esimerkiksi jotakin tilastoaineistoa referoivaa juttua vielä datajournalismina. Usein tällaisia mediaa kiinnostavia tilastoja ovat erilaiset barometrit, yritysten osavuosiselvitykset tai Tilastokeskuksen julkaisemat aineistot.

Ajalla vallitseva käsitys siis on, että jostakin tällaisesta 'datasta' uutisoiminen referoivaan tyyliin, jossa datalla/datalle ei varsinaisesti ole tehty mitään, ei vielä ole datajournalismia. Sen sijaan vaikkapa em. tilastoaineistojen hyödyntäminen jollakin tavalla ja niissä esiintyvän datan analysoiminen, muokkaaminen tai jokin muu käyttäminen osana juttuprosessia on datajournalismia.

Kts. lisää: <https://blogs.uta.fi/datajournalismi/>

Tehdäänkö mediasianne yllä olevaan kuvailuun sopivaa datajournalismia?

Kyllä

Ei

Kyllä, toimituksessamme tehdään datajournalismia

Kuvaus (valinnainen)

Kuvaile, minkälaista datajournalismia toimituksessanne tehdään?

Esim. tehdäänkö datajournalismia vain tietyistä aihepiireistä? Antakaa konkreettisia esimerkkejä ja linkkejä juttuihin jos se on mahdollista.

Pitkä vastausteksti

Mihin kanaviin datajournalismia lehdessänne tehdään?

- Printtiin
- Verkkoon
- Sekä printtiin että verkkoon
- Muu...

Arvioi, kuinka tärkeä osa datajournalismi on toimituksenne työtä?

| | 1 | 2 | 3 | 4 | 5 | |
|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|---------------------|
| Ei lainkaan merkittävä | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Erittäin merkittävä |

Arvioi, kuinka usein datajournalismia tehdään toimituksessanne tai kuinka usein datajournalistisia työtapoja käytetään?

- Päivittäin
- Kerran viikossa
- Kerran kuussa
- Kerran puolessa vuodessa
- Kerran vuodessa
- Harvemmin

Onko toimituksessane erikseen datajournalismiin keskittyviä toimittajia / datatoimitusta?

- Kyllä
- Ei

Kertokaa, miksi toimituksenne tekee datajournalismia / käyttää datajournalistisia työtapoja?

Pitkä vastausteksti

Oletteko tyytyväinen toimitukseenne tekemän datajournalismin laatuun?

Voitte halutessanne perustella vastausta alla olevaan tekstikenttään.

- Kyllä
- En
- En osaa sanoa

Kysymys

Pitkä vastausteksti

Oletteko tyytyväinen toimitukseenne tekemän datajournalismin määrään?

Voitte halutessanne perustella vastausta alla olevaan tekstikenttään.

- Kyllä
- En
- En osaa sanoa

Kysymys

Pitkä vastausteksti

Ei, toimituksessamme ei tehdä datajournalismia

Kuvaus (valinnainen)

Kertokaa, miksi toimituksessanne ei tehdä datajournalismia?

Pitkä vastausteksti

Onko toimituksenne suunnitelmissa alkaa tehdä datajournalismia tulevaisuudessa?

Voitte halutessanne kertoa lisää alla olevassa tekstikentässä.

- Kyllä
- Ei

Kysymys

Pitkä vastausteksti

Oliko datajournalismi teille entuudestaan tuntematon käsite ennen tätä kyselyä?

- Kyllä
- Ei