

Pinnallisista klikkikartoista yhteiskunnalliseen merkittävyyteen

Datajournalismin työprosessi Helsingin Sanomissa ja Ylessä

Ilkka Toikkanen

Journalistiikan pro gradu -tutkielma

Kevät 2014

Viestintätieteiden laitos

Jyväskylän yliopisto

JYVÄSKYLÄN YLIOPISTO

| | |
|---|--|
| Tiedekunta – Faculty HUMANISTINEN | Laitos – Department Viestintätieteiden laitos |
| Tekijä – Author Ilkka Toikkanen | |
| Työn nimi – Title Pinnallisista klikkikartoista yhteiskunnalliseen merkittävyyteen Datajournalismin työprosessi Helsingin Sanomissa ja Ylessä | |
| Oppiaine – Subject Journalistiikka | Työn laji – Level Pro gradu -tutkielma |
| Aika – Month and year Helmikuu 2014 | Sivumäärä – Number of pages 106 + liite |
| <p>Tiivistelmä – Abstract</p> <p>Selvitän pro gradu -työssäni, millainen on datajournalismin työprosessi Helsingin Sanomien, Ylen uutis- ja ajankohtaisohjelmien, sekä Svenska Ylen toimituksissa. Kuvaan toimitusten datajournalistisen prosessin tyypillistä kulkua ja haasteita, joita sen tekijät kohtaavat. Tämän lisäksi luotaan datajournalismin tulevaisuudennäkymiä.</p> <p>Helsingin Sanomien datadeski, Ylen Plus-deski ja Svenska Ylen Grävnavet valikoituivat tutkimuskohteiksi, koska ne ovat Suomen ainoat säännöllisesti datajournalismia tuottavat ryhmät. Tutkimusta varten keräsin tietoa teemahaastatteluilla ja toimituksissa tekemälläni havainnoinnilla.</p> <p>Tutkimuksessa selvisi, että datajournalistien suhtautumista toimittajan työhön ohjaa datan avoimuuden kannattajille tyypillinen ajatus avoimuuden ja läpinäkyvyyden tärkeydestä. Datajournalistit nauttivat toimituksissaan arvostusta alansa osajina, mutta datajournalistisen työprosessin vaatimusten ja mahdollisuuksien heikko tuntemus vaikeuttaa heidän osaamisensa hyödyntämistä. Suomalaistoimitusten datajournalismi on monella tapaa vasta kokeiluvaiheessa: työprosessin hallinta ja tekninen osaaminen, sekä yhteistyö muun toimituksen kanssa kehittyvät jatkuvasti. Datajournalistien toiveena on tehdä tulevaisuudessa nykyistä suurempia ja merkittävämpiä juttuja ja siirtää helppoja ja rutiininomaisia työtehtäviä muiden toimittajien itse tehtäväksi. Toimittajien teknisen osaamisen kehitys viittaakin siihen, että tämä toive voi toteutua.</p> | |
| Asiasanat – Keywords Datajournalismi, verkkojournalismi, visualisointi, data, Yle, Helsingin Sanomat | |
| Säilytyspaikka – Depository Jyväskylän yliopisto | |

Sisällys

| | |
|---|----|
| 1 JOHDANTO | 1 |
| 2 KÄSITTEET | 5 |
| 2.1 Datajournalismi | 5 |
| 2.2 Avoin data – datajournalismin polttoaine | 9 |
| 3 DATAJOURNALISMIN HISTORIA | 13 |
| 3.1 Maailmalla... | 13 |
| 3.2 ... ja Suomessa | 17 |
| 4 DATAJOURNALISMIN TYÖPROSESSI | 20 |
| 4.1 Datan hankinta | 25 |
| 4.2 Datan käsittely | 26 |
| 4.3 Lopputuotteen rakentaminen | 28 |
| 5 DATAJOURNALISMI JA AVOIMUUS | 31 |
| 6 TUTKIMUKSEN KUVAUS JA METODIT | 34 |
| 6.1 Tutkimuskysymykset ja tutkimuksen tarkoitus | 34 |
| 6.2 Tutkimusprosessi ja sen lähtökohdat | 35 |
| 6.3 Tieteenfilosofinen asemointi | 36 |
| 6.4 Metodivalinnan perustelu | 38 |
| 6.4.1 Haastattelu | 40 |
| 6.4.2 Havainnointi | 41 |
| 6.5 Tutkittavien valinta ja aineistonkeruun järjestelyt | 42 |
| 6.6 Aineisto | 44 |
| 6.7 Aineiston analyysi | 45 |

| | |
|---|----|
| 7 DATADESKIT JA TYÖPROSESSI | 48 |
| 7.1 Helsingin Sanomien datadeski | 48 |
| 7.1.1 Datan hankinta | 52 |
| 7.1.2 Datan puhdistus ja analyysi | 53 |
| 7.1.3 Visualisointi tai muu lopputuote | 54 |
| 7.1.4 Luokkakone | 55 |
| 7.2 Svenska Ylen Grävnavet | 57 |
| 7.2.1 Datan hankinta | 62 |
| 7.2.2 Datan puhdistus ja analyysi | 63 |
| 7.2.3 Visualisointi tai muu lopputuote | 63 |
| 7.2.4 Fixa högskolan | 64 |
| 7.3 Ylen Plus-deski | 66 |
| 7.3.1 Datan hankinta | 72 |
| 7.3.2 Datan puhdistus ja analyysi | 73 |
| 7.3.3 Visualisointi tai muu lopputuote | 74 |
| 7.3.4 Tuloerolaskuri | 75 |
| | |
| 8 POHDINTA | 77 |
| 8.1 Työprosessi ja sen haasteet | 78 |
| 8.2 Kommunikaatio | 81 |
| 8.3 Avoimuus | 83 |
| 8.4 Tulevaisuus | 84 |
| 8.4.1 Kannattaako datajournalismi? | 85 |
| 8.4.2 Datajournalismin ulkoistaminen ja ulkopuolinen apu | 87 |
| 8.4.3 Millaista datajournalismia tulevaisuudessa tehdään? | 88 |
| 8.4.4 Datajournalismikoulutus | 92 |
| | |
| 9 LOPUKSI | 95 |
| | |
| KIRJALLISUUS | 97 |
| | |
| LIITE | |

1 JOHDANTO

Ari Heinonen (1997, 78) ennusti aikanaan onnistuneesti, että internetin informaatiolähteet, kuten keskusteluryhmät ja erilaisten tietolähteiden verkkosivut tulevat muuttamaan kommunikaatiokulttuuria ja johtamaan ihmisiä pois päin nykymuotoisesta ammattijournalismista. Tämä kehitys on saavuttanut mittasuhteet, joita oli Heinosen kirjan julkaisun aikaan lähes mahdotonta kuvitella. Pienetkin tekomme luovat datavirtoja, joiden kerääminen, järjestäminen ja tarinoiksi jalostaminen on tähän saakka ollut vaikeaa. Näitä käyttökelpoiseen muotoon muunnettuja virtoja käytetään jatkuvasti enemmän myös journalismin lähteinä. (Bell 2012, 48.)

Informaatiosta ei ole nykymaailmassa pulaa. Yleisön luottamusta taas on vaikea saavuttaa, ja juuri luottamus on journalismille tärkeää. Vaikka luottamusmarkkinoilla on tilaa, suurin osa mediataloista kilpailee edelleen perinteisillä huomiomarkkinoilla (Lorenz, Kayser-Brill & McGhee 2011), eli yrittää houkutella yleisöä tuotteidensa ja mainostensa pariin keinolla millä hyvänsä. Eräs keino ansaita tuo puuttuva luottamus on auttaa yleisöä hahmottamaan maailmaa muodostamalla johtopäätöksiä datapaljoudesta.

Datajournalisti Paul Bradshaw'n (2012b) mukaan datajournalismin erottaa muusta journalismista juuri mahdollisuus yhdistää perinteinen uutisvainu ja tarinankerrontataito digitaalisen informaation valtavaan kirjoon ja määrään. Datajournalismi on luonteeltaan tutkivaa: dataa käyttävän journalistin työ ei ole raportoida tapahtumista ensimmäisten joukossa, vaan kertoa, mitä jokin kehityskulku itse asiassa tarkoittaa (Lorenz 2012).

Yhteiskuntatieteiden määrällisiä metodeja jo 1960-luvulla journalismiin soveltanut tarkkuusjournalismin oppi-isä Philip Meyer pohti ratkaisua journalismin luottamuspulaan kirjassaan *The New Precision Journalism* (1991): ”Puolustautuakseen manipulointia vastaan, media tarvitsee lisää itsevarmuutta, ja paras tie itsevarmuuteen on tiedon hankkiminen.”

Vaikka Meyer tarkoittaa tuolla tiedolla määrällisten tutkimusmenetelmien ja tilastotieteen tuntemusta, jokaisen toimittajan ei tarvitse olla numeronmurskaaja. Monien ilmaisten ja

helppokäyttöisten ohjelmien avulla datajournalismi on viime vuosina tullut teknologiaan vihkiytymättömienkin ulottuville. Usein laadukkaan datajournalismin tuottaminen kuitenkin vaatii toimittajan, ohjelmoijan ja graafikon työpanosta, ja yksittäisellä toimittajalla harvoin on kaikkia näitä taitoja. Erilaista ammatillista kieltä puhuvien osaajien yhteen sovittaminen onkin yksi datajournalismin suurimpia haasteita ja mahdollisuuksia. Myös vaativimpien datajournalististen prosessien ulkoistamista pohditaan jatkuvasti. Esa Väliiverosen (2011, 30) mukaan eräs tämänhetkisen journalistisen kulttuurin kipupisteistä näyttää olevan perinteisen individualistisen toimittajaihanteen ja ryhmätyötä korostavien muutosvaatimusten välinen jännite.

Kulttuurin lisäksi muutoskipuja kokee koko media. Mediayhtiöt etsivät kuumeisesti uutta sisältöä, joka houkuttelisi yleisöä perinteisen median pariin. Esimerkiksi USA:ssa suuri osa lehtiä lukevasta yleisöstä on siirtynyt paperituotteiden parista verkkoon (Nygren, Appelgren & Hüttenrauch 2012, 82). Aitamurron (2011) mukaan uutisorganisaatiot etsivät taloudellisesti kestäviä tapoja käyttää datajournalismia ja esimerkiksi visioivat datavarastoiksi muuttumista sekä maksullisten data-analyysipalveluiden tarjoamista

Kasvavan verkkopanostuksen myötä datajournalismin suosio onkin kasvussa toimituksissa ympäri maailmaa. Reynolds Journalism Institutin kyselyn mukaan yhdysvaltalaisen päivittäin ilmestyvien lehtien toimittajat etsivät jo rutiininomaisesti juttujaan varten dataa verkosta. European Journalism Centren datajournalismikyselyyn vastanneista 108 toimittajasta suurin osa toimii dataprojekteissa. (Sirkkunen 2011, 7.) Lorenz ym. (2011) kuitenkin väittävät, että suurin osa datajournalismin pioneereista ja verkkojournalisteista joutuu edelleen työskentelemään liian pienillä resursseilla ja vääränlaisilla työkaluilla, eikä kaikkia datan mahdollisuuksia siten voida hyödyntää.

Nykymuotoisen datajournalismin edelläkävijät työskentelevät pääosin Yhdysvaltain ja Ison Britannian suurissa toimituksissa. Uraa uurtavia datajournalistisia juttuja ovat julkaisseet esimerkiksi The Guardian, The New York Times, The Texas Tribune ja Die Zeit. Myös pienemmät maat ja pienemmät toimitukset ovat päässeet mukaan kehitykseen. Suomessa datajournalismia käyttävät eniten Yle Uutiset, Helsingin Sanomat sekä Svenska Yle.

Satunnaisesti datajournalismia hyödyntävät myös Suomen Kuvalehti, MTV Uutiset, Iltasanomat, STT ja Kaleva. Datajournalismin työtapojen voi olettaa yleistyvän tulevaisuudessa myös parhaiten resursoitujen toimitusten ulkopuolella, koska viime vuosina datajournalismi- ja taulukkolaskentakurssien määrä on kasvanut sekä toimittajakouluissa että toimituksissa.

Datapaljouden tulkitseminen kiinnostaa myös journalismin ulkopuolella: Suomessa alalla toimivat esimerkiksi talouslukuja visualisoivan Veropuu-sovelluksen kehittänyt yritys Hahmota ja datavisualisointeihin ja informaatio suunnitteluun erikoistunut Flo Apps. Nygrenin, Appelgrenin ja Hüttenrauchin (2012, 82 - 83) mukaan median talousvaikeuksien ja leikkausten aiheuttamaa tutkivan journalismin tyhjiötä on syntynyt täyttämään myös useita journalistisia, voittoa tavoittelemattomia organisaatioita, kuten Pro Publica, Center for Public Integrity, Sunlight Foundation ja California Watch.

Datajournalismin tutkimusta ja aihetta käsittelevää akateemista kirjallisuutta on julkaistu melko vähän. Siksi suuri osa tämänkin tutkielman lähteistä on erilaisia käytännön oppaita ja datajournalistien blogikirjoituksia. Tähän mennessä merkittävimmät tutkimukset ovat Sylvain Parasien ja Eric Dagiralin tutkimus *Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago* (2013) ja Cindy Royalin *The journalist as programmer: A case study of the New York Times interactive news technology department* (2010).

Suomessa on julkaistu Tanja Aitamurron, Esa Sirkkusen, Pauliina Lehtosen selvitys *Trends In Data Journalism* (2011). Jyväskylän yliopistossa on tällä hetkellä käynnissä Turo Uskalin johtama DaJoRe-projekti, jossa tutkitaan datajournalismia Suomessa, Englannissa ja Yhdysvalloissa. Ruotsissa Södertörnin yliopiston tutkijat Gunnar Nygren, Ester Appelgren ja Helge Hüttenrauch tutkivat datajournalismia Datajournalistik-tutkimusprojektissaan. Heidän vuonna 2012 kirjoittamansa artikkeli *Datajournalistik – ett växande område* julkaistiin Nordicom-Information -lehdessä.

Suomessa datajournalismia on käsitelty useissa opinnäytetöissä, viimeisimpänä Minna Rapelin pro gradu -tutkielma *Data Journalism: An Outlook for the Future Processes*

(2013). Käytäntöön painottuvat opinnäytetyöt aiheesta ovat tehneet Riikka Koskenranta (2013) sekä Minna Ala-Kyyny ja Susanna Patrikainen (2013).

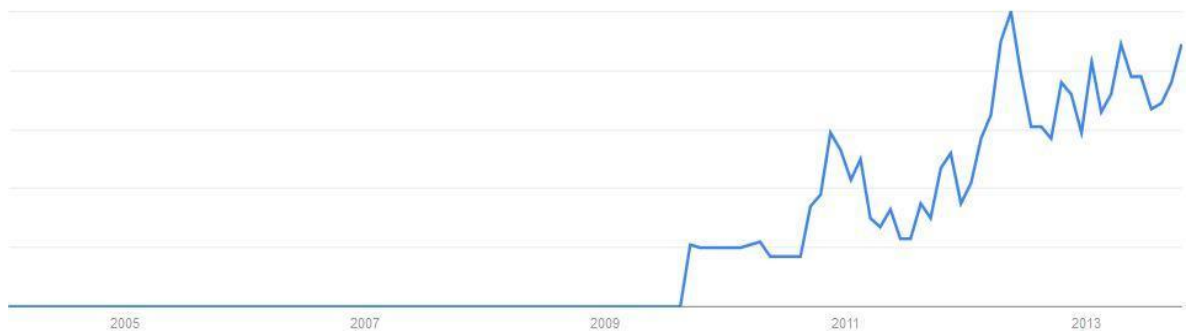
Perusteellisin pelkästään datajournalismiin keskittyvä julkaisu on Jonathan Grayn, Liliana Bounegrin ja Lucy Chambersin toimittama käytännönläheinen opas *The Data Journalism Handbook* (2012). Datajournalismia käsitellään myös useissa journalismioppaissa, kuten Paul Bradshaw'n ja Liisa Rohuman *The Online Journalism Handbookissa* (2011).

2. KÄSITTEET

2.1 Datajournalismi

Tässä tutkimuksessa datajournalismilla tarkoitetaan tietojärjestelmätieteestä, tilastotieteestä ja yhteiskuntatieteistä ammentavia journalistisen työn tapoja, joilla pyritään tiettyjä työkaluja apuna käyttäen hankkimaan, siivoamaan ja järjestelemään dataa muotoon, jossa sitä voi käyttää journalismin lähteenä. Datajournalismi hyödyntää verkonomaisia ilmaisumuotoja, kuten interaktiivisuutta ja pelillisyyttä, ja siihen liittyy kiinteästi yleisön osallistaminen työprosessin avaamisella sekä jutuissa käytetyn data-aineiston julkaisemisella. Tässä luvussa esittelen kirjallisuudesta kumpuavia näkökulmia, jotka johtivat minut tähän määritelmään.

Datajournalismi ei ole journalismin muotona uusi. Samankaltaisia työtapoja käyttäen on paljastettu epäkohtia ja selkeytetty suuria tietomassoja jo vuosikymmeniä. McGregorin (2013) mukaan datajournalismi ja sitä lähellä olevat termit, kuten laskennallinen journalismi (computational journalism) ovat vasta viime vuosina ilmestyneet toimittajien sanavarastoon. Nykymuotoista datajournalismia ovat tehneet tunnetuksi esimerkiksi Afganistanin sotaa koskevan tietovuodon jälkeiset julkaisut, joita vuonna 2010 tekivät sanomalehdet New York Times, Der Spiegel ja Guardian (Sirkkunen 2011). Datajournalismin tunnetuksi tulemistä noiden tapahtumien jälkeen kuvaa hyvin ”Data journalism” -hakusanalla tehtyjen Google-hakujen määrä (kuvio 1).



KUVIO 1 ”Data journalism” -hakusanan suosio Googlessa suhteutettuna kaikkiin hakusanalla tehtyihin hakuihin (Google Trends 2013, haettu 28.10.2013).

Englanninkielisessä kirjallisuudessa datajournalismin työtapoihin viitataan usein myös edellä mainitulla termillä *computational journalism* (McGregor 2013; Cohen, Hamilton & Turner 2011; Cohen, Chengkai & Cong 2011; Flew, Daniel & Spurgeon 2010).

Datajournalismin käsitteen määrittely on vaikeaa, koska sen voi katsoa tarkoittavan tietynlaisia journalistisen työn taitoja, työprosessia tai lopputuotetta (Hirst 2013). Hirst nostaa esille kaksi datajournalismin määritelmää:

Taito löytää datasta tarinoita ja kertoa ne uudelleen. Tähän sisältyy huomio siitä, että datajournalismi on tarinoiden löytämistä tietyn tyyppisestä lähteestä (lähteenä data) ja niiden viestimistä, mutta se ei vaadi tarinan kertomista millään tietyllä tavalla.

Journalismia, jossa data on yksi lähteistä, joita käytetään jutun hankkimiseen tai siihen liitettyä. Tässä tapauksessa datan rooli liittyy jutun lähteyttämiseen tai sen kommunikoimiseen tai jopa molempiin --.

Data Journalism Handbookin (Gray, Chambers & Bounegru 2012) esipuheessa New York Timesin datajournalismitiimin johtaja Aron Pilhofer määrittelee datajournalismin laajasti, hänen mielestään se on sateenvarjomainen termi, joka kattaa alati kasvavan joukon erilaisia työkaluja, tekniikoita ja kerronnan tapoja:

Se voi sisältää kaikkea perinteisestä tietokoneavusteisesta journalismista (datan käyttäminen lähteenä) edistyneimpään datan visualisointiin ja uutissovelluksiin. Yhdistävä päämäärä on journalistinen: tiedon tarjoaminen ja analysointi, jonka tarkoituksena on kertoa meille kaikille päivän tärkeimmistä asioista.

Aitamurron (2011) haastatteleminen datajournalistien mukaan datajournalismi on työkalu, jolla voidaan tilastotieteellisten metodien, tiedon visualisoinnin ja interaktiivisuuden keinoin tuottaa parempaa journalismia. Toimittajien mielestä datajournalismi on tapa löytää kätkeytyjä juttuaiheita. Uutisorganisaatioiden tasolla se taas nähdään tapana palvella yhteistä etua ja vaikuttaa yhteiskuntaan positiivisesti.

Hyvä apukäsite datajournalismin määrittelyssä on tietokoneavusteinen journalismi, eli CAR (Computer Assisted Journalism), joka sai lähtölaukauksensa erityisesti Philip Meyerin 1960-luvulla tekemästä työstä. Garrisonin (1998) mukaan CAR on laaja käsite, joka sisältää

kaikki tavat, joilla tietokoneita käytetään uutisten etsinnän apuna. Se voi sisältää esimerkiksi verkossa tehtyä tutkimusta ja tietokantajournalismia. Nygrenin, Appelgrenin ja Hüttenrauchin (2012, 85) mukaan toimittajat käyttivät Meyerin kehittämiä CAR:in työtapoja pelkästään viranomaisdatan analysointiin. Heidän mielestään nykyaikainen datajournalismi kiinnittää CAR:ia enemmän huomiota tutkimustulosten ja tietokantojen esittämiseen ja tiedon visualisointiin. Datajournalismi on CAR:ia vuorovaikutteisempaa: yleisö on toimittajalle myös tärkeä lähde, tuotannon apu ja julkaistun datapaljouden käyttäjä. Lorenz (2010) pitää CAR:ia tekniikkana, siinä missä datajournalismi on prosessi, joka vaikuttaa perustavalla tavalla koko journalismin työnkulkuun. Myös Lorenz korostaa teknologisen kehityksen, kuten pilvipalveluiden, nopeiden internetyhteyksien ja tehokkaiden tietokoneiden nostaneen datan suodatuksen ja prosessoinnin tasolle, joka olisi ollut sensaatio vielä muutamia vuosia sitten.

Bounegrun (2012) mukaan datajournalismi on tapa vastata uusiin olosuhteisiin, mutta myös osa historiallista jatkumoa. Vaikka datajournalismin ja CAR:in tekniikoissa ja tavoitteissa ei olekaan merkittävää eroa, datajournalismi-termin syntyminen vuosituhaten alussa merkitsee alan siirtymistä uuteen vaiheeseen. Tässä vaiheessa vapaasti saatavilla olevan datan valtava määrä, helppokäyttöiset työkalut ja joukkoistamisen keinot mahdollistavat sen, että yhä useampi ihminen voi työskennellä yhä suurempien datamäärien kanssa aiempaa helpommin.

McGregor (2013) huomauttaa, että tietokoneiden mahdollisuuksien hyödyntäminen on tuonut journalismin kentälle uusia ihmisiä toimitusten ja CAR-perinteen ulkopuolelta. Sen taas voi katsoa vaikuttavan työtapojen, niiden tausta-ajatusten ja siten myös käsitteen muotoutumiseen. Parasie ja Dagiral (2013) huomasivat Chicagolaisia koodaritoimittajia ja heidän työtään tutkiessaan, että nämä ihmiset olivat usein myös datan ja hallinnon avoimuutta ajavia aktiiveja. Taustansa vuoksi heidän käsityksensä tiedon luonteesta ja datan roolista journalismissa poikkesi kolmella tavalla CAR-perinteen edustajien ajattelusta.

1. *Uutiset ovat rakenteista informaatiota.*

Uutinen ei ole vain dataan kätkeyty tarina, vaan koneluettavassa muodossa oleva data on itsessään osa uutista. Jos data on oikealla tavalla jaoteltua ja riittävän kattavaa, se ei kätke mitään. Datan julkaisun tulisi olla osa uutistystä. Kaikki uutistystyön prosessit, tiedonkeruu, analysointi ja lopputuloksen esittäminen, voidaan hoitaa nopeammin tietokoneiden avulla, koska informaatio on rakenteista. Uutta tietoa voidaan tuottaa yhdistelemällä ominaisuuksiltaan erilaisia datasettejä. Lisäksi journalismiin ja ohjelmointiin tulisi toimituksissa suhtautua toisistaan erottamattomina, koska ne molemmat käsittelevät rakenteista tietoa ja tietokantoja.

2. Yleisölle on annettava mahdollisuus ja välineet tutkia dataa.

Yleisölle on annettava mahdollisuus tarkastaa data, yhdistellä sitä haluamallaan tavalla ja käyttää sitä myös omiin tarkoituksiinsa. Yleisö on aktiivinen ja osallistuva ja pystyy löytämään merkityksiä datasta myös ilman toimittajan apua.

3. On vähennettävä riippuvuutta hallinnon agendoista datan luovuttamisen suhteen.

Kehittyneen teknologian ja datan kasvavan määrän vuoksi journalistien työhön kuuluvan hallinnon toimien valvonnan ei enää tarvitse olla riippuvaista hallinnon itse julkaisemasta datasta.

Parasien ja Dagiralin haastatteluissa ilmennyt ajatus ohjelmoinnin ja journalismin erottamattomuudesta saa vastakaikua muiltakin tutkijoilta. Financial Timesin datajournalisti Cynthia O'Murchun mukaan datan analysoiminen ja tulkitseminen on nykyjournalistin tärkeä työväline, mutta ei oma journalismin koulukuntansa. Hän korostaa, että datan analysoinnin ja visualisoinnin ei tulisi olla itsetarkoitus, vaan työkalu, jolla voidaan päästä lähemmäs totuutta. (Gray ym. 2012.) Nykymääritelmän mukaan journalismi on, ainakin ihannetapauksessa, ajankohtaista ja faktapohjaista joukkoviestintää (Kunelius 2009, 21), eikä datajournalismi perusteiltaan poikkea tuosta ajattelusta.

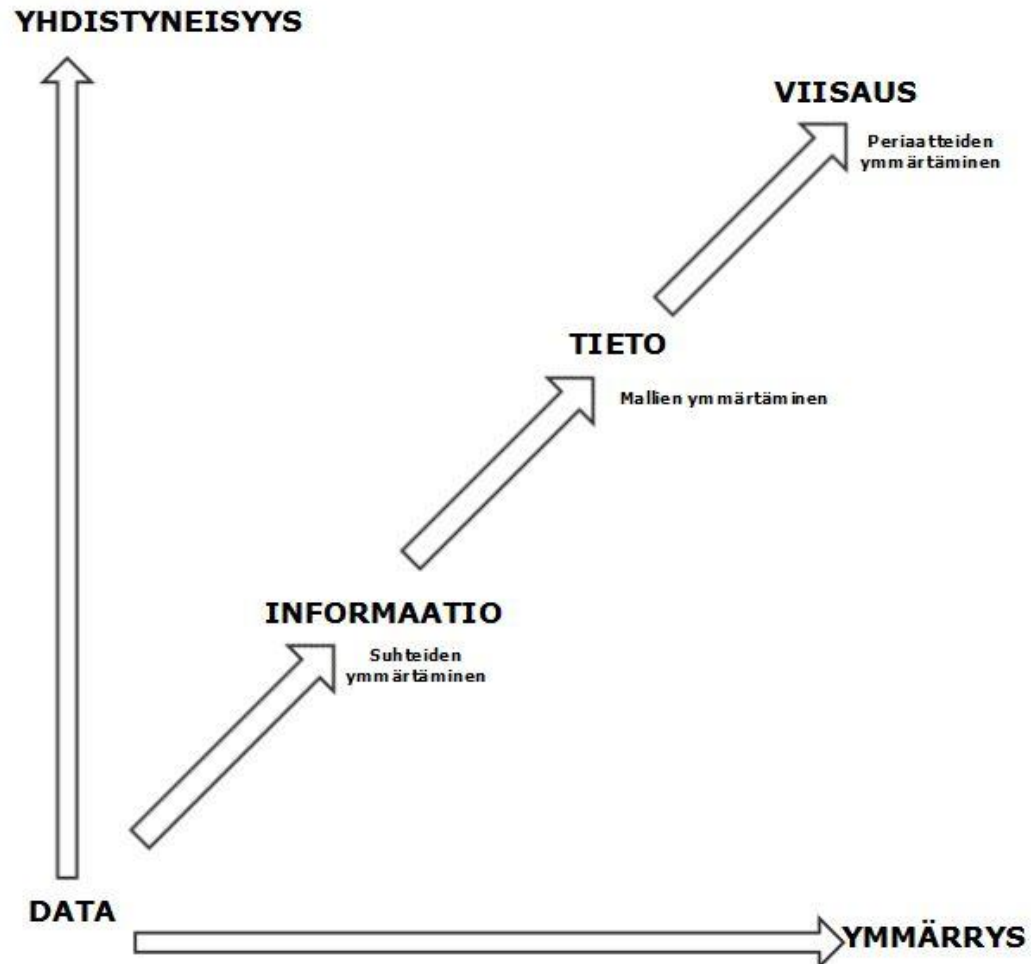
Myös CAR-pioneeri Philip Meyerin (1999, 4 - 5) mukaan tietokone on vain yksi, joskin toimiva apuväline totuuden tavoittelussa ja ihmiselle ominaisen toiveajattelun, taikauskon, valikoivan havainnoinnin ja muiden virheiden voittamisessa:

Tietokoneita käytetään toimituksissa niin monin eri tavoin, että se ei enää määritä meitä, jos se nyt koskaan on määrittänytään. On tullut aika julistaa CAR voittoisaksi ja siirtyä eteenpäin tuoreempaan, kunnianhimoisempaan konseptiin. Me tarvitsemme sitä ja niin tarvitsee myös maailma.

Meyer kyseenalaisti CAR-käsitteen jo viisitoista vuotta sitten: hänen mielestään tietokoneen käyttö toimituksissa oli jo silloin itsestäänselvyys ja tietokoneavusteisesta journalismista puhuminen oli vain este, joka häytti journalismin todellisten kamppailuiden ymmärtämistä.

2.2. Avoin data – datajournalismin polttoaine

Data voi olla datajournalismin raaka-aine, keino jolla journalistinen tarina kerrotaan, tai molemmat (Bradshaw 2012b). Informaatiotutkimuksen käsitteistössä data tarkoittaa merkeistä ja symboleista koostuvaa potentiaalista informaatiota. Datasta voidaan muodostaa informaatiota, joka jalostuu tiedoksi, kun vastaanottaja tulkitsee sen ja antaa sille merkityksen. (Poikola, Kola & Hintikka 2010, 14.) Datan piirteisiin voidaan lukea myös, että se on digitaalisessa muodossa, ja sitä voidaan käsitellä sekä tallentaa tietoteknisesti (Kuutti 2011, 599).



KUVIO 2 Tiedon hierarkia (Bellinger, Castro & Mills 2004).

Tarkemman kuvan datan raaka-ainemaisuudesta saa tutustumalla systeemiteoreetikoiden ajatteluun. Gene Bellingerin, Durval Castron ja Anthony Millsin (2004) systeemiteoreetikko Russel Ackoffin jaotteluun perustuva malli (kuvio 2) jaottelee tiedon tasot ymmärryksen (understanding) ja yhdistyneisyyden (connectedness) asteen mukaisesti. Data tarkoittaa faktaa, lausetta tai tapahtumaa, jolla ei ole suhteita muihin asioihin ("Nyt sataa"). Datan muuttuminen informaatioksi vaatii jonkinlaisten asioiden välisten suhteiden, kuten syy-seuraus-suhteen ymmärtämistä ("Lämpötila laski, ja sitten alkoi sataa"). Tieto taas edustaa mallia, joka yhdistelee tapahtumia ja antaa mahdollisuuden ennustaa tulevaa melko tarkasti ("Jos ilman kosteus on korkea ja lämpötila laskee merkittävästi, ilmakehä ei useinkaan pysty pitämään kosteutta itsessään, joten alkaa sataa"). Viisaus on sitä, että ymmärretään tietoon sisältyvät perusperiaatteet, jotka tekevät tiedosta sen, mitä se on ("Sade johtuu siitä, että sataa. Tämä toteamus sisältää ymmärryksen tapahtumaketjusta,

johon kuuluvat sade, haihtuminen, ilmavirrat, lämpötila, korkeuserot ja uusi sade”). (Bellinger ym. 2004.)

Dataa tuotetaan ja varastoidaan enemmän, kuin koskaan maailman historiassa. Vuonna 2011 dataa arvioitiin tuotetun 1,8 zettatavua, eli noin triljoona gigatavua. Datan määrän myös ennustettiin jatkavan valtavaa kasvuaan. (Kundra 2011, 3.) Saatavilla olevan datan määrän kasvua pidetään yhtenä nykymuotoisen datajournalismin synnyn tärkeimmistä syistä (Bounegru 2012). Monen datajournalistin juuret ovat datan avoimuutta kannattavissa liikkeissä, ja tämä näkyy myös heidän työssään avoimuuden arvostuksena (Parasie & Dagiral 2013, 855). Vaikka kaikki datajournalistien lähteenään käyttämä data ei ole avoimesti saatavilla, suurin osa ainakin suomalaisten datajournalistien työssään käyttämästä datasta täyttää vähintään osan avoimen datan määritelmästä. Avoimen datan puolesta kampanjoivan Open Knowledge Foundationin tiivistetty määritelmä on se, että data tai muu sisältö on avointa, jos kuka tahansa voi käyttää, uusiokäyttää tai jakaa sitä eteenpäin, eikä sen käytölle saa asettaa muita vaatimuksia, kuin lähteen mainitseminen tai muokatun datan edelleen avoimena pitäminen (Open Knowledge Foundation 2014). Viranomaisdatan avaamista käsittelevässä oppaassaan Poikola ym. (2010, 34) listaavat, että avoin julkinen data on kaikkien vapaasti ja ilmaiseksi saatavilla, helposti löydettävissä ja käytettävissä, sekä vapaasti uudelleen käytettävissä. Datan on myös oltava alkuperäistä ja ajantasaista ja sen on oltava tallennettu ymmärrettävään, avoimeen ja koneluettavaan formaattiin.

Yhdysvallat ja Iso-Britannia ovat olleet julkisen datan avaamisen edelläkävijöitä, joiden esimerkkiä muu maailma on seurannut. Euroopan unioni otti askeleen viranomaisdatan avaamisen suuntaan, kun avoimuuden kasvattamiseen tähtäävä PSI (Public Sector Information) -direktiivi hyväksyttiin vuonna 2003. Vuonna 2007 toimeenpantu Inspire-direktiivi taas vapauttaa kansallisen maantieteellisen aineiston käyttöä. Suomessa valtioneuvosto teki vuonna 2010 periaatepäätöksen julkisten tietovarantojen avaamisesta. (Lehtonen 2011, 4-5.)

Datan avaamista on perusteltu sekä demokratian kehittämiseksi että liiketoiminnan tukemisella. Avoin data vapauttaa muuten datan hankkimiseen hukkaantuvia resursseja.

Avoimuuden myötä voidaan tehokkaammin panostaa datan hyödyntämiseen, mikä luo uusia palveluita, sekä uutta tutkimusta ja tietoa. (Poikola ym. 2010, 16.) Kundran (2011, 4) mukaan avoin data voi auttaa yhteiskuntaa korruption kitkemisessä ja hallinnon avoimuuden, läpinäkyvyyden ja osallistavuuden lisäämisessä. Kinnari (2013, 73) esitteli pro gradu -tutkielmassaan suomalaisia yrityksiä ja muita yksityisiä tahoja, joiden liiketoimintamalliin avoin data liittyy. Hän jakoi ne viiteen ryhmään:

1. Kaupalliset avoimen datan julkaisijat (HSL Reittiopas, HS Open)
2. Hankintaa ja muuntamista tekevät tahot (Louhos)
3. Datan analysoijat (Hahmota, Cloud'n'Sci, Asiakastieto)
4. Käyttäjäkokemuksen tarjoajat (mm. Duunitori.fi, Kansanmuisti, ReittiGPS)
5. Tukipalveluiden sekä konsultoinnin tarjoajat (Flo Apps, Logica)

Tällä hetkellä julkisen datan avaamisessa haasteita muodostavat ohjaavan lainsäädännön puute ja esimerkiksi yksityisyyteen liittyvät kysymykset. Myös dataa säilyttävien organisaatioiden vakiintunut työkalutuuri asettaa haasteita, mistä esimerkkinä se, ettei virastoissa välttämättä tiedetä kenen vastuulla datan avoimuuden takaaminen on, eikä kiinnostukseen aina ole suurta. Julkisen datan avaamista hidastavat myös vakiintumattomat datan käsittelyn ja tulkinnan käytännöt. (Lehtonen 2011, 8.)

3. DATAJOURNALISMIN HISTORIA

3.1. Maailmalla...

Uutisia on kaivettu datasta jo kauan ennen tietokoneita ja sähköisiä tietokantoja. Guardian -lehdessä julkaistiin ensimmäinen datajournalismia muistuttava juttu jo vuonna 1821. Se käsitteli Manchesterin kaupungin koulujen oppilasmääriä ja hintoja paljastaen samalla, kuinka moni lapsi sai ilmaisen koulutuksen ja kuinka monta köyhää lasta kaupungissa oli. (Rogers 2011.) Toinen kuuluisa 1800-luvun datajournalistinen työ on sairaanhoitaja Florence Nightingalen infografiikka brittiarmeijan kuolleisuudesta. Nightingale osoitti visualisoinnillaan, että suurin osa kuolemista johtui estettävissä olevista sairauksista, eikä taisteluista. (Bounegru 2012.)

Nykymuotoisen datajournalismin kannalta olennaisin kehitys alkoi kuitenkin vasta tietokoneiden tulosta toimitustyön avuksi (Bounegru 2012). Coxin (2000, 6) mukaan CAR:in aikakausi alkoi, Yhdysvaltain vuoden 1952 presidentinvaaleista, joissa CBS News käytti Remington Rand UNIVAC -tietokonetta Dwight D. Eisenhowerin ja Adlai Stevensonin välisen kisan lopputuloksen ennustamiseen. Ohjelmoijat olivat valmistelleet kaavoja, joilla tulos voitiin ennustaa kun vasta osa äänistä oli laskettu. Kisan oli odotettu olevan tiukka, mutta ennuste osoitti Eisenhowerin voittavan äänivyöryllä, mikä osoittautui myöhemmin todeksi. (Bounegru 2012; Cox 2000, 6.)

Vuoden 1952 vaalien jälkeen tietokoneiden tulo toimituksiin tapahtui kolmen päällekkäin kerääntyvän vaiheen kautta. Alun liiketoimintavaiheeksi kutsuttuna aikana tietokoneita käytettiin pääasiassa tilinpidon ja levikkiasioiden hoitamiseen. 1960-luvun puoliväliin mennessä tietokoneita käytettiin jo hyvin vaihteleviin tarkoituksiin, kuten esimerkiksi varastorekisterin pitämiseen ja sähköisten kirjoittamiseen. 1970-luvulla saapui tuotantovaihe, jossa tietokoneet valjastettiin laajemmin myös sisällöntuotantoon. Siitä siirryttiin informaatiovaiheeseen, jossa tietokoneet valjastettiin myös tiedonhaun avuksi. (Reavy 1996 Coxin 2000, 6 mukaan.)

1960-luvun lopulla toimittaja Philip Meyer alkoi soveltaa työssään yhteiskuntatieteistä ja tilastotieteistä tuttuja mittaamisen ja tutkimuksen tapoja. Hänen ensimmäinen kuuluisa tutkiva työnsä käsitteli vuoden 1967 mellakoita Detroitissa. Siinä hän osoitti kyselytutkimuksen ja IBM 360 -tietokoneella tehdyn analyysin avulla, että koulutetut osallistuivat mellakointiin yhtä lailla kuin kouluttamattomatkin. Meyerin työtä ja hänen vuonna 1973 ensipainoksensa saanutta kirjaansa *Precision Journalism: A Reporters' Introduction to Social Science Methods* pidetään CAR-perinteen peruskivinä (Parasie & Dagiral 2013, 856; Bounegru 2012; Cox 2000, 7). Meyerin mielestä toimittajan piti osata käyttää tietokantoja ja käsitellä ja analysoida dataa. Tarkkuusjournalismiksi kutsuttujen työtapojen ja tieteellisten metodien avulla toimittajat voisivat tulla vähemmän riippuvaisiksi poliitikkojen agendoista:

Passiivisen ja viattoman toimittajan ongelma on se, että häntä on liian helppo käyttää hyväksi, kuten ketä tahansa passiivista ja viatonta ihmistä. Suurimmassa osassa modernia journalismin kritiikkiä perustava teema on se, että mediaa manipuloivat liian helposti voimakkaat poliitikot ja heidän taitavat pr-ihmisensä, ja heidän tahtonsa määrittelee liian vaivattomasti sen, mikä on uutinen ja mikä ei. Puolustautuakseen manipulointia vastaan media tarvitsee lisää itseluottamusta, ja paras tie itseluottamuksen hankkimiseen on tieto. (Meyer 1991.)

Tarkkuusjournalismin voidaan siis katsoa olleen vastaus journalismin luotettavuuden kriisiin ja riippuvuuteen ulkopuolisista lähteistä ja tiedotteista. Se oli myös vastavoima 1970-luvulla yleistyneelle “uudelle journalismille”, joka haki inspiraatiota kaunokirjallisuudesta. (Bounegru 2012.) Tarkkuusjournalismin suosioon vaikuttivat myös tietokoneiden yleistyminen ja lehtien levikin laskusta kummunnut kustantajien huoli journalismin laadusta (Meyer 1991). Philip Meyerin lisäksi tietokoneita ja tieteellistä mittaamista käyttäneisiin pioneereihin kuuluivat esimerkiksi Elliot Jaspin, Clarence Jones, David Burnham, Rich Morin, Fred Tasker, Don Bartlett, ja James Steele. Tarkkuusjournalismin voidaan katsoa syntyneen vasta kun heidän käyttämänsä tekniikat levisivät laajempaan käyttöön. (Cox 2000.)

Pian Philip Meyerin Detroitin mellakoita käsittelevän työn jälkeen, vuonna 1969, The Miami Herald -lehden Clarence Jones avustajineen syötti tietokoneelle 13 000 reikäkortille tallennetut tiedot ja paljasti epäkohtia Daden piirikunnan rikosoikeusjärjestelmästä. Tämä

oli ensimmäinen kerta, kun viranomaisilta saatua dataa käytettiin tietokoneavusteisen journalismin lähteenä. (Cox 2000, 17.)

Tietokoneavusteisen journalismin käytäntöjen yleistymistä voi hahmottaa innovaation leviämisen s-käyrän avulla. Malli, jonka mukaisesti innovaatio, kuten uusi teknologia, idea tai toimintatapa, leviää sosiaalisessa järjestelmässä, on tarkasti dokumentoitu (Davenport, Fico & DeFleur 2002, 6). Hitaan alun jälkeen uuden innovaation leviäminen kiihtyy ja sitten taas tasoittuu, kun innovaatio on levinnyt kaikkien alan toimijoiden käyttöön. Ensimmäiset innovaation käyttöön ottajat ovat usein suuria toimijoita, joilla on varaa kantaa investoinnin riskit. (Garrison 1998.)

CAR oli Meyerin ensimmäisistä jutuista aina 1990-luvulle asti vain suurten lehtien työväline. Tietokoneet, tietokannat ja asiantuntemus maksoivat, eikä pienemmillä toimituksilla ollut niihin varaa. Mikrotietokoneiden yleistyminen 1980-luvulla madalsi tietokoneavusteisen journalismin kynnystä: ensin toimittajat ostivat koneita omaan käyttöönsä ja pikkuhiljaa myös toimitukset alkoivat hankkia uusia, pienikokoisia tietokoneita. Myös tietokantojen käyttö yleistyi ja jotkut toimitukset alkoivat kerätä omia tietokantojaan. Vuosikymmenen puolivälissä tietokoneita käytettiin jo yleisesti tutkivan journalismin apuna. (Garrison 1998; Cox 2000.) Tim Millerin vuonna 1986 Yhdysvalloissa tekemän kyselytutkimuksen mukaan verkosta materiaalia etsivien sanomalehtien määrä oli nelinkertaistunut edellisten neljän vuoden aikana. Tietokantoja käytti tietolähteinä jo yli puolet kyselyyn vastanneista sanomalehdistä. Myös sähköinen media siirtyi online-aikaan: vuoden 1988 alussa tehdyn kyselyn mukaan yli 25 prosenttia sadasta Yhdysvaltain markkinoiden johtavasta kanavasta käytti kaupallisia, maksullisia tietokantoja. (Miller 1988, 36.)

Varsinkin varhainen CAR-journalismi oli usein tutkivaa ja sen aiheet liittyivät korruptioon ja vähemmistöjen oikeuksiin. Hyvä esimerkki tästä on Atlanta Journal Constitution -lehden Bill Dedman, joka vuonna 1989 voitti Pulitzer-palkinnon *The Color of Money* -juttusarjasta, jossa hän paljasti Atlantan alueen pankkien epäävän lainoja ihonvärin perusteella. Tuloksensa Dedman sai vertaamalla Yhdysvaltain tilastokeskukselta (US Census Bureau)

ja kansalliselta finanssitarkastuslautakunnalta (Federal Financial Institution Examination Council) hankkimiaan datasettejä. (Parasie & Dagiral 2013, 856; Cox 2000, 11.)

Vuonna 1986 Providence Journal -lehdessä työskennellyt Elliot Jaspin vertasi ristiin hankkimiaan tietokantoja ja huomasi, että koulubussien kuljettajina työskenteli tuomittuja huumekauppiaita, ja että paikalliset viranomaiset antoivat itselleen alennuksia kiinteistöveroista. Todistettuaan aiheista tekemillään jutuilla, että relaatiotietokannat ja tietokantahauissa käytettävä SQL (Structured Query Language) -kyselykieli olivat hyvä uutisten lähde, hän matkusti ympäri Yhdysvaltoja opettaen tietokantataitoja. Vuonna 1989 hän perusti Missourin yliopistossa Missouri Institute for Computer Assisted Reporting -organisaation, josta sai alkunsa tietokoneavusteista journalismia edistävä yhdistys NICAR (National Institute for Computer-Assisted Reporting). (Cohen, Hamilton & Turner 2011; Paul 1999, 2.)

Tietokoneavusteisten juttujen määrä kasvoi räjähdysmäisesti 1990-luvulle tultaessa. Tietokantoja käyttävän journalismin määrää kasvattivat koneluettavan datan määrän kasvu ja internetin leviäminen. (Cox 2000, 11; Parasie & Dagiral 2013, 856). Myös keskisuurten ja pienten uutisorganisaatioiden CAR-projektien määrä kasvoi. CAR oli vuonna 1998 joissakin toimituksissa sulautettu kaikkeen toimitustyöhön ja sen työkalut olivat kaikkien toimittajien käytössä. (Garrison 1998.) 1990-luvun tietokoneavusteista journalismia määrittivät uudet työkalut, kuten taulukkolaskentaohjelmat, tietokantaohjelmat, tilastolaskentaohjelmat ja verkossa olevat lähteet, kuten sähköposti, keskusteluryhmät, tietokantakirjastot ja BBS:t (Bulletin Board System) (Houston 1996 Coxin 2000, 11 mukaan). 1990-luvulla journalismikoulutus laahasi CAR-taidoissa työelämää jäljessä, koska uusien työtapojen opettamiseen ei löytynyt rahaa, laitteita, intoa eikä riittävän taitavia kouluttajia (Davenport, Fico & DeFleur 2002, 6 - 7).

Tietokoneavusteisen journalismin työkalut ovat pysyneet pääpiirteittäin samoina Philip Meyerin *Precision Journalism* -kirjan ensijulkaisusta aina nykyaikaan asti: uutisia etsitään datasta taulukkolaskennan, tietokantaohjelmien ja verkon tarjoamien keinojen avulla (McGregor 2013). Nykyisessä datajournalismissa lopputuote on kuitenkin usein paperilehden artikkelin sijasta, tai lisäksi, visualisointi tai uutissovellus.

2000-luvulla merkittäviä datajournalistisia juttuja on tehty liikaa lueteltavaksi. Datajournalistitutkija Liliana Bounegru (2012) nostaa Adrian Holovatyn vuonna 2006 aloittaman Everyblock-palvelun yhdeksi nykymuotoisen datajournalismin ensimmäisistä ilmaantumisista. Everyblockin avulla käyttäjä pystyi selvittämään kartalta, mitä hänen lähiseudullaan tapahtuu. *Data Journalism Handbook* -kirjassa (Gray, Bounegru & Chambers 2012) listataan joitakin viime vuosien vaikuttavimmista töistä. Niihin kuuluu esimerkiksi Las Vegas Sun -lehden vuonna 2010 tekemä juttukokonaisuus Do No Harm, jossa paljastettiin useita vältettävissä olleita kuolemantapauksia ja loukkaantumisia sairaaloiden laskutustietoja analysoimalla. Tekstin lisäksi löydöksiä on esitetty interaktiivisilla visualisoinneilla ja kartoilla sekä perinteisemmällä taulukoilla. Hiukan erilaista datajournalismia edustaa AP:n Jonathan Strayn ja Julian Burgessin tekstianalyysin ja erilaisten algoritmien keinoin luoma visualisointi, jossa teemoiteltiin medialle vuodetuissa Irakin sodan asiakirjoissa esiintyviä sanoja.

Suurelle yleisölle datajournalismia ovat tehneet tunnetuksi esimerkiksi vuoden 2010 Afganistanin sotaa koskevan tietovuodon jälkeiset julkaisut, joita tekivät sanomalehdet New York Times, Der Spiegel ja Guardian (Sirkkunen 2011, 7; Bounegru 2012). Datajournalistisiin läpimurtoihin kuuluu myös Nate Silverin kohua herättänyt, mutta tarkkaan paikkansa pitänyt ennuste Barack Obaman voitosta vuoden 2012 presidentinvaaleissa (McGregor 2013).

3.2. ... ja Suomessa

Suomessakin tutkivat toimittajat ovat käyttäneet CAR-työtapoja jo pitkään ennen nykymuotoisen datajournalismin rantautumista. Ensimmäisen vaalikoneen julkaisi Yle vuonna 1996 (Aalto 2005), jonka jälkeen vaalidataa käyttävät sovellukset alkoivat raivata tietään suomalaiseen journalismiin. Helsingin Sanomat oli edelläkävijä verkkokerronnan ja interaktiivisuuden alalla, kun toimitus perusti 1990-luvun lopulla Verkkoliite-ryhmän, joka rakensi erilaisia webortaaseiksi kutsuttuja verkkoa varten suunniteltuja sisältöjä (Mäkinen,

haast. 16.12.2013). 2000-luvun alussa webortasit olivat jo kehittyneitä ja sisälsivät esimerkiksi erilaisia testejä ja Flash-animaatioita.

Helsingin Sanomien nykyinen datadeski sai alkunsa Esa Mäkisen aloitteesta.

Kulttuuritoimituksessa työskennellyt, datajournalismista kiinnostunut Mäkinen alkoi tehdä dataan pohjautuvia juttuja, joista ensimmäinen, kirjakustantamoiden laatuvertailu, julkaistiin vuoden 2011 alkupuolella. Vuosina 2011 ja 2012 Helsingin Sanomat myös järjesti HS Open -tapahtumia, joissa toimittajat, ohjelmoijat ja graafikot rakensivat yhteistyössä uutissovelluksia esimerkiksi vaalidatasta. Helsingin Sanomat osti joitakin näistä sovelluksista ja julkaisi ne verkkosivuillaan. Kesäkuussa 2012 Helsingin Sanomat perusti nykyisen datadeskinsä. (Mäkinen, haast. 16.12.2013.)

Yle avasi vuonna 2010 beta.yle.fi-sivuston, joka esitteli uudenlaisia verkkokerronnan muotoja, joukossa myös muutamia datapohjaisia toteutuksia (beta.yle.fi 2014). Ylen säännöllisempi datajournalismi alkoi Svenska Yleltä, kun Teemo Tebest aloitti ohjelmoijana Svenska.yle.fi-kehitystiimissä vuoden 2012 alussa. Verkkokehitystyön ohella Tebest teki visualisointeja, kuten tutkivaan journalismiin keskittyvän Radar-ohjelman verkkouutisessa julkaistun kartan, joka esitteli Suomen kuntien alkoholin kulutusta. Pohjana tuossa Ylen datajournalismi-innostusta lietsoneessa Google Maps -visualisoinnissa käytettiin Helsingin Sanomien avoimena lähdekoodina julkaisemaa kuntakonetta (Tebest 2012). Syksyllä 2012 Svenska Ylen uutisissa päätettiin lisätä datajournalismipanostusta ja Alexander Ugglan aloitti työn uutisten tietokantatoimittajana. Hän työskenteli neljä kuukautta ja lähti sitten kahdeksan kuukauden virkavapaalle, jonka ajaksi hänen sijaisekseen tuli Peter Sjöholm. Syksyllä 2013 perustettiin nykymuotoinen Grävnavet-ryhmä, jossa datajournalismista vastaavat nämä kaksi toimittajaa yhdessä tuottaja Eva-Maria Koskisen kanssa. (Sjöholm, haast. 18.12.2013.) Aloitteen Ylen uutis- ja ajankohtaisohjelmien kanssa työskentelevän Plus-deskin perustamiseen teki uutis- ja ajankohtaistoiminnan internet-päällikkö Mika Rahkonen, ja henkilöstöä alettiin rekrytoida marraskuussa 2012. (Salminen, haast. 19.12.2013.) Plus-deski perustettiin keväällä 2013 osana Yle Uutisten panostusta verkkosisältöjen tuottamiseen. Tuolloin deskiin siirtyi myös Svenska Ylellä työskennellyt Teemo Tebest.

Tällä hetkellä säännöllisesti datajournalismia tuottavat ainoastaan Yle ja Helsingin Sanomat. Niiden lisäksi datajournalismia ovat hyödyntäneet ainakin Taloussanomat, Suomen Kuvalehti, MTV Uutiset, Ilta-Sanomat, STT ja Kaleva. Useat suomalaismediat ovat myös ostaneet datajournalistisia juttuja freelance-toimittaja Jens Finnäsilta, joka nykyään työskentelee kansainvälisessä datajournalismitoimisto Journalism++:ssa.

4 DATAJOURNALISMIN TYÖPROSESSI

Alan tekijät ja tutkijat korostavat, että datajournalismi on ennen kaikkea journalismia ja se korostuu myös datajournalistisessa työprosessissa (O'Murchu 2012; Aitamurto 2011, 10; Nygren, Appelgren & Hüttenrauch 2012, 83). Pohjimmiltaan datajournalismin työkulku noudattaa perinteisen journalistisen prosessin kaavaa, jonka vaiheet Sirkkusen ja Kuusiston (1999, 20 - 21) mukaan ovat ideointi ja aiheen valinta; jutun tekeminen, joka sisältää tiedon hankinnan, jutun kokoamisen ja editoinnin; sekä prosessin kontrollointi, jonka hoitavat päälliköt ja toimitussihteerit. Näiden vaiheiden jälkeen seuraa jutun jakelu toimituksen käyttämän kanavan kautta.

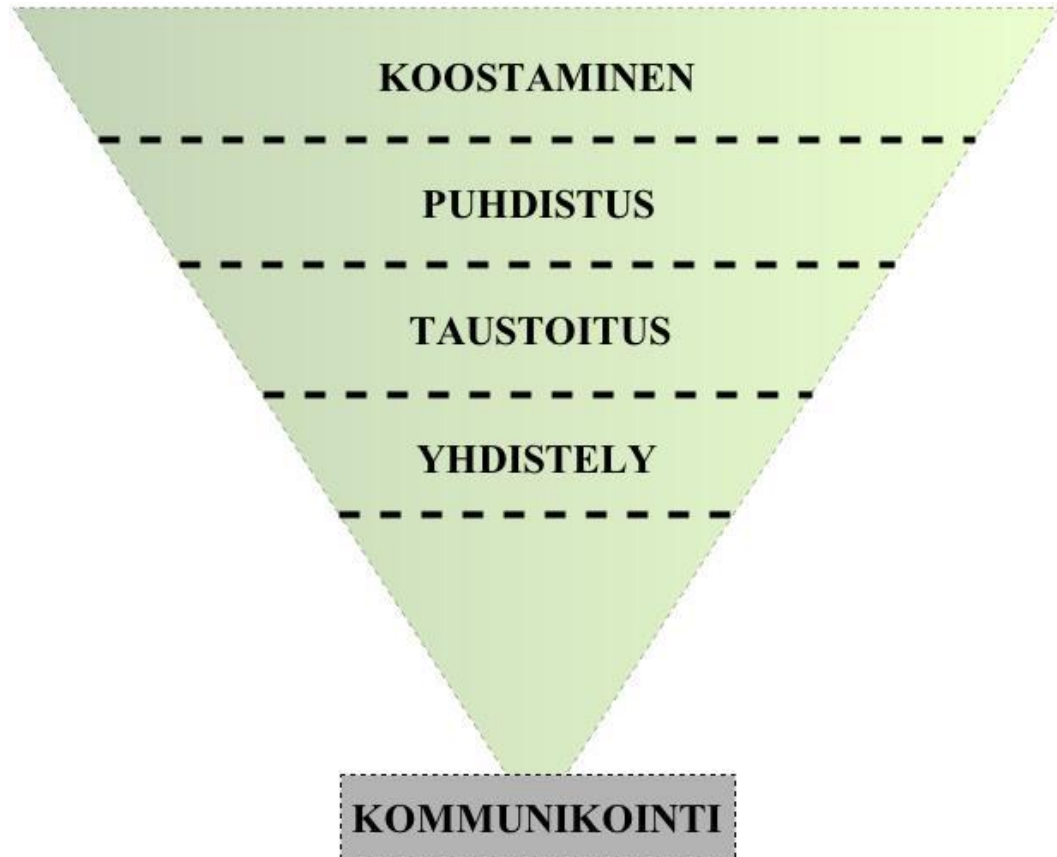
Perinteiseen journalismiin verrattuna datajournalistisessa työprosessissa kuitenkin korostuvat monialainen tietotekninen osaaminen ja tiimityö (Aitamurto 2011, 12; Royal 2010). Datajournalismin pioneerit ovat usein olleet monilahjakkuuksia, jotka pystyvät itse hoitamaan tuotantoprosessin kaikki vaiheet, tai ainakin useita niistä (Lorenz 2011). Nykyäänkin datajournalistina oleminen voi tarkoittaa monia asioita: datajournalisti voi olla esimerkiksi tilastotieteen osaaja, vuorovaikutteisia sovelluksia luova suunnittelija tai datan käsittelyn taitava toimittaja – roolit kehittyvät jatkuvasti (Bell 2012).

Toimituksiin on kuitenkin perinteisesti hakeutunut vain harvoja sekä tietotekniikan että journalismin alalla lahjakkaita tekijöitä, joten yleensä datajournalismin voi väittää olevan tiimityötä. Erilaisten osaajien yhteistyön kulttuuriset ja tekniset raja-aidat ovatkin hitaasti kaatumassa ja tiimityötä vaativien journalististen projektien määrä kasvaa (Bradshaw 2012a, 2).

Aitamurron mukaan datajournalismitiimeissä on yleensä ihmisiä, joilla on osaamista joko journalismissa, web-kehityksessä, data-analyysissa, visualisoinnissa tai tilastotieteessä. Kaikki eivät välttämättä ole toimittajia, mutta yleensä ainakin tiimin johtaja on toimittaja, jolla on myös data-analyysiosaamista. Tiimin toiminnassa korostuu jatkuva oppiminen ja uusien työkalujen kokeileminen. (Aitamurto 2011, 12.)

On selvää, että datajournalismitiimin roolit ovat monelta osin päällekkäisiä ja varsinkin pienemmissä toimituksissa sama ihminen hoitaa useaa roolia. Kirk (2012) esittää kahdeksan erilaista roolia, jotka datan visualisointiin tähtäävässä prosessissa esiintyvät.

1. *Aloitteentekijä (Initiator)*. Hänellä on kysymys, johon hän tahtoo vastauksen.
2. *Datan käsittelyn asiantuntija (Data scientist)*. Hän kokoaa ja valmistelee datan.
3. *Toimittaja (Journalist)*. Hän hoitaa tarinankerronnan, etsii kärjen ja näkökulman ja hankkii vastaukset kysymyksiin. Hän myös miettii, mihin kysymyksiin lopputuotteen tulee vastata.
4. *Tietojenkäsittelyn asiantuntija (Computer scientist)*. Hän luo lopputuotteen teknisen toteutuksen ja hoitaa projektin vaatiman ohjelmoinnin sekä muut vaikeat tekniset prosessit.
5. *Suunnittelija (Designer)*. Hän suunnittelee lopputuotteen ulkoasun ja käyttäjäkokemuksen.
6. *Kognitiotieteen asiantuntija (Cognitive scientist)*. Hän varmistaa, että visuaalinen kommunikaatio on tehokasta.
7. *Kommunikoiija (Communicator)*. Hän hoitaa neuvottelun asiakkaan kanssa sekä järjestää tiedonkulun asiakkaan ja suunnittelijatiimin välillä. Hän myös hoitaa julkaisun.
8. *Projektinjohtaja (Project manager)*. Hän valvoo projektin etenemistä ja huolehtii aikatauluista. Hän myös tarkastaa kaiken toimivuuden ja tarkkailee eettisten periaatteiden toteutumista.



KUVIO 3 Datajournalistinen työprosessi (Bradshaw 2011a).

Paul Bradshaw (2011a) mallintaa datajournalistista työprosessia kärjellään seisovalla kolmiolla (kuvio 3). Hänen mukaansa muoto on sopiva, koska prosessin alussa käsittelyssä on suuri määrä informaatiota, jota jalostetaan ja fokusoidaan, kunnes päästään tuloksiin. Bradshaw jakaa prosessin viiteen osaan, joista viidettä, kommunikointia, avaam tarkemmin datajournalistisen työn lopputuotetta käsittelevässä luvussa (s. 28).

1. *Koostaminen*. Prosessi alkaa joko ideasta, joka vaatii tuekseen dataa tai datasetistä, josta juttuaihetta lähdetään etsimään. Tämä vaihe määrittelee työn datajournalismiksi ja on pohjana myöhemmille vaiheille.

2. *Puhdistus*. Jotta datasta voidaan uskottavasti löytää tarina, siihen täytyy luottaa. Datan puhdistuksessa poistetaan inhimilliset virheet ja muutetaan data muotoon, joka on käyttökelpoinen ja sama, kuin muussa datassa, jota projektissa käytetään.

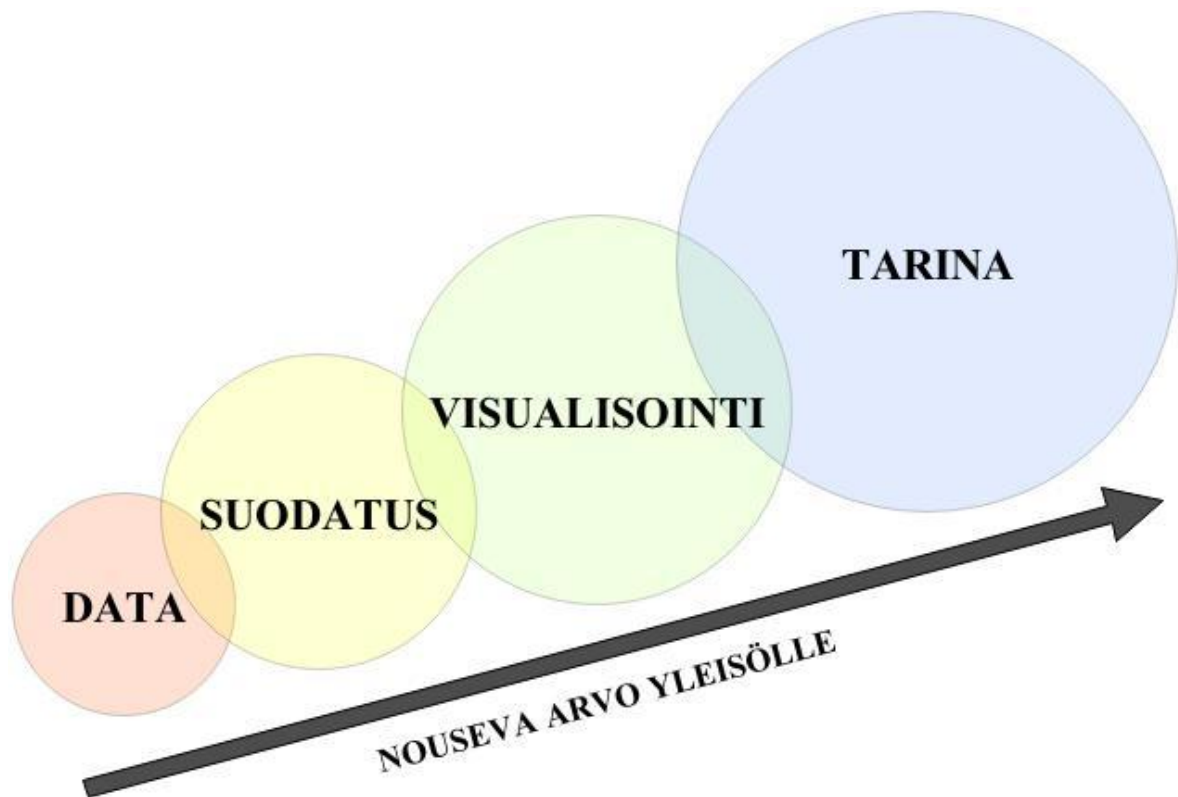
3. *Taustoitus*. Toimittajan on otettava selvää datan taustoista, kuten mitä tahansa journalismin lähdettä käsiteltäessä: kuka on kerännyt sen, miksi ja miten? Myös datassa esiintyvät termit ja lyhenteet on selvitettävä. Tässä vaiheessa voi myös herätä tarve hankkia lisätietoa ja etsiä datalle kontekstia esimerkiksi muista dataseteistä. Esimerkiksi kaupungissa tehtyjen rikosten määrää voi verrata rikosten määrään viisi vuotta sitten.

4. *Yhdistely*. Usein hyvä juttuaihe vaatii useamman datasetin yhdistämistä. Esimerkki tästä on datasetin yhdistäminen karttaan tai datasettien yhdistäminen yhteisten datapisteiden, kuten kuntanumeroiden tai nimien avulla.

5. *Kommunikointi*. Prosessin tulokset esitellään yleisölle esimerkiksi tekstin, visualisoinnin tai uutissovelluksen avulla. Aikaansaannokset jaetaan eri kanavien kautta.

Aitamurto (2011,11) jakaa datajournalistisen työprosessin seitsemään osaan, jotka ovat melko lailla samat kuin Bradshaw'n mallissa.

1. Jutun ytimen tunnistaminen ja datan roolin määrittäminen.
2. Jutun kannalta oikeiden datasettien löytäminen ja hankkiminen.
3. Datan muokkaaminen analysointikelpoiseksi, eli virheiden korjaaminen.
4. Datan analysointi oikeilla työkaluilla. Datan yhdistäminen muihin datasetteihin, mikäli se on tarpeellista.
5. Jutun tuottaminen. Tekstin, visualisointien ja interaktiivisten elementtien rakentaminen.
6. Analyysissa käytettyjen datasettien julkaiseminen.
7. Lukijoiden kutsuminen osallistumaan datan uudelleen käyttöön, jutun kommentointiin ja jakamiseen sosiaalisessa mediassa sekä lisäsisällön lähettämiseen esimerkiksi Flickrin, Facebookin ja Twitterin kautta.



KUVIO 4 Datajournalismi ja datan merkityksen kasvu (Lorenz 2010)

Datajournalisti Mirko Lorenz (2010) kuvaa datajournalismia jalostusprosessiksi, jossa raaka data muutetaan joksikin merkitykselliseksi (kuvio 4). Prosessin tuloksena datan arvo yleisölle kasvaa, kun monimutkaiset asiat muuttuvat selkeäksi tarinaksi, joka on helppo ymmärtää ja muistaa. Lorenzin ajattelussa data kasvattaa arvoaan samaan tapaan kuin systeemiteoreettisessa ajatuksessa tiedon hierarkiasta (ks. kuvio 2).

Datajournalistista työprosessia kuvaavat mallit noudattavat pääpiirteittäin tätä samaa kaavaa, jossa ideasta tai datasetistä luodaan puhdistuksen, muokkaamisen ja yhdistelemisen jälkeen lopputuote, joka on yleensä visualisoinnin ja tekstin yhdistelmä, mutta saattaa olla muutakin. Yleisön rooli prosessissa saattaa olla pelkkä lopputuotteen käyttäjä, jakaja ja kommentoija tai yleisö saatetaan kutsua apuun joko tiedon tuottajaksi tai käsittelijäksi. Esimerkkejä muista tähän kaavaan sopivista malleista ovat Nygrenin, Appelgrenin ja Hüttenrauchin (2012) malli ja Rogersin (2011) The Guardian -lehden datajournalismia kuvaava malli. Datan analysointi-, taustoitus- ja siivousvaiheiden järjestys ja vaiheiden

kuvauksen yksityiskohtaisuus saattavat kuitenkin vaihdella. Erittelen seuraavissa luvuissa näiden mallien perusteella tunnistamani datajournalistisen työprosessin vaiheita.

4.1 Datan hankinta

Datajournalistinen työ alkaa joko juttuideasta, jonka innoittamana etsitään sopivaa dataa, tai datasetistä, jota “kuulustellaan” juttuideoiden toivossa. Datan hankintavaihe määrittelee, onko työ datajournalismia vai ei. Datan hankintavaihe on tärkein vaihe myös siksi, että kaikki muut työn vaiheet rakentuvat sen päälle. Tähän vaiheeseen palataan, jos myöhemmin huomataan tarvittavan lisää dataa. (Bradshaw 2011a.)

Bradshaw listaa esimerkkejä erilaisista tavoista, joilla dataa voidaan hankkia.

1. Jokin organisaatio toimittaa datan suoraan.
2. Data löydetään hallinnon verkkosivuilta kehittyneillä etsintätekniikoilla.
3. Data hankitaan verkossa piilevistä tietokannoista ruudun raavinnan avulla. Ruudun raavinta tarkoittaa datan keräämistä verkkosivuilta siihen tarkoitettuja ohjelmia käyttäen.
4. Data hankitaan muuntamalla dokumentteja analysoitavaan muotoon.
5. Data hankitaan ohjelmointirajapintoja (Application Programming Interface, API) käyttäen.
6. Data kerätään käyttäen esimerkiksi kyselyitä tai muuta joukkoistamista.

Uutisorganisaatioilla on kaksi pääväylää datan hankkimiseen. Dataa joko hankitaan organisaatioilta, kuten valtion virastoilta lataamalla julkista dataa verkosta, tekemällä tietopyyntöjä tai ruudun raavinnalla, tai sitä kerätään itse eri lähteistä. On tärkeää saada data oikeassa, koneluettavassa muodossa, eikä esimerkiksi pdf-tiedostoina. Myös datan tarkkuuden on oltava riittävä: puuttuva metadata, eli datasta itsestään kertova data, voi vähentää datasettien luotettavuutta. (Aitamurto 2011, 10, 14.)

Data voi olla datajournalismin lähde tai se voi olla keino, jolla tarina kerrotaan. Kuten mikä tahansa työkalu, se muokkaa lopputulosta tiettyyn suuntaan. (Bradshaw 2012b.) Tulipa data mistä tahansa, se on aina tavalla tai toisella ihmisen kokoamaa ja sen syntymisen ja

jaottelun takana piilee useita ihmisten tekemiä päätöksiä, joten siihen on suhtauduttava samalla kriittisyydellä kuin muihinkin lähteisiin (Bradshaw 2013, 34). Esimerkiksi hallinnon tietokannat on koottu tietyn hallinnon alueen tiedontarpeisiin, jotka eivät välttämättä ole samat kuin toimittajan tiedontarpeet – lähtökohtaisesti kaikki data on siis liikaista (Morris 1999, 35).

CAR:in ja datajournalismin ongelma oli aiemmin datan vähäinen määrä ja vaikea saatavuus. Nykyäänkään kaikki data ei ole vapaasti saatavilla, uudeksi ongelmaksi on kuitenkin noussut se, että myös saatavilla olevaa dataa on valtava määrä, eivätkä toimittajat osaa hyödyntää sitä kaikkea. (Bounegru 2012.) Muun muassa taloudellisista syistä toimitukset pyrkivät yleensä käyttämään ilmaista dataa (Aitamurto 2011, 10), jotkin tahot saattavat kuitenkin pyytää maksua tietyn datan luovuttamisesta.

4.2 Datan käsittely

Kun data on hankittu, se on saatava käyttökelpoiseen ja mahdollisimman luotettavaan muotoon, eli se on puhdistettava. Datan puhdistaminen tarkoittaa yleensä ihmisestä johtuvien virheiden poistamista ja datan muuttamista muotoon, joka on yhteensopiva muun käytetyn datan kanssa. Erilaisia etsittäviä virheitä ovat esimerkiksi kaksoismerkinnät, tyhjät merkinnät, väärät muotoilut, kirjoitusvirheet ja puuttuva data. Datan puhdistamiseen sopivia työkaluja ovat esimerkiksi OpenRefine, Google Driven taulukkolaskentaohjelma ja Microsoftin taulukkolaskentaohjelma Excel. (Bradshaw 2011.) Jos data jää puutteelliseksi puhdistamisesta huolimatta, se on huomioitava analyysivaiheessa ja mahdollista virhemarginaalia arvioitaessa (Kuutti 2011, 609).

Dataa analysoidessaan toimittajan on oltava selvillä datan, kuten minkä tahansa muunkin lähteen taustoista. On syytä selvittää, kuka datan on kerännyt, milloin, miten ja mitä tarkoitusta varten. On myös tiedettävä, mitä datan mahdollisesti sisältämät termit, koodit ja ammattikieli tarkoittavat. Tässä vaiheessa tulee usein myös tarve löytää kontekstoivaa dataa, kuten vaikkapa vanhempaa rikosdataa uudemman tueksi, auttamaan mahdollisten uutisaiheiden, kuten muutoksen tai alueellisen vaihtelun havaitsemisessa. Erilaisten

datasettien yhdistäminen on vaikea työvaihe, jossa tilastotieteellinen osaaminen on tarpeen virheiden välttämiseksi. Datasettejä yhdistellään yhteisen datapisteen avulla, se voi olla esimerkiksi molemmissa dataseiteissä esiintyvä nimi, henkilötunnus, postinumerotieto, koordinaatit tai muu tunnistetieto. (Bradshaw 2011.) Myös datan visualisoiminen muuhun muotoon kuin taulukoksi auttaa usein huomaamaan uusia asioita analyysivaiheessa (Aich 2012).

Journalisteilla on datan käsittelyyn erilaiset motiivit kuin esimerkiksi tutkijoilla. Akateemisessa tai kaupallisessa ympäristössä työskentelevät tutkijat etsivät laajoista dataseiteistä tilastollisia trendejä ja malleja, joiden perusteella voidaan luoda vaikkapa uusia lainsäädännön suosituksia tai markkinointitapoja. Journalistit taas etsivät tarinan alkua tai epätavallisia asioita, jotka voivat johtaa juttuaiheiden luokse. Journalistien datan käsittelyllä on usein tarkat aika- ja resurssirajat, eikä yhteen projektiin käytetä kauempaa, kuin on pakko. (Cohen, Hamilton & Turner 2011.)

Bradshaw (2013, 5) listaa erilaisia tarinoita, joita dataseiteistä voidaan analyysin avulla löytää:

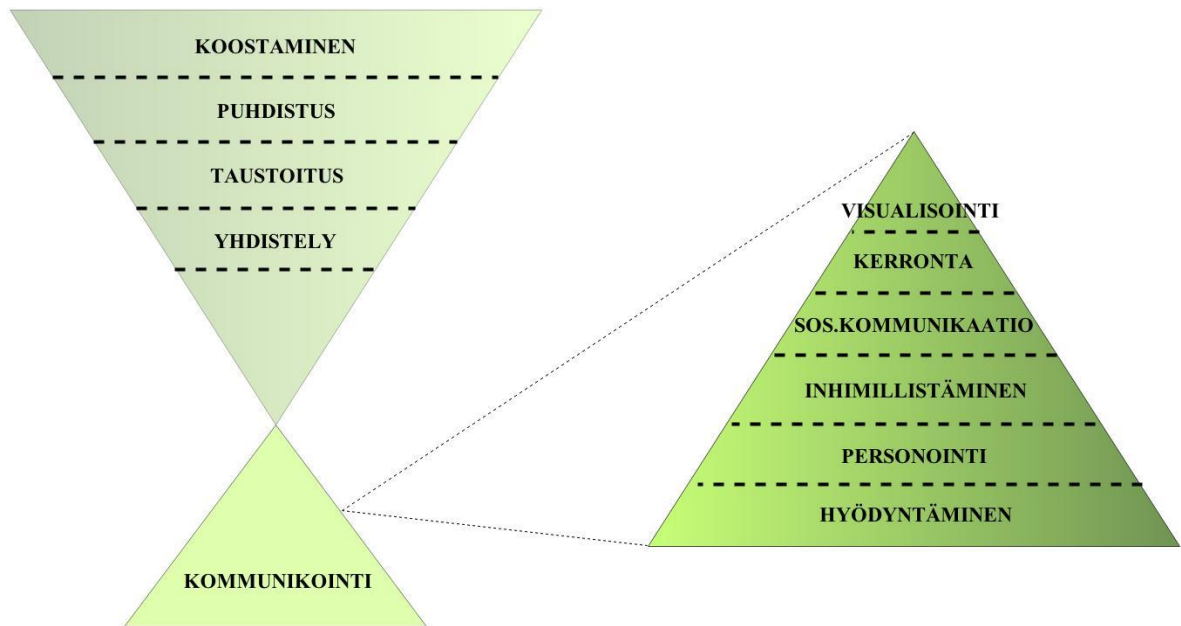
1. Mikä on verrattavista kohteista paras tai huonoin?
2. Mitä data kertoo meistä maana, alueena tai toimialana?
3. Paljastuuko datasta epätavallinen muutos tai sattumus? Mikä muuttui datassa tiettyinä aikavälinä ja miksi?
4. Tarkastetaan poliitikon tai muun tahon väitteiden paikkansapitävyys datasta.
5. Trendi: mikä nousee ja mikä laskee?
6. Kohdellaanko eri alueilla asuvia ihmisiä eri tavoin?
7. Onko datan kokoajalla jotain salattavaa? Onko data luotettavaa ja täydellistä? Yrittääkö joku estää ihmisiä näkemästä sen?
8. Mikä on datan vaikutus oikeassa elämässä? Millaisia ihmiskohtaloita ja tarinoita dataan sisältyy?
9. Asiat tapahtuvat yhdessä. Esimerkiksi jokin ominaisuus laskee, kun toinen nousee.
10. Onko jonkin tahon tekemällä muutoksella luvatonlainen vaikutus?

4.3 Lopputuotteen rakentaminen

Kun data on etsitty, puhdistettu, yhdistely ja kontekstoitu, rakennetaan sen avulla itse datajournalistinen lopputuote. Ehkä leimallisin ja jopa kliseeksi muodostunut datajournalistinen tuote on Googlen työkalujen avulla tehty karttavisualisointi. Lopputuote voi kuitenkin olla lähes mitä tahansa, kunhan sen taustalla on dataa. Tämän voi havaita esimerkiksi tarkastelemalla datajournalismipalkintojen kategorioita. Norjalainen innovaatioverkosto NxtMedia ja ruotsalaisten mediatalojen ja Södertörnin yliopiston datajournalismiprojekti palkitsivat Pohjoismaiden parhaat datajournalistiset työt ensimmäistä kertaa marraskuussa 2013. Nordic Data Journalism Awardsin lopputuotteen perusteella luokitellut kilpailukategoriat olivat:

1. *Tutkiva datajournalismi*. Työn täytyy olla yhteiskunnallisesti merkittävä. Sekä journalismin tason että käyttäjäkokemuksen on oltava korkeaa tasoa. Yhden tai useamman datasetin käytön on oltava työssä tärkeässä roolissa.
2. *Tarinan kertominen datan avulla*. Tarinankerrontaa, joka tarjoaa oivalluksia asettamalla datan kontekstiin. Sisältää tekstiä, grafiikkaa, videota, karttoja tai muuta innovatiivista datan esittämistä.
3. *Dataa käyttävät sovellukset*. Datan jakamista yleisön kanssa sovelluksella, joka tarjoaa oivalluksia sekä selkeän ja helppotajuisen navigaation, ja jota on helppo käyttää uudelleen ja jatkojalostaa.

Jako on selkeä, mutta kuitenkin niin laaja, että lähes jokainen kuviteltavissa oleva datajournalistinen tuote mahtuu johonkin kategoriaan. Myös Global Editors Networkin ja Googlen vuonna 2013 järjestämä Data Journalism Awards jakoi erilaiset lopputuotteet näihin samoihin kategorioihin (Filloux 2013).



KUVIO 5 Datajournalistisen lopputuotteen kommunikointi.

Paul Bradshaw (2011b) jakaa datajournalistisen työprosessin viimeisen vaiheen, löydösten ja aikaansaannosten kommunikoinnin, kuuteen osaan (kuvio 5).

1. *Visualisointi (Visualise)*. Visualisointivaiheessa datasta rakennetaan visuaalinen esitys, joka voi olla esimerkiksi grafiikka, video, interaktiivinen visualisointi tai uutissovellus. Tyypillisesti visualisoinnin työkaluina käytetään erilaisia piirto-ohjelmia, kuten Adobe Illustringia, koodikirjastoja, kuten JavaScriptin Chart.js:aa, tai helppokäyttöisiä visualisointisovelluksia, kuten Google Fusion Tablesia.

2. *Kerronta (Narrate)*. Perinteinen teksti ja tarina.

3. *Sosiaalinen kommunikaatio (Socialise)*. Datajournalismin levittämistä esimerkiksi jakamisen, kilpailun, kampanjoinnin ja yhteistyön kautta. Sosiaalisen median verkostojen käyttämistä ja joukkoistamista voidaan tehdä kaikissa datajournalismin vaiheissa.

4. *Inhimillistäminen (Humanise)*. Tapausesimerkkien ja tarinoiden etsiminen numeroiden takaa.

5. *Personointi (Personalise)*. Käyttäjälle tulee tarjota mahdollisuuksia tarkastella tietoa haluamiensa kriteerien, kuten vaikkapa postinumeroalueen, mukaisesti. Käyttäjälle voi antaa mahdollisuuden kokeilla, miten hänen omat ratkaisunsa vaikuttaisivat jonkin asian toteutumiseen, esimerkiksi, kuinka käyttäjä leikkaisi valtion budjettia ja mitä se aiheuttaisi.

6. *Hyödyntäminen (Utilise)*. Dataa käyttäen voi luoda erilaisia käyttäjää hyödyttäviä työkaluja ja sovelluksia, kuten laskureita. Näiden rinnalla voidaan käyttää muita tarinankerronnan muotoja.

Bradshaw korostaa, että datan kommunikoinnissa visualisointi on niin vallalla oleva toimintatapa, että se saattaa saada unohtamaan muut tavat, joilla datatarinoita voi kertoa ja levittää. Pelkkä datajournalistisen työn julkaisu ei kuitenkaan johda parhaaseen saavuttavuuteen.

5 DATAJOURNALISMI JA AVOIMUUS

Yleisesti ollaan sitä mieltä, että sekä journalismissa että tieteessä on oltava avoin ja annettava muidenkin toistaa sama tutkimus, jotta voitaisiin varmistua sen paikkaansa pitävyydestä (esim. Meyer 1991). Toimitukset julkaisevat datajournalismissa käyttämiään datasettejä pääasiassa kahdesta syystä. Ensimmäinen syy on läpinäkyvyys: kun data on kaikkien saatavilla, kuka tahansa voi testata toimituksen tekemää analyysia ja varmistua siitä, että tulos on oikea. Toinen syy on osallistaminen: datan julkaisulla toimitus kutsuu yleisön tutkimaan dataa ja tuottamaan uutta tietoa sen pohjalta. Näin datasta voi olla yhteiskunnalle laajempaa hyötyä, kuin mitä se pelkän toimituksen käytössä tuottaisi. (Aitamurto 2011, 11; Nygren, Appelgren & Hüttenrauch 2012, 85.) Bradshaw (2013, 40) tiivistää avoimuusajattelun osuvasti: ”Mitä avoimempia toimitukset ovat resurssiensa jakamisessa, sitä todennäköisemmin myös muut tahot ovat avoimia omien taitojensa ja aikansa antamisessa.”

Avoimuudella on myös liiketoiminnallinen ulottuvuus: muutamat uutisorganisaatiot, kuten BBC ja NPR ovat jo muutamien vuosien ajan jakaneet joitakin sisältöjään, kuten sää- ja uutissyötteitä, avoimien ohjelmointirajapintojen (Open API) kautta. Aitamurto ja Lewis (2012) nostavat esiin Henry Chesbrough'n avoimen palveluinnovaation ajattelun, jonka mukaan yrityksen ulkopuolisille toimijoille ja käyttäjäyhteisölle jaettu tieto kasvattaa kokeiluiden ja siitä syntyvien innovaatioiden määrää tasolle, jota pelkkä yrityksen sisäinen kehitystyö ei voi saavuttaa (Chesbrough 2010, 24). Avoimen innovaation odotetaan yleistyvän tulevaisuudessa. Tähän mennessä avoimien ohjelmointirajapintojen käyttö ei kuitenkaan ole tuonut uutisorganisaatioille merkittäviä taloudellisia hyötyjä. (Aitamurto & Lewis 2012, 328.)

Aitamurron (2011) tutkimissa uutisorganisaatioissa on määritetty käytäntöjä, joita onnistuneessa datajournalistisessa työprosessissa tulisi noudattaa. Näitä ovat joustavuus ja nopea analyysi, kaiken raakadatan tarjoaminen kätevästi saataville, saman datan käyttö moneen projektiin, jos se on mahdollista, ja riittävän metadatan lisääminen julkaistuun

dataan. Olennaista on myös luoda yleisölle useita osallistumiskanavia, joiden kautta he pääsevät mukaan datajournalismin tekemiseen.

Ajatus datajournalistisissa töissä käytetyn datan jakamisesta yleisön kanssa herättää kysymyksen siitä, eikö toimittajan tehtävä ole etsiä tietomassasta yleisön kannalta olennaiset asiat, eikä vain asettaa näytille raakamateriaalia. Avoimuusajattelun lisääntymisen voi liittää digitaalisen median mukanaan tuomaan muutokseen yleisön asemassa. Teknologian mahdollisuuksien kasvaessa ainakin osa yleisön jäsenistä muuttuu pelkistä passiivisista vastaanottajista toimiviksi lähettäjä-vastaanottajiksi (Coldry 2009, 438). Merrinin (2009, 24) mukaan koko termi yleisö on menettänyt arvoaan liian passiivisena. Termin merkitys vähenee, koska yleisönä olo on nykyään paitsi mediatuotteiden vastaanottoa ja niihin reagoimista, myös arvostelua, kommentointia, jakamista ja luomista. Pietilän ja Ridellin (2008, 29) mukaan yleisö on kuitenkin yhä tarpeellinen käsite verkkomedian tutkimuksessa, kunhan yleisöyys ymmärretään rooliksi, jossa vastaanottotilanteessa toimitaan, mutta siitä voidaan siirtyä myös muihin, kuten tuottajan rooleihin.

Parasien ja Dagiralin (2013, 854) haastattelemien datajournalistien käsityksissä yleisö on aktiivinen ja osallistuva ja pystyy löytämään datasta merkityksiä myös ilman toimittajan apua. Heidän mielestään kaikki data tulisikin luovuttaa rajoituksitta yleisön käyttöön. Vaikka datan luovuttaminen on ollut verkkojournalistien pohdinnan aihe aina 1990-luvulta asti, tämä uusi ajatus avoimesta datan luovuttamisesta haastaa voimakkaasti perinteisen ajattelun, jossa journalistin tehtävä on esittää yleisölleen tulkintoja alati monimutkaistuvasta maailmasta. Datan luovuttamista on kritisoitu myös toteamalla, ettei suurimmalla osalla yleisöä ole taitoja datan tulkitsemiseen ja että data itsessään ei aina välttämättä ole täydellistä ja manipuloimatonta. (Parasie ja Dagiral 2013 867 - 868.) The Guardianin toimittaja Simon Rogers huomauttaa Paul Bradshaw'n (2012, 17) haastattelussa, että asia ei aina välttämättä ole näinkään:

Sen sijaan että vain heitettäisiin se (data) ”kiitolliselle yleisölle” on ymmärrettävä, että suurimmassa osassa tapauksia yleisössä on asiantuntijoita, jotka luultavasti osaavat analysoida tiettyjä datasettejä paremmin kuin sinä.

Vuonna 2010 The Guardian joukkoisti brittikansanedustajien kulukuittien tutkimisen yleisölleen. Tätä projektia tutkineet Daniel ja Flew (2010) korostavat myös, että datan jakaminen yleisön kanssa on enemmänkin tiedon hallinnointia kuin journalismia. Journalismia syntyy vasta, kun yleisön avulla tuotettuun tai suodatettuun tietoon sovelletaan journalistista etiikkaa, päätöksentekoa ja faktantarkistusta.

Bradshaw'n (2014, 40) mukaan journalismissa on meneillään kulttuurinen kamppailu juttujen omistajuuteen uskoviin toimittajien ja verkon yhteisöjen avoimen yhteistyön kulttuurin välillä. Datajournalistit hyödyntävät usein työssään avointen verkkoyhteisöjen tietoa ja teknistä osaamista. Datajournalistien omat lähtökohdat vaikuttavat muutenkin avoimuuden arvostukseen. Parasien ja Dagiralin (2013) haastattelemien datajournalistien juuret olivat avointa dataa ja avointa demokratiaa kannattavissa liikkeissä. Avoimuus ja jakaminen kuuluvatkin perustavasti verkon luonteeseen. Bounegrin (2012) mukaan digitaalinen teknologia ja verkko muuttavat informaation julkaisun tapaa. Datajournalismi on yksi osa työkalujen ja toimintatapojen ekosysteemiä, joka on syntynyt datasivustojen ja palvelujen ympärille. Lainaaminen ja lähdemateriaalin jakaminen on ominaista verkon hyperlinkkirakenteessa ja kuuluu nykyisiin tapoihimme luovia informaation keskellä. Hyperlinkkirakenteen perusteet taas kumpuavat akateemisen maailman viittauskäytännöistä. (Bounegrin 2012.)

6 TUTKIMUKSEN KUVAUS JA METODIT

Tässä luvussa kuvaan tutkimukseni etenemistä, käyttämiäni metodeja ja niiden taustalla piileviä ajatustapoja ja periaatteita. Luvun rakenteen innoittaja on John W. Creswell (2013, 5), joka korostaa, että tutkimusta suunniteltaessa on pohdittava filosofisen maailmankatsomuksen mukanaan tuomia oletuksia, maailmankatsomukseen liittyvää tutkimussuunnitelmaa ja metodeja, jotka tuovat tutkimussuunnitelman määrittämän lähestymistavan käytäntöön. Tämä pohdinta lisää tutkimuksen läpinäkyvyyttä ja luotettavuutta.

6.1 Tutkimuskysymykset ja tutkimuksen tarkoitus

Tutkimukseni tarkoitus on tuottaa ajantasaista tietoa datajournalistisesta työprosessista Ylessä ja Helsingin Sanomissa. Samalla kartoitan datajournalismin tekijöiden työssään kohtaamia haasteita. Lisäksi luotaan haastattelu- ja havainnointiaineiston ja kirjallisuuden avulla datajournalismin tulevaisuutta.

Tutkimuskysymykseni ovat:

- Millainen on datajournalistinen työprosessi Helsingin Sanomien, Svenska Ylen ja Ylen uutis- ja ajankohtaistoiminnan toimituksissa?
- Millaisia haasteita toimitukset kohtaavat datajournalismin tuottamisessa?
- Kuinka datajournalismia voidaan toimituksissa tulevaisuudessa edistää?

Laadullisessa tutkimuksessani ei ole varsinaista hypoteesia, mutta kirjallisuuden perusteella olen muodostanut joitakin esioletuksia datajournalistin työstä. Käsitykseni datajournalismista ja sen työprosessin jakautumisesta kuvasin luvuissa 2 (s. 5) ja 4 (s. 20). Toinen laajempi esioletukseni on se, että datajournalismiin läheisesti liittyvä teknologian kulttuuri tuo journalismin kulttuuriin erilaisia ajatuksia objektiivisuudesta, läpinäkyvyydestä, tiedon ja taustamateriaalin jakamisesta ja työtavoista (Parasie & Dagiral 2013; Royal 2010). Käsittelin datajournalismin kulttuuriin liittyvää avoimuusajattelua

luvussa 5 (s. 31). Uskon, että nämä ajatukset vaikuttavat datajournalismin tekijöiden toimintaan toimituksissa ja perinteisessä journalismin kulttuurissa.

6.2 Tutkimusprosessi ja sen lähtökohdat

Datajournalismin nousu Suomen mediassa on tapahtunut samoina vuosina, joina itse aloitin matkani koulun penkiltä työelämään. Tutkimukseni lähtökohta oli datajournalismia kohtaan tuntemani kiinnostus, joka heräsi, kun seurasin Ylen ja Helsingin Sanomien datajournalistien jatkuvasti kehittyviä tuotoksia.

Tuo kiinnostus johti tutkimusaiheen valintaan ja ohjasi alan kirjallisuuden pariin. Tutustuminen tietokoneavusteisen journalismin historiaan, datajournalismin tutkimukseen ja erilaisiin työnkuvauksiin sekä datajournalistien kirjoittamiin oppaisiin ja blogeihin auttoi minua hahmottamaan tutkimuskysymykseni, valitsemaan tutkimukseni kohteet ja metodit, joilla saisin mahdollisimman tehokkaasti hankittua kokemuksia ja tietoa datajournalistin työstä.

Datajournalismin työtapoja on tutkittu melko vähän ja tutkimukset painottuvat Yhdysvaltain suuriin toimituksiin. Suomen toimituksissa ja toimituskulttuurissa datajournalismia ei ole muutamaa opinnäytetyötä (ks. s. 3) lukuun ottamatta tutkittu lainkaan. Siksi perinteisen tieteellisen kirjallisuuden rooli lähdemateriaalissani jäikin melko pieneksi. Tutkimukseni muotoutuikin laajaksi ja perustietoa tuottavaksi juuri siksi, että sille oli selkeästi tarvetta. Tilanteessa, jossa uutta ja vähän tutkittua ilmiötä täytyy tutkia ja ymmärtää, laadullinen tutkimus on perusteltu lähestymistapa (Creswell 2013, 20). Oma taustani toimittajana ja tieteenalani käytännönläheinen painotus puoltavat myös laadullista lähestymistä, koska se tarjoaa metodivalikoiman, jolla suomalaisesta datajournalismista on mahdollista tuottaa sekä käytännön journalistista työtä hyödyttävää että jatkotutkimuksen suuntia ohjaavaa tietoa.

Metodikirjallisuuteen ja valitsemieni metodien, etnografisen havainnoinnin ja haastattelun, taustoihin perehtyminen ohjasi pohdintaani tutkimukseni tieteenfilosofisista lähtökohdista.

Tutkimukseni tavoitteeksi oli muodostunut tuottaa tietoa datajournalistisesta työprosessista ja siinä koetuista haasteista, joten en nähnyt tarpeelliseksi ottaa tiettyä teoriaa ohjaamaan tutkimustani.

Pertti Alasuutarin (2011, 261) mukaan teoreettinen viitekehys tulee erottaa asioiden paikallisesta selittämisestä, eli siitä, että empiirisestä aineistosta löydettyjä ilmiöitä tehdään ymmärrettäviksi. Tutkimuksessani pyrin teorian muodostamisen sijasta kuvaamaan datajournalistien työtä tietyissä toimituksissa, tietyllä hetkellä sekä tuomaan esille heidän näkemyksiään työn tulevaisuudesta. Koostin jokaisen toimituksen työprosessista ja siinä ilmenevistä haasteista omat luvut (alkaen s. 48) ja esittelin jokaiselta toimitukselta yhden datajournalistisen jutun konkretisoidakseni työprosessia. Pohdinta-luvussa (s. 77) esittelen laajempaa yhteenvetoa tuloksistani ja niiden perusteella tekemiäni havaintoja datajournalistisesta työprosessista ja sen haasteista. Aineistoa analysoidessani huomasin, että kirjallisuudesta tutut ajatukset datajournalismin tulevaisuudesta ja ihanteista toistuivat myös suomalaisten datajournalistien haastatteluissa, joten päädyin rakentamaan Tulevaisuus-luvun käyttäen aineistoani ja vieden pohdintaa pidemmälle kirjallisuuden avulla.

6.3 Tieteenfilosofinen asemointi

John W. Creswell (2013, 5 - 6) kuvaa maailmankatsomusta yleisenä filosofisena suhtautumisena maailmaan ja tutkimuksen luonteeseen. Tuo suhtautuminen sisältää epistemologiset ja ontologiset käsitykset. Vaikka tutkimukseni tarkoitus onkin kuvata käytännönläheisesti datajournalistista työprosessia, on tärkeää tunnistaa maailmankatsomuksen asettamat esioletukset, koska tutkija tuo aina maailmankatsomuksensa mukaan tutkimukseensa.

Laadullinen tutkimus on lähestymistapa, jolla selvitetään ja ymmärretään ihmisten ja ihmisryhmien luomia merkityksiä, joita he liittävät yhteiskunnallisiin ja inhimillisiin ongelmiin. Se nojaa induktioon, yksilöllisen merkityksen korostamiseen ja tilanteiden monitulkintaisuuden hahmottamiseen. (Creswell 2013, 4.) Nämä merkitykset ilmenevät

esimerkiksi ihmisen tekoina ja päämäärien asettamisina, suunnitelmina, hallinnollisina rakenteina, yhteisöjen toimina ja päämäärinä (Varto 1992, 24).

Konstruktivistinen maailmankuva tukee näitä laadullisen tutkimuksen tavoitteita, koska sen mukaan ihmisten sosiaaliset ja historialliset perspektiivit ohjaavat merkitysten luomista, eli tutkimuksen tapauksessa aineiston keräämistä ja tulkintaa. Laadullista tutkimusta tekevä tutkija voi pyrkiä ymmärtämään tutkimuksen osallistujien ympäristöä ja heidän toimintansa kontekstia esimerkiksi vierailemalla heidän luonaan ja keräämällä tietoa henkilökohtaisesti. Tutkijan päämäärä on siis tulkita merkityksiä, joita muilla on maailmasta. (Creswell 2013, 8 - 9.) Ihmistä tutkiva ei voi koskaan olla ulkoinen tarkkailija, koska tutkija on aina osa tutkimaansa merkitysyhteyttä, hänen tapansa ymmärtää tutkimansa kysymykset vaikuttaa koko ajan tutkimuksen taustalla (Varto 1992, 26).

Edellä mainitut näkemykset johtavat ajatukseen, ettei tutkimukseni lopputulosta voi yleistää koskemaan muita tutkimuskohteita. Varto (1992, 39) summaakin, että laadullisessa, ideografisessa tutkimuksessa pyritään selittämään yksittäistä tapahtumaketjua tai ilmiötä, eikä ajatella, että sen perusteella luotu selitys olisi yleistettävä.

Konstruktivismiin asettamat taustaoletukset huomioon ottaen suhtaudun aineistooni Pertti Alasuutarin (2011) ajattelun mukaisesti faktanäkökulmasta. Faktanäkökulma ohjaa tutkijaa käyttämään kapeaa käsitystä merkityksellisestä tiedosta: aineistoja tutkiessaan tutkija haluaa tavoittaa arkiajattelun mukaisen käsityksen todellisuudesta. Tutkijaa kiinnostaa tutkittavien todellinen käyttäytyminen ja tutkittavien haastatteluista koostuvan aineiston käyttökelpoisuus riippuu siten heidän rehellisyydestään. (Alasuutari 2011, 90 - 91.) Tutkimuksessani tämä ei kuitenkaan tarkoita positivistista suhtautumista maailmaan, vaan ainoastaan ohjaa minua aineiston hankinnassa ja käsittelyssä. Faktuaalinen luenta ei ole naiivin realistista, vaan materiaallinen maailma ja muut toimijat, eli tutkijat sekä haastateltavat, pyrkivät estämään minkä tahansa esittämisen faktana (Alastalo & Åkerman 2010, 377).

Konstruktivistinen ajattelu opastaa tutkijaa skeptisyyteen valmiita tutkimushahmotelmia kohtaan ja kannustaa esimerkiksi haastattelemaan avoimin kysymyksin, jotta

merkityksellistämisen prosessia olisi helpointa hahmottaa (Warren 2001). Varsinkin etnografiassa on tärkeää lähestyä aineistoaan mahdollisimman avoimesti. Kuitenkaan yksikään laadullinen tutkimus ei ala pelkästä havainnoinnista, vaan kaikki alkaa aiemman tutkimuksen läpikäymisestä ja metodeihin tutustumisesta. (Alasuutari 2011, 262, 66.) Taustalla vaikuttavat myös sekä tutkijan että tutkittavan ajatukset, kokemukset ja ominaisuudet (Warren 2001).

6.4 Metodivalinnan perustelu

Tutkimusmetodi tarkoittaa niitä käytäntöjä ja operaatioita, joiden avulla havaintoja tuotetaan sekä niitä sääntöjä, joita noudattaen tuotettuja havaintoja tulkitaan ja muokataan (Alasuutari 2011, 82). Käyttämäni tutkimustapa on etnografia, jossa aineiston hankkimisen keinoina toimivat havainnointi ja haastattelut.

Etnografia on antropologiasta ja sosiologiasta kumpuava tutkimuksen tapa, jossa tutkija selvittää yhtenäisen kulttuurisen ryhmän jakamia käytöksen, kielen ja toiminnan malleja heidän luonnollisessa ympäristössään pitkän ajan kuluessa (Creswell 2013, 14). Patersonin (2008) mukaan etnografinen tutkimus on paras lääke itsestänselvyyksien kyseenalaistamiseen. Uutistuotannon tutkimuksessa etnografiset metodit antavat perustietoa kulttuuristen tuottajien työideologioista ja käytännöistä ja tarjoavat mahdollisuuden tarkkailla kulttuurin tuotannon perustalla olevia sosiaalisia käytäntöjä teorian informoimana (Schlesinger 1980, 363, Patersonin 2008, 2 mukaan).

Verkkjournalismia tutkineen Domingon (2005) mukaan tutkijat ovat verkon ja sen tutkimuksen kehittyttyäkin säilyttäneet fiksaation sisältöön ja tutkimuskirjallisuuden suurin osa käsittelee sisältöä ja sen vastaanottoa, mutta vain pieni osa tuotannon analyysia ja vielä harvempi tuotannon analyysia etnografisia metodeita käyttäen. Domingon (2003) listaus (kuvio 6) selvittää etnografian etuja ja heikkouksia.

Etnografisen metodin edut ja heikkoudet

Edut

Tutkija voi koota valtavan määrän rikasta ensikäden dataa

Tutkija voi havainnoida sosiaalisten suhteiden ja teknologian toimintoja, rutiineita ja määritelmiä

Tutkija voi saavuttaa tutkittavien luottamuksen ja saada siten ”sisäpiiriläisten” näkökulmia

Tutkija voi tarkkailla konflikteja ja evoluutioprosesseja

Heikkoudet

Havainnointi vie aikaa ja tutkijan läsnäolo saattaa häiritä tutkittavia

Ei ole helppoa merkitä muistiin kaikkea näkemäänsä, erityisesti tekniset toiminnot ovat vaikeita

Tutkittavat voivat pyytää olemaan siteeraamatta tutkijalle tekemäänsä tunnustusta

Tuloksia ei saa yleistää suoraan. On riski luulla satunnaista mainintaa tai puhetta säännöksi

Tutkijan on oltava tietoinen ennakkoluuloistaan välttääkseen niiden vaikutuksen tutkimukseen

KUVIO 6 Etnografisen metodin edut ja heikkoudet (Domingo 2003).

Etnografisessa tutkimuksessa on tavallista, että tutkittava ympäristö ja sen ihmiset ovat tutkijalle tuttuja (Warren 2001). Toimitusten työrutiinit, sekä osa niissä käytetyistä järjestelmistä ja tutkimistani henkilöistä ovat minulle ennalta tuttuja, mutta katson sen vain auttavan minua kohdistamaan huomioni tutkimuksen kannalta oikeisiin asioihin. Puijikin (2008, 40) mukaan tutkimuksen tekeminen digitaalisessa ympäristössä vaatii usein sitä, että tutkija tuntee tietokoneet edistyneellä tasolla ja joskus jopa sitä, että hän tietää, millaisia ohjelmia toimittajat käyttävät. Vaikka tutkimukseni ympäristö ei olekaan suoraan digitaalinen, verkkomaailma ja tietotekniikka liittyvät kiinteästi datajournalistien työhön. Taustatieto kentästä siis on paitsi hyväksi, monesti myös välttämätöntä.

On yleistä yhdistää havainnoinnilla hankittua aineistoa haastatteluaineistoon ja esimerkiksi dokumenteista hankittuun tietoon (Warren 2001). Useamman metodin käyttöä kutsutaan metodiseksi triangulaatioksi (Singer 2008, 165; Tuomi & Sarajärvi 2009, 145).

Triangulaatio voi auttaa suojautumaan yhden metodin käytön aiheuttamilta mahdollisilta vääristymiltä (Singer 2008, 166). Haastattelun ja havainnoinnin yhdistelmä siis auttaa ylittämään kummankin metodin heikkouksia. Haastattelussa säännöt ja normit korostuvat, ja havainnointi tuo esille niihin liittyvän käyttäytymisen. (Tuomi & Sarajärvi 2009, 81; Grönfors 2001, 127.)

6.4.1 Haastattelu

Osana etnografista tutkimusstrategiaani käytän puolistrukturoitua teemahaastattelua, joka metodina sopii tiedonhankintaan siitä, miten tutkittavat kokevat toimintansa datajournalistisen työprosessin eri vaiheissa. Tutkimusongelma sitoo haastattelun teemat yhteen (Eskola & Vastamäki 2001, 33). Rakensin haastattelurungon (liite) tutkimuskysymysteni ympärille, eli siinä on omat useita kysymyksiä sisältävät teemaosuudet eri datajournalismin työvaiheille, työn haasteille, ja haastateltavien näkemyksille tulevaisuudesta. Lisäksi teemoihin kuuluu joukko yleisiä kysymyksiä toimitusten datajournalismin synnystä ja nykytilasta sekä haastateltavien taustasta.

Teemahaastattelu on keskustelu, joka tapahtuu tutkijan aloitteesta ja hänen ehdoillaan. Siinä tutkija yrittää selvittää haastateltavilta häntä kiinnostavia asioita. Teemahaastattelun alueet, eli teemat, on määritelty ennen haastattelutilannetta käyttäen apuna esimerkiksi kirjallisuutta tai aikaisempaa tutkimusta. Haastatteliija voi kuitenkin käsitellä asioita tilanteen vaatimassa järjestyksessä. (Eskola & Vastamäki 2001, 24, 26.) Haastattelun etu tiedonhankintametodina onkin ennen kaikkea joustavuus: mahdollisuus kysyä uudestaan, oikaista väärin ymmärrettyjä kysymyksiä ja vaihtaa kysymysten suunniteltua järjestystä (Tuomi & Sarajärvi 2009, 73). Haastatteluja tehdessäni esimerkiksi jätin väliin kysymyksiä, jos haastateltavat vastasivat niihin jo aiempien kysymysten yhteydessä. Ylen Plus-deskistä ja Svenska Ylen Grävnavetista haastattelin molemmista sekä toimittajaa että tuottajaa, joten muokkasinkin haastattelurunkoja vähentääkseni päällekkäisyyttä ja saadakseni haastateltavien

vastattavaksi juuri ne kysymykset, joihin uskoin heidän parhaiten pystyvän vastaamaan. Helsingin Sanomissa sama toimittaja työskentelee datajournalistina ja tuottajana, joten katsoin, että hänellä on paras näkemys koko deskin toiminnasta.

Haastattelulla hankittavaan aineistoon vaikuttavat haastattelijan ja haastateltavan perspektiivit, joita muodostavat yhteiskunnalliset ja haastattelutilanteessa syntyvät hierarkiat ja sosiaaliset positiot (Warren 2001). Tutkimukseni haastatteluissa näitä perspektiivejä muodostivat varmasti ainakin haastattelijan ja haastateltavien työkokemuksen, teknisen osaamisen ja ammatillisen statuksen erot.

6.4.2 Havainnointi

Toinen osa etnografista tiedonhankintaani on havainnointi ilman osallistumista. Jensenin (2002, 242) mukaan havainnointi tarkoittaa tutkimustapojen joukkoa, johon yleensä liittyy tutkijan pitkäaikainen läsnäolo yhdessä paikassa. Eläytyvä havainnoija on tutkimuksen keskeinen instrumentti, jonka tiedonhankinta nojaa useisiin erilaisiin aistihavaintoihin ja median kautta saatuun tietoon. Havainnointi kytkee tiedon oikeaan kontekstiin muita menetelmiä paremmin (Grönfors 2001, 127; Tuomi & Sarajärvi 2009, 81).

Havainnointia voidaan tehdä tarkoituksenmukaisuusperiaatetta noudattaen, eli kiinnittäen huomiota asioihin, joiden katsotaan kuuluvan tutkimuksen aihepiiriin (Grönfors 2001, 136). Olenaiseen huomion kiinnittämisessä auttaa huolellinen alkuvalmistautuminen kirjallisuuden, keskustelujen ja vapaan alkuhavainnoinnin avulla (Aarnos 2001, 146). Havainnoinnin dokumentaation on oltava perusteellista ja itsereflektioivaa, koska muuten on vaarana, että tutkija tekee liian pitkälle meneviä tulkintoja (Bruhn Jensen 2002, 242).

Jensen (2002, 243) jakaa havainnoinnista tehtävät kenttämuistiinpanot kolmeen osaan.

1. *Ensisijaiset muistiinpanot* (substantive notes), jotka dokumentoivat tutkittavan paikan tai kohteen representaatioita. Vastaavat kysymykseen “mitä?”.
2. *Logistiset muistiinpanot* (logistical notes), jotka antavat lisätietoa tilanteesta, jossa data kerättiin. Vastaavat kysymykseen “miten?”.

3. *Reflektiiviset muistiinpanot* (reflexive notes), jotka käynnistävät analyysiprosessin ja teoretisoinnin havaintojen ja muun aineiston perusteella. Vastaavat kysymykseen “miksi?”.

Seurasin datajournalistien työtä yrittäen itse häiritä heidän toimintaansa mahdollisimman vähän. Juttelin tutkittavieni kanssa ja tarvittaessa kysyin, mitä he tekevät, mutta muuten pyrin vaikuttamaan heidän työhönsä mahdollisimman vähän. Bruhn Jensen (2002, 242) kuitenkin muistuttaa, että havainnoijan on mahdotonta olla täysin ulkopuolinen: hän osallistuu joka tapauksessa aineiston syntyyn olemalla paikalla havainnointitilanteessa ja tulkitsemalla tapahtumia omalla tavallaan. Grönforsin (2001, 131) mukaan osallistumaton havainnointi on perusteltua esimerkiksi, kun osallistuminen ei toisi tutkimukseen merkittävää lisänäkökulmaa tai sisältöä. Tässä tutkimuksessa tutkijan tekninen osaaminen ei riittäisi osallistumaan työhön häiritsemättä merkittävästi tutkittavien omaa työskentelyä. Tavallisen uutistoimittajan roolissa osallistuminenkin veisi tutkijan keskittymisen pois tutkittavien seuraamisesta.

Aineistonkeruuvaiheessa kaikissa kolmessa toimituksessa valmistauduttiin joululoman alkamiseen, joten en päässyt havainnoimaan aivan tavanomaisia työpäiviä. Helsingin Sanomien datadeskissä ja Yle Uutisten Plus-deskissä tehtiin uutissovelluksia ja erikoistaittoja loman aikana julkaistavaksi. Helsingin Sanomissa pääsin havainnoimaan ehkä tyypillisintä työntekoa, koska samana päivänä julkaistavia juttuja ja visualisointeja tehtiin vielä melko paljon. Havainnointipäivieni aikana suuria juttupaketteja tyypillisesti tekevä Svenska Ylen Grävnavet-ryhmä valmistautui joululoman viettoon ja käsitteli edellisen suuren projektinsa loppuraporttia, joten en päässyt juurikaan havainnoimaan heidän arkityöskentelyään. Siksi keskustelu ja erilaiset pienemmät haastattelut sekä vanhempien sovellusten ja visualisointien esittely saivat tiedonhankinnassa suuremman roolin.

6.5 Tutkittavien valinta ja aineistonkeruun järjestelyt

Ilmiön kuvaamiseen pyrkivässä laadullisessa tutkimuksessa tiedonantajien valinnan on palveltava tutkimuksen tarkoitusta, eli tässä tapauksessa heidän on tärkeää tietää

tutkittavasta ilmiöstä mahdollisimman paljon (Tuomi & Sarajärvi 2009, 85 - 86).

Tutkittavien valinnassa painotin sitä, että heidän tulisi tietää mahdollisimman paljon datajournalismista omassa toimituksessaan, datajournalismin työtavoista ja nykytilasta sekä kehityksestä. Suomen datajournalismikenttä on edelleen pieni, joten tutkimuskohteiden valinta ei tuottanut vaikeuksia.

Helsingin Sanomat on nykymuotoista datajournalismia kauimmin tehnyt mediatalo, jonka uutispäällikkö Esa Mäkinen on alan pioneeri ja siten tärkeä haastateltava. Toisen suuren uutistoimijan, Ylen, Plus-deskin perustaminen taas oli näkyvä panostus datajournalismiin ja uudelleen verkkosisällön tekemiseen, ja Ylellä jo aiemmin datajournalismin parissa työskennellyt Teemo Tebest on varmasti paras ihminen kertomaan, miten työtä toimituksessa tehdään. Svenska Ylen datajournalisti Peter Sjöholmiin tutustuin marraskuussa 2013 Södertörnin yliopistolla Huddingessa järjestetyssä datajournalismikonferenssissa, ja kun kuulin toimituksen suuresta datajournalismipanostuksesta, Grävnavet-ryhmän perustamisesta, päätin sisällyttää myös heidät tutkimukseeni. Plus-deskin tuottaja Juho Salmisen ja Grävnavetin tuottaja Eva-Maria Koskisen haastattelemisesta päätin, kun toimitusten työtä havainnoidessani huomasin, että heillä on arvokasta tietoa muun muassa datajournalismiryhmien yhteistyöstä muun toimituksen kanssa. Kyseessä oli siis lumipallo-otanta, jossa avainhenkilö johdatti minut uusien haastateltavien luo (Tuomi & Sarajärvi 2009, 85-86).

Viiden haastateltavan aineisto on melko pieni, varsinkin, kun tutkimuksessa on tarkoitus kuvata myös jokaisen toimituksen työprosessia erikseen. Haastattelun ja havainnoinnin yhdistelmällä ja eniten asiasta tietävien haastateltavien valinnalla kuitenkin sain hankittua datajournalistin työstä niin paljon tietoa, kuin pro gradu -tutkielmani rajoissa oli mahdollista. Haastateltavien vähäinen määrä ja toimitusten datajournalismiryhmien eroavuudet tekivät haastateltavien anonymiteetin säilyttämisestä mahdotonta, varsinkin, kun tutkimukseni tarkoitus oli myös kuvata datajournalistista työprosessia kussakin toimituksessa. Haastatteluiden ei aina tarvitsekaan olla anonymoituja, mistä esimerkki ovat tässäkin tutkimuksessa käytetyt asiantuntijahaastattelut. Asiantuntijat tunnistavat toisensa

helposti ja lisäksi rajatulla asiantuntijakentällä haastateltavien anonymisoiminen on harvoin mahdollista niin, ettei asiasisältöä katoa (Alasalo & Åkerman, 383).

Kun olin syksyn aikana tutustunut riittävästi alan kirjallisuuteen, otin haastateltaviin yhteyttä sähköpostilla marraskuussa 2013 ja sovin alustavasti havainnointijaksoista. Tein kaikissa toimituksissa kahden päivän havainnointijakson joulukuussa 2013. Havainnointien yhteydessä tein valitsemilleni viidelle tiedonantajalle puolistrukturoidun teemahaastattelun. Havainnoidessani kirjoitin muistiin myös käymiämme keskusteluja ja pyysin tutkittaviani esittelemään ja kuvailemaan tekemiään asioita.

6.6 Aineisto

Viidestä haastattelusta kertyi nauhoitettua materiaalia yhteensä noin neljä ja puoli tuntia. Esa Mäkisen haastattelusta osa jäi nauhoittamatta, koska nauhurin muistikortti täyttyi kesken haastattelun. Tämän vuoksi käytin Helsingin Sanomien datajournalismin kuvauksessa myös Datajournalismin työkäytännöt -tutkimuksen tutkijoiden Heikki Kuutin ja Turo Uskalin aiemmin tekemiä Mäkisen haastatteluja, joissa käsiteltiin myös minun tutkimiani teemoja.

Auki kirjoitettuna koko aineistoni käsitti:

- 19 sivua Helsingin Sanomien Esa Mäkisen haastatteluja, joista itse tekemääni oli noin neljä sivua, Turo Uskalin haastattelua 12 sivua, ja Heikki Kuutin haastattelua noin kolme sivua.
- Kahdeksan sivua Helsingin Sanomien havainnointimuistiinpanoja.
- Seitsemän sivua Svenska Ylen Peter Sjöholmin haastattelua ja kuusi sivua Eva-Maria Koskisen haastattelua.
- Yhdeksän sivua Svenska Ylen havainnointimuistiinpanoja.
- 13 sivua Ylen Teemo Tebestin haastattelua ja yhdeksän sivua Juho Salmisen haastattelua.
- Seitsemän sivua Ylen Plus-deskin havainnointimuistiinpanoja.

6.7 Aineiston analyysi

Valitsin aineiston analyysitavakseni laadullisen sisällönanalyysin, koska se on riittävän väljä teoreettinen kehys, joka on helposti liitettävissä erilaisiin analyysikokonaisuuksiin (Tuomi & Sarajärvi 2009, 91). Tarkoitukseni on tuottaa uutta tietoa aiemmin vähän tutkitusta asiasta, joten joustavuus ja aineistolähtöisyys on tälle tutkimukselle hyödyllisempää kuin täyteen teoriaohjaavuuteen sitoutuminen. Analyysini runko muodostuu kirjallisuuden avulla rakentamani teemahaastattelurungon perusteella, joten täydestä aineistolähtöisyydestäkään ei toki voida puhua. Pyrin kuitenkin aineistoa purkaessani ja luokittelua suunnitellessani säilyttämään herkkyyden aineistosta nouseville asioille. Päättelyni on siis abduktiivista, eli aineistolähtöisyys ja valmiit mallit vuorottelevat ajatusprosessissani (Tuomi & Sarajärvi 2009, 97). Tämän kautta päädyin esimerkiksi sisällyttämään työn haasteiden kuvauksen jokaisen toimituksen työnkuvaukseen, enkä rakentanut niistä pelkkää yhteenvetoa, kun huomasin jokaisen toimituksen työprosessin ja työn ongelmien olevan hyvin erilaisia, eikä niitä voinut yleistää.

Analyysini eteneminen noudattelee Ruusuvuoren, Nikanderin ja Hyvärisen (2010, 12) esittelemiä haastattelun analyysin vaiheita. Tutkimusongelman asettamisen, aineistonkeruutavan valinnan ja aineiston keruun jälkeen aloitin aineistoon tutustumisen ja sen järjestämisen. Litteroin haastatteluaineiston sanatarkasti, jättäen pois väliäänähdykset, yskähdykset ja joitakin asiaan liittymättömiä osuuksia. Asiantuntijahaastattelussa kun kiinnostuksen kohteena ei ole asiantuntija sinänsä, vaan hänellä oleva tieto, jonka avulla pyritään kuvaamaan ilmiökenttää faktuaalisesti (Alasalo & Åkerman 2010, 373, 375). Kirjoitin myös lyijykynällä muistivihkoon kirjoittamani havainnointimuistiinpanot tekstitiedostoiksi. Joukossa on myös havainnoinnin aikana tekemiäni lyhyitä haastatteluja. Etnografinen haastatteluaineisto ei aina olekaan litteroidussa haastattelun muodossa, ja riittävän tarkasti kenttäpäiväkirjaankin kirjatut haastattelut ovat käyttökelpoisia analyysia rakennettaessa (Huttunen 2010, 41).

Järjestin litteroidut haastattelut ja havainnointimuistiinpanot teemahaastatteluteemojeni pohjalta muodostamiini kategorioihin, jotka ovat:

1. Haastateltavan tausta
2. Datajournalismi toimituksessanne
3. Datajournalistinen työprosessi (yleistä)
 - 3.1 Datan hankinta
 - 3.2 Datan puhdistus ja analyysi
 - 3.3 Visualisointi tai muu lopputuote
4. Työn haasteet
 - 4.1 Työkalut
 - 4.2 Tekniset taidot
 - 4.3 Resurssit
 - 4.4 Kommunikaatio
5. Tulevaisuus
 - 5.1 Tulevaisuus teidän toimituksessanne
 - 5.2 Tulevaisuus Suomessa
 - 5.3 Ulkoistamisen mahdollisuudet
 - 5.4 Koulutus
6. Esimerkkiprojekti

Kategorioita on paljon ja osaan kohdista materiaalia kertyi vain yhden maininnan verran. Lisäsin haastattelurunkoon pohjautuvaan jaotteluun Esimerkkiprojekti-kategorian, johon keräsin tietoja yhdestä toimituksen datajournalistista työprosessia hyvin ilmentävästä työprojektista. Koodasin haastatteluaineistosta ihmisten selontekoja, jotka ovat heidän kokemuksia omasta työstään (Ruusuvuori, Nikander & Hyvärinen 2010, 20-21). Haastatteluaineiston koodaaminen oli valmiin teemoittelun vuoksi helppoa ja noudatin koodauksessa teemoittelun järjestystä. Jaottelin myös havainnointimuistiinpanot samoihin kategorioihin haastattelumateriaalin yhteyteen.

Muodostin kunkin toimituksen materiaalista oman jaottelun, jonka pohjalta kirjoitin yhtenäisen kuvauksen kunkin toimituksen datajournalismista, datajournalistisesta työprosessista sekä siinä koetuista yleisimmistä haasteista. Lisäksi esittelen kultakin toimitukselta yhden datajournalismin työtapoja hyödyntävän projektin. Tämä kuvaus vastaa ensimmäiseen ja osittain toiseenkin tutkimuskysymykseeni, eli ”millainen on

datajournalistinen työprosessi Helsingin Sanomien ja Yle toimituksissa?” ja “millaisia haasteita toimitukset kohtaavat datajournalismin tuottamisessa?”. Tiedostan, että tieteellisessä tutkimuksessa empiirisiä havaintoja ei kuitenkaan itsessään pidetä tutkimuksen tuloksina, vaan johtolankoina, joita tulkitsemalla pyritään pääsemään havaintojen taakse (Alasuutari 2011, 78). Toimitusten datajournalistisen prosessin kuvaus on kuitenkin jo käynyt läpi analyysin: kirjoitan yhteenvedot haastatteluaineiston ja oman havainnointini tuottamasta materiaalista tekemäni tulkinnan pohjalta, enkä suinkaan vain aineistoa esitellen. Lisäksi asiantuntijahaastattelun erityispiirteisiin kuuluu se, että aineiston keruu ja analyysi kietoutuvat yhteen, ja analyysia tehdään jo aineistoa kerätessä ja purettaessa (Alasalo & Åkerman 2010, 373).

Kokosin analyysini yleisiä tulkintoja ja kaikkia toimituksia yhdistäviä havaintoja ja näkökulmia, sekä keskustelun aiheita Pohdinta-lukuun. Pyrin ankkuroimaan esittämäni väitteet aineistooni nostamalla sieltä esimerkkejä ja korostamaan siten tutkimukseni läpinäkyvyyttä. Kirjoitin kirjallisuuden sekä datajournalismin tulevaisuutta käsitteleviin kategorioihin koodaamani materiaalin pohjalta Tulevaisuus-luvun, jossa vastaan viimeiseen tutkimuskysymykseeni “kuinka datajournalismia voidaan toimituksissa tulevaisuudessa edistää?”. Tässä luvussa analyysiaineistoni pääsee keskusteluun datajournalismia käsittelevän kirjallisuuden ja muiden alan pioneerien näkemysten kanssa. Päätäntökappaleessa pohdin tutkimukseni onnistumista ja esitän aiheita jatkotutkimukselle.

7 DATADESKIT JA TYÖPROSESSI

7.1 Helsingin Sanomien datadeski

Helsingin Sanomien toimituksessa datajournalismista vastaa datadeski, jossa työskentelevät tuottaja Esa Mäkinen, koodari-graafikko Jarmo Lundgren ja graafikko Minna-Liisa Linjala. Datadeski sijaitsee keskeisellä paikalla toimituksessa, verkkotoimituksen ja printtitoimituksen vieressä. Mäkisen mukaan datadeskin toimenkuvassa yhdistyvät anglosaksisessa maailmassa usein erotetut datan käsittely ja interaktiivisten elementtien luominen. Datajournalismin ohella deski tuottaakin erikoistaittoja, pelejä ja testejä sekä juttuihin lisättäviä interaktiivisia elementtejä, kuten karttoja ja aikajanoja. Työnkuva ei muutenkaan ole järin tarkkarajainen, vaan deskin työntekijät toimivat myös muun toimituksen apuna teknistä osaamista vaativissa tehtävissä ja vastaavat toimittajien tietotekniikkakysymyksiin.

Datadeskin nykyisessä (joulukuu 2013) kokoonpanossa Esa Mäkinen on tuottaja ja toimii samalla Helsingin Sanomien organisaatiossa datajournalismista vastaavana uutispäällikkönä, joten hän osallistuu laajasti toimituksen verkon kehittämiseen ja edustaa deskiään toimituksen kokouksissa. Datadeskin tuottajana Mäkinen kommunikoi toimituksen muiden osastojen ja toimittajien kanssa sekä hallinnoi juttuprosesseja. Tuottajan tehtävien lisäksi hän on koodaava toimittaja, joka hankkii ja käsittelee dataa sekä ohjelmoi uutissovellusten toiminnallisuutta. Hän myös tekee juttujen ja niiden elementtien ulkoasuun liittyvää työtä, kuten rakentaa Helsingin Sanomien verkkosivuilla käytettyjä erikoistaittoja. Jarmo Lundgren on ohjelmointitaitoinen graafikko, joka keskittyy usein suurempien projektien, kuten Luokkakoneen (s. 55) ulkoasun ja toiminnallisuuden rakentamiseen. Minna-Liisa Linjala on graafikko, jonka työnkuvaan kuuluvat verkkograafikon rutiinomaisemmat työt sekä verkkojuttujen ulkoasukehitys. Helsingin Sanomat myös etsii deskiin toista koodari-graafikkoa. Mäkisen mukaan datadeskillä on kuitenkin tälläkin hetkellä riittävä tekninen osaaminen toteuttaa halutunlaista datajournalismia. Aikaakin työtehtäviin on riittävästi, eivätkä deadlinet ole epärealistisia:

datadeskin työajat ovat suhteellisen joustavat ja usein töitä tehdään erilaisten suurempien projektien aikataulujen mukaisesti.

Helsingin Sanomat perusti datadeskin saadakseen verkkosivuilleen uniikkeja sisältöjä ja niiden avulla kilpailuetua. Deskin alkuvaiheista kerrotaan tarkemmin luvussa 3.2 (s. 17). Mäkisen mukaan jo huhtikuussa 2012, kun deskin perustamista suunniteltiin, lehden verkkopalvelun tiedettiin muuttuvan maksulliseksi. Datadeskin tekemillä jutuilla on viikoittaiset kävijätavoitteet, jotka on tähän saakka saavutettu hyvin. Mäkinen uskoo, että uuden työntekijän tulo deskiin nostaisi heidän tuotanto- ja kävijätavoitteitaan. Juttuun lisätty interaktiivinen elementti ei vielä sinänsä lisää jutun kävijämäärää, vaan tärkeämpi tekijä on onnistunut otsikko. “Datavirityksen” sisältäviä juttuja kuitenkin luetaan Helsingin Sanomien analytiikan mukaan noin minuutti pidempään, kuin juttuja, joissa sellaista ei ole.

Helsingin Sanomissa käytetään verkkojulkaisujärjestelmä FatWirea, jossa datadeskin visualisoinnit lisätään verkkojuttujen yhteyteen HTML-, CSS- ja JavaScript-koodina. Mäkisen mukaan datadeski on saanut vaikuttaa julkaisujärjestelmän kehittämiseen omalla “roadmapillaan”. Järjestelmään esimerkiksi lisättiin heidän toiveestaan ominaisuus, jonka avulla verkkojuttuihin voi lisätä standardikokoa suurempia kuvia tekemättä sitä työläästi käsin HTML-upotuksena. Pientä hienosäätöä FatWire kuitenkin vielä Mäkisen mielestä kaippaa. Datadeski saa valita itse käyttämänsä työkalut ja heillä on järjestelmävalvojan oikeudet toimituksen tietokoneisiin. He pyrkivät käyttämään työssään mahdollisimman paljon avoimen lähdekoodin ohjelmia.

Eniten datadeski tuottaa erilaisia valmiisiin malleihin perustuvia yksinkertaisia visualisointeja, jotka voivat olla esimerkiksi staattisia tai interaktiivisia karttoja ja diagrammeja. Rutiininomaisesti julkaistaan esimerkiksi viikoittaisten Mieli- ja mielipideilmasto-kyselyiden visualisointeja. Niitä varten kerätään ensin kyselyaineisto Helsingin Sanomien verkkosivuilla julkaistavalla kyselylomakkeella, jonka tulokset siirtyvät Google Driveen. Sitten tuloksista rakennetaan visualisointeja avoimen lähdekoodin Datawrapper-ohjelmaa käyttäen. Lähes kaikki interaktiivisia elementtejä tai minkäänlaista erikoistaittoa sisältävät verkkojutut toteutetaan datadeskissä, joten heidän tehtäviinsä kuuluu paljon erilaista

yksinkertaista rutiinitoimintaa. Poikkeus tähän ovat toimittajien Arkku-työkalun avulla tekemät yksinkertaiset visualisoinnit (ks. s.54). Graafikko Linjalan tulo deskiin on kuitenkin vähentänyt Mäkisen tarvetta käyttää aikaansa rutiinitehtäviin. Ajallisesti vaativimpia ovat laajat uniikkiprojektit, kuten erikoistaitot tai tässä luvussa (s. 55) esiteltävä sovellus Luokkakone.

Valmiiksi suunniteltujen projektien lisäksi datadeski tarttuu tarvittaessa päiväkohtaisiin aiheisiin sekä uutistapahtumiin, ja tyypillistä onkin, että työn painopiste siirtyy useiden päällekkäin etenevien projektien ja eteen tulevien kiireisten työtehtävien välillä. Mäkisen mukaan erimittaisten projektien hallinta on kehittynyt datadeskin olemassaolon aikana. Hän arvioi, että painotus on siirtynyt suurempiin juttuprojekteihin, kuten sunnuntaisivujen jutuista julkaistaviin uutissovelluksiin. Juttujen teon ja suunnittelun lisäksi Mäkisen tuottajarooliin kuuluu myös paljon erilaisia kokouksia ja paljon aikaa vievää sähköpostinvaihtoa toimittajien ja muiden yhteistyökumppaneiden kanssa. Datadeskissä käy myös paljon toimittajia etsimässä apua yksinkertaisiin teknisiin kysymyksiin esimerkiksi julkaisujärjestelmään liittyen. Mäkinen kuitenkin pyrkii opettamaan ja kannustamaan ihmisiä tekemään vaikeiltakin tuntuvia asioita itse. Vaikka työ koko ajan keskeytyykin, hän ei koe “tukitehtävien” ja kokousten merkittävästi häiritsevän datadeskin toimintaa, vaan katsoo niiden kuuluvan työhön. Lisäksi hänellä on uutispäällikkönä tarpeen vaatiessa mahdollisuus kieltäytyä liiallisuuksiin menevistä vaatimuksista.

Projektinhallinnassa datadeski käyttää Trello-ohjelmaa, johon listataan kaikki käynnissä olevat ja suunnitellut projektit, sekä niiden osanottajat. Trello on toimituksen sisällä avoin kaikille kiinnostuneille, mutta käytännössä sen käyttö rajoittuu datadeskin sisälle. Sovitut projektit merkitään myös toimituksen päivälistalle, jossa ne ovat kaikkien tarkasteltavissa. Toimituksen sisäiseen yhteydenpitoon datadeski käyttää sähköpostia ja puhelinta.

Usein idea datajournalistisesta jutusta syntyy, kun Mäkinen löytää päivälistalta aiheen, johon datapohjainen lähestymistapa sopisi. Ideoita tulee myös osastojen esimiehiltä ja toimittajilta, jotka tulevat datadeskiin ehdottamaan juttuaiheita ja esittelemään erilaisia

datasettejä tai pyytämään interaktiivisia elementtejä juttujensa yhteyteen. Juttuidean ohella datadeskin projektit saavat usein alkunsa mielenkiintoiselta vaikuttavasta datasetistä.

Tyypillisesti datadeskin jutun tekoon osallistuu tuottaja Mäkisen lisäksi joko koodari-graafikko tai graafikko, sekä jutun “omistava” toimittaja, jonka vastuulla on aineiston hankkiminen ja kirjoitustyö. Mäkisen mukaan toimituksessa on kolmisenkymmentä toimittajaa, jotka ovat tehneet datadeskin kanssa useita dataan pohjautuvia juttuja, sekä muutamia toimittajia, jotka ovat muita edistyneempiä datapohjaisten juttujen tekemisessä. Juttuihin osallistuvien ihmisten määrä luonnollisesti vaihtelee projektin laajuuden mukaan. Toimituksen ulkopuolisten freelancereiden kanssa datadeski työskentelee melko vähän ja ulkopuolisilta tilataan töitä, joita toimituksessa ei itse osata tai ehditä tehdä, kuten aikaa vievää datan puhdistusta tai tiettyjä vaikeita ohjelmointitöitä. Kokonaisia datajournalistisia juttujakin deski on ostanut, mutta kokemukset niistä ovat olleet huonoja, koska ulkopuolisen tekemiä toteutuksia on vaikea saada toimimaan juuri Helsingin Sanomien käyttämissä järjestelmissä, juuri oikealla tavalla.

Vaikka datadeskin työ on luonteeltaan tavallista toimittajantyötä teknisempää, Mäkinen ei koe vaikeaksi kommunikoida muiden toimittajien kanssa, koska hänellä itselläänkin on toimittajan eikä ohjelmoijan tausta. Hän kiertää Helsingin Sanomien eri toimituksissa kertomassa datadeskin työtavoista ja työkaluista sekä pitää Excel-työpajoja. Hänen mielestään on helpointa tehdä yhteistyötä Excelin käytön taitavan toimittajan kanssa. Silti nykyäänkään harva toimittaja tulee datadeskiin valmiiksi puhdistetun datan kanssa. Datadeskin olemassaolon aikana joidenkin toimittajien tekniset taidot ja näkemys ovat kuitenkin kehittyneet ja he ovat esimerkiksi oppineet käyttämään Arkku-työkalua tai ainakin Mäkisen osaamista hyväkseen. Yksi datadeskin perusajatuksia onkin opettaa teknistä osaamista ja yrittää lopulta minimoida tarve Mäkisen avulle.

7.1.1 Datan hankinta

Helsingin Sanomien datadeskin juttujen pohjalla on useimmiten data, joka on suoraan saatavilla verkosta, esimerkiksi Tilastokeskuksen tai eri virastojen sivuilta. Datan helppo saatavuus määrittelee juttuaiheen valintaa, ja Mäkisen mukaan usein poimitaan “rusinat pullasta”, eli tehdään juttuja aiheista, joista dataa on helppo hankkia. Viranomaiset antavat mieluummin erilaisia tilastoja ja tulkintoja datasta, kuin varsinaista dataa, mutta juuri data mahdollistaa toimituksen oman tulkinnan. Helsingin Sanomilla ei ole juurikaan kilpailua datapohjaisissa jutuissa, ja ainakin Helsingin kaupungin dataa he ovat saaneet käydä läpi ilman kilpailun painetta. Dataa onkin Mäkisen näkemyksen mukaan saatavilla enemmän, kuin juttuja ehtii tehdä.

Suoraan verkosta noutamisen lisäksi dataa hankitaan tietopyynnöillä ja harvemmin ruudun raavinnalla. Joukkoistamista Helsingin Sanomat tekee peruskyselyiden muodossa, mutta toiveissa on tehdä tulevaisuudessa kunnianhimoisempia joukkoistusprojekteja. Helsingin Sanomien digitaalisista sisällöistä vastaavan toimituspäällikön Perttu Kauppisen mukaan joukkoistaminen ei ole autuaaksi tekevä asia, eikä se välttämättä tuota käyttökelpoista dataa: ei voida luottaa siihen, ketkä kyselyihin ja joukkoistusyryksiin vastaavat ja mikä vastausten laatu on.

Hankintavaiheen haasteisiin kuuluu se, että viranomaiset antavat joskus dataa hankalasti käsiteltävässä muodossa, kuten pdf-tiedostoina. Tämä on kuitenkin aiempaa pienempi ongelma, koska tiedostojen oikeaan muotoon muuntamiseen on kehitetty erilaisia työkaluja ja työtapoja. Suurempi haaste on se, että datan irrottaminen ja käsitteleminen luovutuskelpoiseen muotoon saattaa maksaa enemmän, kuin mitä toimitus on siitä valmis maksamaan. Helsingin Sanomat liittää tietopyyntöihinsä maininnan jossa kehoitetaan viranomaista ottamaan erikseen yhteyttä, jos luovutuskustannukset ovat yli 500 euroa, minkä jälkeen datan hankintaa harkitaan tilanteen mukaan. Mäkisen mukaan ongelmaa pahentaa valtionhallinnon it-toimintojen ulkoistaminen Accenturen kaltaisille alihankintayhtiöille. Myös kuntatason viranomaisten toiminnan yhtiöittäminen on ongelma, joka saattaa siirtää dataa pois julkisuuslain luovutusvelvollisuuden piiristä. Datan hankintaa saattaa toisinaan vaikeuttaa myös viranomaisten heikko julkisuuslain tuntemus, joka

pitkittää hankintaprosessia. Mäkisen mukaan on tärkeää mainita jutussa, jos viranomaiselta on yritetty saada tärkeää datasettiä, mutta sitä ei ole saatu, eikä viranomaisen tekemää tulkintaa ole siten pystytty vahvistamaan.

7.1.2 Datan puhdistus ja analyysi

Datan puhdistamis- ja analysointivaiheet ovat usein juttuprosessin työläimmät ja hankalimmat. Pääasiallinen työväline puhdistamisessa on taulukkolaskentaohjelma Excel, jota Mäkinen käyttää myös monissa muissa tehtävissä, jotka vaativat tietojen, kuten kyselytulosten järjestelyä, tai esimerkiksi tietyn verkossa ulkoasua muokkaavan koodinpätkän liittämistä moneen eri kuvatiedostoon. Toisinaan datan puhdistukseen käytetään myös Googlen OpenRefine -työkalua ja Access-tietokantatyökalua. Puhdistuksessa ja analyysissä saatetaan käyttää myös muita ohjelmia, jotka joko rakennetaan itse tarkoitukseen sopivalla ohjelmointikielellä tai etsitään verkosta. Datan virheettömyyden varmistamisen pääasiallinen keino on kuitenkin kerääjätahoon ja keräystapaan kohdistuva journalistinen lähdekritiikki: vaikkapa Tilastokeskuksen datan voi olettaa olevan luotettavaa, mutta jos datan toimittaja on tuntemattomampi taho, datasta voidaan etsiä yksittäisiä polkuja, joita seurataan ja tarkastetaan tiedon paikkansapitävyys. Ihmisten keräämässä datassa on luonnollisesti myös erilaisia kirjoitusvirheitä ja väärin täytettyjä kohtia, jotka täytyy poistaa mahdollisimman huolellisesti. Kuitenkin myös erilaisten rajapintojen kautta saatavassa datassa, kuten Helsingin kaupungin liikennemääriä esittävässä datassa, voi olla virheitä. Jos datan luotettavuutta ei voida täysin varmistaa, siitä on kerrottava lukijalle.

Useampia datasettejä yhdistetään, mikäli jutun idea niin vaatii. Mäkisen arvion mukaan useamman datasetin juttuja tehdään kerran kahdessa viikossa. Yhdistämisvaiheessa ongelmia tuottavat esimerkiksi datasettien puutteelliset tunnistetiedot. Yhdistämis- ja analyysivaiheessa myös aineiston suuri koko saattaa hidastaa työtä, mikä huomattiin esimerkiksi, kun verrattiin koneellisesti 75 000 panamalaisen veroparatiisiyrityksen tietoja 10 000 Suomen rikkaimman ihmisen tietoihin. Analyysissä pyritään melko harvoin

osoittamaan tilastollista korrelaatiota, koska se on työlästä ja epäonnistumisen riski on suuri. Mäkisen mukaan deski tarvitsisikin lisää tilastotieteellistä osaamista.

7.1.3 Visualisointi tai muu lopputuote

Suurin osa datadeskin toteuttamista visualisoinneista tehdään Mäkisen kehittämää Arkku-työkalua käyttäen. Arkun avulla dataa, kuvia ja tietoja voi nopeasti ja yksinkertaisesti lisätä valmiille visualisointipohjille, kuten kartoille tai aikajanoille. Arkun on tarkoitus olla työkalu, jota toimittajat voivat käyttää myös ilman datadeskin apua. Mäkisen mukaan työkalun avulla on tähän mennessä tehty noin 30 juttua, joissa hän ei ole avustanut. Arkun lisäksi datadeski käyttää visualisointivaiheessa erilaisia avoimen lähdekoodin koodikirjastoja, kuten Charts.js:aa. Data-animaatioita datadeski ei juurikaan käytä ja niiden tekemiseen tarvittaisiinkin Mäkisen mukaan lisää osaamista.

Visualisointivaihetta nopeuttaa deskin toiminnan aikana muodostunut valmis koodipohja. Erilaisia projekteja varten on perusinfrastruktuurin, kuten mobiilitoiminnallisuuden, sisältäviä pohjia, eikä työtä siten tarvitse koskaan aloittaa täysin alusta. Kaikenlaisten valmiiden pohjien (template) käyttöä pyritään lisäämään entisestään. Visualisointivaiheen haasteet liittyvät yleensä erilaisten tietoteknisten ongelmien ratkaisemiseen, mistä esimerkkinä vaikkapa sisältöjen toimivuus Internet Explorer -selaimen vanhalla 8-versiolla tai mobiililaitteissa. Helsingin Sanomien verkkosivut tukevat Internet Explorerin versiota 8 ja sitä uudempia versioita, mutta Mäkinen yrittää saada toimitusta luopumaan IE8-tuesta, koska se aiheuttaa merkittävästi lisätyötä uutissovellusten rakentamisessa. Niin visualisointeihin, kuin muihinkin työvaiheisiin liittyvien ongelmien ratkaisussa auttavat erilaiset ohjelmointisivustot ja -verkkoyhteisöt, kuten Stack Overflow. Toisten ohjelmoijien asiantunteva vertaisapu takaa sen, että jokaiseen kysymykseen löydetään yleensä vastaus.

Sovelluksia ei julkaisun jälkeen juurikaan päivitetä, koska joitakin poikkeuksia lukuun ottamatta niiden lukijamäärät laskevat nopeasti julkaisun jälkeen. Mäkinen ei pidä

päivitetävien datajournalismipalveluiden tekemistä ja ylläpitämistä työmäärään nähden järkevänä.

Itsestään päivittyviä, reaaliaikaista dataa käyttäviä sovelluksia ei juurikaan julkaista. Tähän asti ainoa on Helsingin ruuhkakartta, joka käyttää kolmannen osapuolen luoman ohjelmointirajapinnan kautta saatavaa Liikenneviraston dataa. Se on julkaisemisen jälkeen liitetty muihinkin liikenneaiheisiin juttuihin. Mäkisen mukaan reaaliaikaista dataa voidaan tulevaisuudessakin käyttää, mutta toistaiseksi sitä tarjoavia lähteitä on vielä melko vähän.

Mäkisen mukaan Helsingin Sanomat julkaisee datajournalistisissa jutuissa käyttämänsä datan liian harvoin, ja hänen mielestään sitä tulisi tulevaisuudessa tehdä enemmänkin. Syy julkaisematta jättämiseen on usein se, että siitä on liikaa vaivaa ja datadeskillä on jo kiire siirtyä seuraavan projektin pariin.

7.1.4 Luokkakone

Helsingin Sanomien tähän asti suurin ja monimutkaisin datajournalistinen projekti, Luokkakone, julkaistiin lehden verkkosivuilla 3.1.2014. Luokkakone on BBC:n The Great British class calculator -sovelluksesta ideansa lainannut uutissovellus, jossa käyttäjä täyttää kyselyn, jonka tulosten mukaan häntä verrataan valmiista kyselyaineistosta muodostettuihin kulttuurisiin ja yhteiskunnallisiin luokkiin. Luokkakoneen julkaisun yhteydessä 4.1.2014 julkaistiin Helsingin Sanomien Kuukausiliitteen juttu, jossa toimittaja Saska Saarikoski kutsui kuusi kirjailijaa Säätytalolle keskustelemaan Suomen yhteiskunnallisista jakolinjoista.

Esa Mäkisen mukaan luokkakoneen suunnitteluprosessi alkoi sopivan yhteistyökumppanin etsimisellä. Turun yliopiston professori Jani Erola suostui mukaan ja toi projektin suunnitteluun tarvittavaa tilasto- ja yhteiskuntatieteellistä osaamista. Luokkakoneen pohja-aineisto kerättiin verkkokyselyllä, johon Erola ja Mäkinen suunnittelivat kysymyksiä, jotka kuvaisivat mahdollisimman hyvin vastaajan yhteiskunnallista ja kulttuurista luokkaa.

Koodari-graafikko Jarmo Lundgren rakensi kyselyn työlääksi osoittautunutta teknistä toteutusta noin viikon. Lopulta luokkakysely julkaistiin poikkeuksellisen näkyvästi. Mäkisen mukaan Sanoman tyhjää mainoskapasiteettia käytettiin tuon näkyvyyden hankkimiseen, ja kyselyn mainosta ajettiin voimallisesti esimerkiksi mainosbannereissa. Näkyvyystavoite saavutettiin ja kyselyyn tuli noin 42 000 vastausta, joille Mäkinen teki alustavan puhdistuksen etsien virheellisiä vastauksia ja kirjoitusvirheitä OpenRefinellä.

Erola muodosti vastaajista kuusi luokkaa käyttäen tietoja vastaajan tuloista, koulutuksesta, koetusta luokka-asemasta sekä vastaajan ja hänen mahdollisen puolisonsa, vanhempiensa sekä parhaan ystävän ammatista. Datan analyysi vaati paljon työtä: vastaajien ammattinimikkeet jaoteltiin käsityönä ISCO88-ammattiryhmiin ja sen jälkeen Eriksonin-Goldthorpen ammattipohjaisiin yhteiskuntaluokkiin. Aineisto luokiteltiin myös tulojen ja koulutuksen perusteella, eläkeläiset ja opiskelijat rajattiin pois, ja aineistosta otettiin 5 000 tapauksen otos. Lopullinen luokittelu kuuteen yhteiskuntaluokkaan - eliitti, asiantuntijat, kipuajat, duunarit, leipätyöläiset ja vähäosaiset - rakennettiin latent class -analyysillä, jonka jälkeen laskettiin ennustetut todennäköisyydet kuulua kuhunkin luokkaan. (Erola 2014.) Luokittelussa Erolalla oli apunaan kolme tutkimusavustajaa.

Luokittelun jälkeen Mäkinen rakensi Luokkakoneen toiminnallisuuden, eli laskentamoottorin, joka vertaa käyttäjän vastauksia luokiteltuun tausta-aineistoon. Ulkoasun suunnittelun tekivät Lundgren ja Linjala. Ennen julkaisua Luokkakonetta testattiin Helsingin Sanomien toimituksessa ja havaittuja vikoja korjattiin. Haasteita projektiin toi esimerkiksi kyselydatan likaisuus: kyselyyn syötettyjen ammattien joukossa oli paljon erilaisia kirjoitusvirheitä ja vitsivastauksia. Tämän vuoksi Mäkinen palkkasi tutkimusavustajan, joka tarkasti käsin 10 000 Excel-riviä ammatteja. Kaikkien vastausaineiston ammattinimikkeiden asiallisuus piti varmistaa, koska Luokkakone tarjoaa ammatti-kohtaa täyttävälle käyttäjälle ammattia automaattisen täydennyksen avulla.

Kaiken kaikkiaan projektiin osallistui kymmenen ihmistä: tuottaja-toimittaja Esa Mäkinen, toimittaja Saska Saarikoski, professori Jani Erola, neljä tutkimusavustajaa, Luokkakoneen

ulkoasun tehneet graafikot Minna-Liisa Linjala ja Jarmo Lundgren, sekä kuvituksen piirtänyt graafikko Klaus Welp.

Toimituksessa päätettiin, ettei Luokkakone tietoturvan vuoksi tallenna vastauksia. Myös tausta-aineistona käytetty data päätettiin jättää julkaisematta, koska se sisälsi ihmisten varallisuudesta kertovia tietoja postinumeroalueen tarkkuudella. Professori Erola kuitenkin sai datan käyttöönsä mahdollista tulevaa tutkimusta varten. Luokkakoneen jaettavuutta sosiaalisessa mediassa parannettiin tekemällä Facebook- ja Twitter-jakoteksteihin muuttuva ominaisuus, joka kertoo jakajan testistä saaman tuloksen. Ennen julkaisua Mäkinen arveli, että luokkakone tulee herättämään kiivasta keskustelua, mikä toteutuikin useiden blogikirjoitusten ja kommenttien muodossa. Hän arveli myös metodin keräävän kritiikkiä. Lopulta Luokkakoneen kuvaukseen kirjoitettiin maininta siitä, ettei sen antama tulos ole tieteellisen tarkka, vaan ainoastaan suuntaa antava. Luokkakoneesta odotettiin suurta hittiä, joten Mäkinen ei kokenut tarpeelliseksi hakea valmiille työlle erityistä näkyvyyttä mainonnan kautta. Kuukausiliitteen juttu toimi osaltaan sovelluksen markkinointikanavana. Sovellus olikin poikkeuksellisen suosittu: jos se olisi julkaistu vuoden 2013 puolella, se olisi ollut vuoden eniten näyttökertoja saanut juttu.

7.2 Svenska Ylen Grävnavet

Svenska Ylessä suurin osa datajournalismiosaamisesta on keskitetty Grävnavet-nimiseen ryhmään, jossa työskentelee tuottaja Eva-Maria Koskisen alaisuudessa kaksi ohjelmointitaitoista toimittajaa, Alexander Ugglan ja Peter Sjöholm. Ryhmä työskentelee omassa toimistossaan Ylen Radiotalossa hieman erillään ruotsinkielisten uutisten muusta toimituksesta ja Svenska.yle.fi-verkkokehitystyimestä. Grävnavetin projektit ovat dataa ja tilastoja hyödyntävää uutistystä, jossa näkyy usein toimittajien tausta tutkivan journalismin parissa. Ryhmän tyypillisintä toimintaa on laajojen, samasta aihepiiristä kumpuavien juttupakettien suunnittelu ja toteuttaminen, mutta sen ohella tehdään toisinaan pienempiä uutisia ja visualisointeja ja tarvittaessa autetaan muuta toimitusta teknistä, erityisesti Excel-osaamista, vaativien ongelmien ratkaisemisessa. Svenska Ylen uutisissa ja tutkimisissa

ohjelmissa on tehty datapohjaisia esimerkiksi vaalikonedataan pohjautuvia juttuja jo ennen Grävnavetin perustamista. Tuottaja Koskisen mukaan ryhmän perustamiseen johtivat toimituksen halu tehdä verkonomaista journalismia sekä nousta päivittäisuutisten virran yläpuolelle lisäämällä omaa uutishankintaa. Tarkemmin Svenska Ylen datajournalismin alusta kerrotaan luvussa 3.2 (s. 17).

Tuottaja Eva-Maria Koskinen hoitaa Grävnavetin työn organisointia ja kommunikoi muun toimituksen sekä Svenska Ylen aluetoimitusten kanssa. Lisäksi hän toimii koko kotimaan reportteritoimituksen esimiehenä ja istuu kotimaan toimituksen johtoryhmässä. Kokousten ja organisointityön lisäksi Koskinen osallistuu toimituksen oman uutistuotannon ideointiin ja tekee myös itse uutistyötä aikataulun salliessa. Alexander Uggl ja Peter Sjöholm ovat koodaavia toimittajia, joilla on molemmilla myös jonkin verran graafista ja tilasto-osaamista. Kenelläkään Grävnavetin toimittajista ei ole muodollista tietotekniikka-alan koulutusta. Jos ryhmä tarvitsee verkkototeutuksissa graafikkoa, he voivat pyytää apuun tv-uutisten, tai Svenska.yle.fi-kehitystiimin graafikkoa, jotka auttavat jos ehtivät. Suurimman osan graafisesta toteutuksesta ryhmä kuitenkin tekee itse. Lisäksi he auttavat muuta toimitusta visualisointien toteuttamisessa. Syksyllä 2013 Svenska.yle.fi-verkkokehitystiimiin palkattiin ohjelmoija Petri Kola, jonka työtehtäviin kuuluu myös datajournalismi ja datan käytön tuominen osaksi kaikkia Svenska Ylen ohjelmia. Muiden tehtäviensä ohella Kolan on tarkoitus tulevaisuudessa työskennellä myös Grävnavetin apuna vaativampaa ohjelmointia vaativissa projekteissa. Hänen osaamistaan on käytetty esimerkiksi Grävnavetin Fixa högskolan -juttupaketin (ks. s. 64) uutissovelluksessa, jolla käyttäjä voi tarkastella eri korkeakouluista valmistuneiden tulevaa palkkatasoa ja työllistymistä. Freelancerien ja muiden toimituksen ulkopuolisten apua ryhmän työssä ei ole käytetty. Ryhmän viikkotyöaika on 37,5 tuntia, mutta käytännössä se venyy projektien aikataulujen mukaisesti. Syksyn suurien projektien, kuten Fixa högskolan -joukkoistusprojektin aikana toimittajille kertyi paljon ylitöitä, minkä vuoksi he pitivät vapaapäiviä ennen joulua. Grävnavetin työtila on rauhallinen, eikä ulkopuolista häiriötä juurikaan ole. Toimittajat pyytävät datajournalisteilta usein apua teknisten ongelmien ratkaisemisessa ja esimerkiksi Excelin käytössä, ja tuottaja Koskisen mukaan myös tuon avun antaminen kuuluu ryhmän työhön.

Grävnavetin työ on suurimmaksi osaksi erilaisten saman aiheen ympärille rakennettujen, datapohjaisten juttupakettien tuottamista. Hyvä esimerkki tällaisesta on syksyllä energiaviikolla julkaistu sähkön siirtohintoja käsittelevä 14 jutun paketti (pääjuttu). Ryhmän tavoite on julkaista yksi juttupaketti viikossa ja aiheiden on tarkoitus kantaa lähetysten ykkösjuttuina. Pakettiin kuuluu juttuja verkkoon, radioon ja televisioon ja lähes kaikki Grävnavetin jutut julkaistaan uutisohjelmissa. Juttupakettien lisäksi päiväkohtaisia juttuja pyritään tekemään yhä enemmän, mutta tällä hetkellä niiden osuus on kuitenkin melko vähäinen. Tuottaja Koskisen mukaan Grävnavetin jutut ovat verkossa muuta sisältöä suosittumia ja niiden parissa viihdytään kauemmin. Datajuttuja lukeva usein myös jatkaa oleskelua Svenska Ylen verkkosivuilla, koska juttupaketin luonteen mukaisesti niihin linkittyy muita aiheeseen liittyviä juttuja. Grävnavetin juttujen suosiota pidetään toisinaan yllä lisäämällä niihin erilaisia kainalojuttuja pitkin päivää. Svenska Ylen sivuilla näkyvyyttä saavat suositut jutut, eikä datajournalistinen panostus itsessään nosta juttua sivuston prioriteetissa.

Juttujen aiheet painottuvat ruotsinkielistä yleisöä kiinnostaviin aiheisiin. Kielipolitiikkaan ja suomenruotsalaisiin yhteisöihin liittyvät aihevalinnat ovat yleisiä. Tekijöiden mukaan pienehkö kohderyhmä usein helpottaa yhteiskunnallisesti merkittävien juttujen tekemistä, mistä esimerkkinä verkostoanalyysi suomenruotsalaisten säätiöiden hallitusten jäsenistä. Joskus aiheet ovat myös koko Suomea koskevia. Ryhmän tutkivan journalismin painotus näkyy myös talous- ja politiikka-aiheiden suuressa määrässä, kevyempiä aiheita Grävnavetin jutut eivät juurikaan käsittele. Koskisen mukaan datapohjaisissa jutuissa aihevalintoihin vaikuttaa usein se, millaista dataa on saatavissa, minkä vuoksi he pyrkivät tekemään myös muita kuin datajuttuja.

Tuottaja Koskinen hoitaa suureksi osaksi kokoukset ja työn koordinoinnin, kun taas Ugglan ja Sjöholm keskittyvät toimittajan työhön ja juttujen toteutuksen rakentamiseen. Monissa suurissa projekteissa ovat mukana Svenska Ylen uutisten aluetoimitukset, joten merkittävä osa Koskisen työstä on tehtävien organisointia alueiden kanssa. Grävnavetin toiminnan alkuvaiheessa kaikki toimittajat hoitivat koordinointia, mutta työhön keskittyminen

helpottui, kun järjestelytyö keskitettiin tuottajalle. Projektinhallinta hoidetaan puhelimen, sähköpostin ja koko toimituksen käytössä olevan Lotus Notes -kalenterin avulla.

Grävnavetin suuremmat projektit etenevät tyypillisesti niin, että ryhmä tekee päätutkimuksen ja lähettää tuloksistaan eräänlaisen tiedotteen kaikille yhteistyötoimittajille, eli alueille, sekä television ja radion valtakunnallisille uutisille. Sen jälkeen ryhmä tekee päänettijutun ja usein myös valtakunnallisen radiojutun. Tv-version tekee yleensä uutisten toimittaja. Projektin edetessä Koskinen pohtii erilaisia näkökulmia aiheeseen ja koordinoi, miten ja missä järjestyksessä mikäkin toimitus voi aiheeseen tarttua. Valtakunnallisten ruotsinkielisten radio- ja tv-uutisten lisäksi aluetoimitukset tekevät aiheista omia, paikallisesti näkökulmattuja juttujaan. Suuret verkkovisualisoinnit pyritään julkaisemaan maanantai-aamuna ja niitä puffataan päivän radio- ja tv-uutisten jutuissa. Joskus Grävnavetin toimittaja saattaa myös esitellä uutissovellusta tai visualisointia tv-uutislähetyksessä. Koskisen mukaan datajournalismia esitetään televisiossa tv-grafiikan keinoin. Sekä televisiossa että radiossa data-aineistoa hyödynnetään etsimällä ihmisiä ja tapauksia, joita datan kuvaama asia koskettaa. Lisäksi projekteja mainostetaan sosiaalisessa mediassa, ja aina kun kyseessä on jonkinlainen sovellus, josta käyttäjä saa henkilökohtaisen tuloksen, siihen pyritään liittämään sosiaalisessa mediassa jaettavuutta helpottavia elementtejä.

Grävnavet tekee yhteistyötä myös Yle Uutisten Plus-deskin kanssa ja kääntää joskus heidän tekemiään juttuja ruotsiksi, sekä tarjoaa heille parhaita juttuja omasta tuotannostaan. Plus-deskin kanssa ryhmä kommunikoi kasvotusten lounailla ja iltapäiväkahveilla sekä yhteisen Google-kalenterin kautta.

Tiimi saa juttuideoita oman ideoinnin ohella aluetoimituksilta ja ideointikokouksista. Sjöholmin mukaan muiden toimittajien tietoisuus Grävnavetin olemassaolosta on kasvanut jatkuvasti, ja ihmiset osaavat jo olla heihin yhteydessä, kun kaipaavat juttuihinsa datapohjaisia elementtejä. Juttupaketti viikossa -tavoite kuitenkin pakottaa priorisoimaan, eikä kaikkiin, varsinkaan hahmottomampiin, ideoihin voida tarttua. Sjöholm kuitenkin korostaa, että yhteistyö eri toimittajien kanssa on tärkeä keino tehdä tutuksi

datajournalismia ja ryhmän työtapoja, ja sen vuoksi pienempiäkin juttuja pitäisi pystyä tekemään enemmän. Koska kaikki Grävnavetin jäsenet ovat itsekin ennen kaikkea toimittajia, ryhmän työn tekninen luonne ei aiheuta vaikeuksia kommunikoida muun toimituksen kanssa. Ugglan ja Sjöholmin ovat pitäneet Pasilassa ja aluetoimituksissa kursseja, joissa he kertoivat datajournalismista. Sjöholm myös visioi, että tulevaisuudessa voitaisiin luoda toimituksen käyttöön työkaluja, jotka tekisivät esimerkiksi datan visualisoinnista entistä helpommin lähestyttävää. Vaikka tieto datan käytön mahdollisuuksista ja muista työtavoista leviää, tällä hetkellä tilanne on kuitenkin vielä se, että harva toimittaja tulee Grävnavetin työhuoneeseen hyvin puhdistetun datasetin kanssa.

Tuottaja Koskisen mielestä ryhmän perustaminen vaati johdolta rohkeutta, koska sen vuoksi päivittäisen uutistoiminnan käytössä on kolme toimittajaa vähemmän. Svenska Ylen uutisten johdossa siis ymmärretään datajournalismin tärkeys ja muu toimitus tukee Grävnavetia riittävästi. Koskisen mukaan ryhmän toiminnalle kuitenkin asetettiin aluksi liian suuret tavoitteet tuotannon määrässä ja siinä, että juttujen pitäisi aina kantaa uutisten pääaiheina. Sittemmin odotukset ovat vakiintuneet realistisemmiksi. Sjöholm arvioi, että muu toimitus ei välttämättä aina tiedä, mikä on realistinen aikataulu, koska datajournalistisen prosessin työläyttä on vaikea arvioida. Hän kokee yhden juttupaketin viikkotahdin sopivaksi, mutta koska työ vaatii myös jatkuvaa uusien työkalujen opettelua ja osaamisen päivittämistä, ylitöitä kertyy. Ryhmälle syksyn aikana kertyneiden ylityötuntien määrä osoittaakin, että projektinhallinnassa on vielä kehittämistä. Toimituksen joustavuus kuitenkin näkyy siinä, että ylityötunnit on saatu pidettyä pois pitkinä vapaina. Ryhmän olemassaolon aikana Grävnavetin toimittajat ovat oppineet entistä paremmin arvioimaan, paljonko mikäkin työprosessi vie aikaa, sekä tutustuneet toistensa vahvuuksiin, joten he uskovat tulevien projektien ajanhallinnan olevan tulevaisuudessa helpompaa. Lisäksi ryhmän toiminta on kehittynyt teknisen osaamisen ja valmiin koodipohjan osalta. Tulevaisuudessa Grävnavet tarvitsee Sjöholmin mukaan lisää osaamista ohjelmoinnissa ja tilastotieteessä.

Svenska Ylellä käytetään sisällönhallintajärjestelmä Drupalia, jonka kautta myös Grävnavetin tekemät visualisoinnit ja muut interaktiiviset elementit julkaistaan. Drupal on

avoimen lähdekoodin ohjelmisto, jota Svenska Ylen kehitystiimi voi muokata tarpeen mukaan. Julkaisujärjestelmäongelmia kuitenkin on melko paljon esimerkiksi dynaamisten, muotoaan muuttavien elementtien, kuten kasvavien taulukoiden julkaisemisessa. Sjöholmin mukaan tekniselle osastolle voi esittää toiveita Drupalin kehittämisestä, ja keskusteluilmapiiri on hyvä, mutta koko Svenska Ylen tietoteknisiä ongelmia ratkovat verkkokehittäjät ovat kiireisiä, eikä muutoksia välttämättä ehditä tehdä uutisjuttujen julkaisun vaatimalla tahdilla. Koskisen mukaan verkkokehittäjien ajattelu eroaa deadlinejen paineessa työskentelevän toimittajien ajattelusta siten, että he eivät välttämättä pidä aikaansaannostaan, kuten verkkosivua koskaan valmiina, kun toimittajan taas on pakko päästä juttunsa käsistään, jos aikataulu vaatii. Julkaisujärjestelmä pois lukien Grävnavetin toimittajat voivat valita vapaasti datajournalismin teossa käyttämänsä työkalut. Uusien työkalujen käytön opettelu erilaisten verkko-oppaiden ja -yhteisöjen avulla onkin merkittävä osa heidän työtään. Googlen ilmaisia työkaluja, kuten Google Visualization API:a ja Fusion Tablesia käytetään paljon esimerkiksi hyvän mobiilitoimivuuden vuoksi. Sjöholmin mukaan niistä pitäisi tulevaisuudessa päästä eroon, koska ilmaistyökalujen ei voi aina luottaa olevan kunnossa. Hänen mielestään yleisradioyhtiön ei muutenkaan ole hyvä liikaa suosia yhden, suuren yrityksen työkaluja.

7.2.1 Datan hankinta

Grävnavet hankkii suurimman osan käyttämästään datasta pyytämällä sitä erilaisilta virastoilta ja organisaatioilta tai yksinkertaisesti verkosta hakemalla. Esimerkiksi Tilastokeskuksen dataa he käyttävät paljon, vaikka se usein onkin muutaman vuoden vanhaa. Maksullista dataa ryhmä ei ole juurikaan käyttänyt. Ruudun raavintaa ei ole datan hankinnassa käytetty kuin kerran. Joukkoistusta ja erilaisia kyselyitä Grävnavet sen sijaan käyttää melko paljon, ja Sjöholmin mukaan joukkoistuksen mahdollisuudet on Svenska Ylellä huomattu laajemminkin. Käyttökelpoisen kvantitatiivisen aineiston hankkiminen joukkoistamalla on kuitenkin vaikeaa, ja Sjöholmin mukaan joukkoistus toimiikin paremmin laadullisen datan, kuten erilaisten tarinoiden tai juttuaiheiden keräämisessä.

Suurimpia ongelmia datan hankinnassa aiheuttavat viranomaisten puutteellinen julkisuuslain tuntemus, sekä se, että osa materiaalista on virastoissa edelleen vain paperimuodossa, ja sen saaminen maksaa, varsinkin jos halutuissa asiakirjoissa on myös ei-julkista tietoa. Suomenruotsalaisten säätiöiden hallitusten jäsenten suhteita käsittelevän jutun aineistoa hankkiessaan Sjöholm istui Patentti- ja rekisterihallituksen toimistossa ja kirjoitti tietoja muistiin säätiörekisteristä, koska niiden tulostamisesta olisi veloitettu tuhansia euroja ja tiedot olisivat silti olleet vaikeasti käytettävissä pdf-muodossa. Väärässä tiedostomuodossa annettu data ei kuitenkaan Sjöholmin mielestä ole yleensä kovin suuri ongelma, koska sitä tulee harvoin, ja sen muuntamiseen on hyviä työkaluja.

7.2.2 Datan puhdistus ja analyysi

Suurin osa datan puhdistus- ja analyysityöstä tehdään Excelissä, joissain tapauksissa on käytetty myös OpenRefinea. Vaikka ryhmän ulkopuoliset toimittajat usein hankkivat datasettejä, puhdistuksen ja analyysin tekevät usein Grävnavetin toimittajat, koska heillä on paras Excel-osaaminen. Datan virheettömyys pyritään varmistamaan siten, että useampi kuin yksi ryhmän jäsen kävisi dataa läpi ja tekisi satunnaistarkastuksia. Sjöholmin mukaan datan tarkistaminen on yksi ryhmän suurimpia kehittämiskohteita. Useita datasettejä yhdistetään lähes kaikissa suuremmissa projekteissa.

7.2.3 Visualisointi tai muu lopputuote

Visualisointien rakentamisessa Grävnavet käyttää paljon Googlen työkaluja, kuten Google Visualization API:a ja Google Mapsia. Lisäksi työtä helpottavat erilaiset JavaScript-koodikirjastot, kuten jQuery ja List.js. Varsinkin yksinkertaisia visualisointeja varten on tehty jonkin verran valmiita pohjia, joissa on valmiina esimerkiksi Svenska Ylen sivustolle sopivat värit. Vanhaa koodia pyritään käyttämään hyödyksi myös visualisointien toiminnallisuuden luomisessa.

Uutissovelluksia ja visualisointeja ei päivitetä julkaisun jälkeen, mutta niiden käsittelemiä aiheita saatetaan kuitenkin tuoda pinnalle tekemällä jatkojuttuja. Interaktiivisia elementtejä käytetään myös uudestaan, kun se on mahdollista. Tästä esimerkki on eniten eri säätiöiden hallituksessa toimivista ihmisistä tehty lista, joka käytettiin uudelleen veropäivänä, kun verkostoituneimpien henkilöiden tulot tahdottiin uutisoida.

Lopputuotteen rakentamisessa suurimmat haasteet tulevat Sjöholmin mukaan siitä, että sekä hänellä että Ugglalla on suhteellisen vähän ohjelmointikokemusta. Sen vuoksi toteutus voi kestää melko kauan. Myös visualisointien saaminen Drupaliin tuottaa usein ongelmia, jotka johtuvat esimerkiksi julkaisujärjestelmän omien koodikirjastojen ja visualisoinneissa käytettyjen JavaScript-kirjastojen yhteensopivuusongelmista.

Jutuissa käytettyjä datasettejä ei tähän asti ole jaettu avoimena, Sjöholm kuitenkin toivoo, että tulevaisuudessa ainakin kaikki itse kerätty data jaettaisiin, kunhan se ei sisällä esimerkiksi henkilötietoja. Syy jakamatta jättämiseen on ajan puute: kun projekti loppuu, pitäisi olla jo tekemässä seuraavaa juttua. Toisten, kuten virastojen, dataa käyttävässä jutussa hän ei pidä jakamista merkittävänä.

7.2.4 Fixa högskolan

Fixa högskolan on Grävnavetin tähän saakka suurin dataa ja joukkoistamista hyödyntävä projekti. Sen tarkoitus oli kannustaa yleisöä tuomaan esille Suomen ruotsinkieliseen korkeakoulumaailmaan liittyviä epäkohtia, joita toimitus voisi sitten tutkia. Lisäksi juttupakettiin kuului sovellus, jolla käyttäjä voi tarkastella eri korkeakoulujen ja tiedekuntien opiskelijoiden työllistymistä ja palkkatasoa. Laskuri ja vihjelomake julkaistiin 25.11.2013 ja sen jälkeen vihjeiden perusteella tehtiin ja julkaistiin juttuja kahden viikon ajan. Joukkoistus tuotti 140 vastausta, joiden perusteella tehtiin yli 40 juttua esimerkiksi ruotsin- ja suomenkielisen koulutuksen saaneiden ekonomien palkkaeroista ja Helsingin yliopistossa Google Translaten avulla kelvottomasti ruotsiksi käännettyistä tenttikysymyksistä. Lisäksi jutut keräsivät yhteensä noin 300 kommenttia. Uutislähetysten

ja verkon lisäksi projekti oli esillä Efter tolv -radio-ohjelmassa ja Obs debatt -keskusteluohjelmassa televisiossa. Projektiin osallistui kaikkiaan yksitoista henkilöä: Grävnavetin tuottaja Eva-Maria Koskinen, toimittajat Peter Sjöholm ja Alexander Ugglä, sekä Svenska.yle.fi-verkkokehitystiimissä työskentelevä ohjelmoija Petri Kola, ja heidän lisäksi juttuja teki viisi toimittajaa aluetoimituksista, yksi uutistoimituksesta sekä yksi Svenska Ylen tutkivasta ryhmästä.

Kola rakensi opintolaskurin Grävnavetin keräämästä datasta. Sjöholmin mukaan datan saattaminen oikeaan muotoon oli suuritöistä, koska lopullista laskuria varten yhdistettiin kymmeniä datasettejä, jotka eivät kaikki olleet merkinnöiltään yhteensopivassa muodossa. Suurin ongelma oli se, että osa korkeakouluista valmistuneiden tiedoista oli esimerkiksi tutkintotasolla ja osa tiedekuntatasolla. Alunperin tarkoitus oli saada kaikki data laskuriin tutkintotasoisena, mutta lopulta päädyttiin tiedekuntatasoiseen esittämiseen. Eri tasoisen datan käyttäminen olisi tehnyt laskurista liian sekavan. Datan puhdistustyö vaati paljon taustaselvitystä eri opintoalojen ja nimikkeiden suhteista. Laskuria oli tarkoitus tarjota myös suomenkieliselle puolelle, mutta ajanpuutteen vuoksi se jäi tekemättä.

Joukkoistuksen onnistumiseen vaikutti tekijöiden mukaan se, että kohderyhmä oli pieni ja tiivis, eli noin 14 000 ruotsinkielisten korkeakoulujen opiskelijaa ja työntekijää. Kohderyhmällä oli myös kokemusta sosiaalisen median käytöstä, joten kynnyks osallistua oli matalampi kuin vanhemmalla yleisöllä. Nuorten tavoittamisessa Grävnavet sai apua myös radiokanava X3M:ltä, jonka kautta kyselyä saatiin markkinoitua. Joukkoistamista tehostettiin myös siten, että ennen vihjelomakkeen julkaisua toimittaja Ugglä loi valmiin verkoston soittamalla 50 eri ainejärjestön puheenjohtajalle ja kehottamalla heitä levittämään tietoa joukkoistuksesta järjestöjen jäsenille. Fixa högskolanin Twitterissä herättämän keskustelun perusteella saatiin myös hankittua opiskelija-aktiivi osallistumaan Obs debatt -ohjelmaan.

Sjöholmin mukaan Fixa högskolan -sivuston rakentaminen Drupalin sisällä oli vaikeampaa, kuin mitä se olisi ollut vaikkapa Wordpress-sivupohjan avulla, mutta pilottiprojektin

tekeminen omilla työkaluilla kannatti: nyt kokemusta voidaan käyttää jatkossa apuna ja monen asian tekeminen on nopeampaa.

Suomenruotsalaisessa korkeakoulumaailmassa projekti herätti luonnollisesti paljon keskustelua, ja esimerkiksi Svenska Handelshögskolanissa yliopiston henkilökunnalle lähetettiin viesti, jossa kehoitettiin muistamaan koulun kunnia ja ilmoittamaan toimittajien yhteydenotoista koulun johdolle. Fixa högskolan sai paljon myönteistä huomiota myös Ylen sisällä, ja Ajankohtainen Kakkonen aloitti sen innoittamana oman Peruskouluremppea-projektinsa. Lähes kaikki juttupaketin jutut olivat Svenska Ylen sivuilla keskivertoa luetumpia ja suosituimmat Fixa högskolanin jutut, kuten opintolaskuri, nousivat poikkeuksellisen luetuiksi.

7.3 Ylen Plus-deski

Plus-deski suunnittelee ja toteuttaa Yle Uutisten verkkosivuille verkkoekstroja, eli juttuja, jotka hyödyntävät datajournalismin ja erikoistaittojen kaltaista verkonomaista kerrontaa. Kyseessä ei ole datadeski sanan varsinaisessa merkityksessä, koska datajournalismi on vain yksi osa heidän tuotantoaan. Uutis- ja ajankohtaissisältöjen ohella Plus-deski tuottaa Yle Uutisten sivuille myös kevyempää sisältöä, kuten erilaisia laskureita, testejä ja tietovisoja. Lisäksi he tekevät eri toimitusten käyttöön verkkoon sopivaa infografiikkaa. He työskentelevät Yle Uutisten pääuutishuoneen reunassa, venäjän-, englannin- ja viittomakielisten uutisten deskien vieressä. Plus-deskin perustamisesta kerrotaan tarkemmin luvussa 3.2 (s. 17).

Tuottaja Juho Salmisen mukaan deski täytti ensimmäisenä toimintavuotenaan kaikki sille asetetut tavoitteet. Plus-deskin tuottamia elementtejä sisältävien juttujen parissa vietetään noin kaksinkertaisesti aikaa muihin Yle Uutisten juttuihin verrattuna. Sivuston vuoden 2013 viidestä luetuimmasta jutusta kolmessa oli jollakin tavalla mukana Plus-deskin panos. Salminen kokee, että ryhmä on olemassaolonsa aikana tullut myös jatkuvasti tunnetummaksi Ylen toimittajien keskuudessa. Vuonna 2013 deski osallistui tavalla tai

toisella yhteensä noin 350 jutun tekemiseen. Kaikki nämä jutut on merkitty Google Drive -dokumenttiin, johon on listattu myös kunkin projektin onnistumiset ja niissä huomattut kehittämiskohteet.

Tuottaja Salminen organisoii deskin yhteistyötä uutistoiminnan kanssa, eli osallistuu uutisten viikkokokouksiin ja välittää juttuideoita ja projekteja muun toimituksen ja Plus-deskin välillä. Tuottaja Mika Pippurin toimenkuvaan taas kuuluu yhteistyön organisointi Ylen ajankohtaistoiminnan kanssa sekä ajankohtaistoiminnan verkon kehittäminen. Teemo Tebest on deskin datajournalisti, joka vastaa datan käsittelystä sekä applikaatioiden ja visualisointien toteutuksesta. Graafikot Juha Rissanen ja Stina Tuominen suunnittelevat ja toteuttavat verkkosisältöjen ulkoasua ja erilaisia graafisia esityksiä, kuten infografiikkaa. Plus-deskin työntekijöiden keskinäistä kommunikaatiota helpottaa se, että he kaikki ymmärtävät jonkin verran ohjelmointia ja verkkosivujen rakennetta, mutta Tebest, tietotekniikan diplomi-insinööri, on ainoa, jolla on aiheeseen varsinaista koulutusta. Jos projekti vaatii lisää ohjelmointi- ja verkkokehitysosaamista, sitä haetaan Ylen verkkokehitystiimiltä, joka osallistui esimerkiksi Plus-deskin Arvokone-projektiin. Salminen arvioi, että vuoden 2014 aikana deskiin pyritään rekrytoimaan lisää ohjelmointiosaamista. Tälläkin hetkellä deskin työntekijöiden tekniset taidot ovat hyvät, ja he pystyvät tuottamaan halutunlaisia sisältöjä. Tulevaisuudessa Plus-deskillä voisi kuitenkin olla käyttöä esimerkiksi tilastotieteen, geanalyysin ja paikkatietoanalyysin osaamiselle.

Plus-deskillä ei ole työvuorolistoja, vaan työaika on joustava ja mukautuu projektien aikatauluihin sekä tarvittaessa henkilökohtaisiin tarpeisiin. Tebest korostaa tätä joustavuutta: deskissä työt hoidetaan loppuun ja ylityöt merkitään muistiin ja pidetään vapaina, kun mahdollista. Kokoukset eivät juurikaan keskeytä deskin työskentelyä, koska niissä käymisestä ja edustamisesta vastaavat tuottajat. Yle Uutisten muut toimittajat pyytävät Plus-deskiltä paljon apua yksinkertaisissakin teknisissä asioissa, kuten verkkojulkaisujärjestelmän edistyneempien ominaisuuksien ja taulukkolaskentaohjelma Excelin perustoimintojen käytössä. Tebest kokee, että toimittajien avustaminen ei aiheuta ylimääräistä häiriötä ja korostaa, että se kuuluu paitsi hänen toimenkuvaansa, myös yleisiin

hyviin tapoihin. Lisäksi toimittajien avustaminen Excelin käytössä tehostaa usein merkittävästi koko työprosessia.

Plus-deski tekee yhteistyötä koko Ylen uutis- ja ajankohtaistoiminnan kanssa. Yksikköön kuuluvat Yle Uutisten kotimaan-, talous-, politiikka-, kulttuuri- ja ulkomaantoimitus, uutispäivystystä tekevä Nyt-deski, Ylen aluetoimitukset sekä A-ohjelmat. Eniten yhteisprojekteja tehdään kotimaantoimituksen ja taloustoimituksen kanssa. Tyypilliseen projektiin osallistuu juttua tekevän toimittajan lisäksi toinen Plus-deskin tuottajista, koodari Tebest sekä toinen graafikoista. Jotkut projektit eivät välttämättä vaadi koodarin tai graafikon työpanosta. Projekti voi olla joko alusta asti verkkoekstraksi suunniteltu juttu tai nopeampi uutistapahtumaan tarttuminen, kuten myrskytuhoja esittelevä karttavisualisointi, jolloin tilaus tulee uutisten tilannejohdolta. Varsinkin pidemmissä, useita päiviä kestävässä projekteissa ongelmia aiheuttaa Plus-deskin ja eri toimitusten toimittajien erilainen työaika: esimerkiksi kotimaan toimituksen reporterit saattavat tehdä viikossa muutaman 12-tuntisen työpäivän ja sitten pitää useita vapaapäiviä, kun plusdeskiläiset taas ovat toimituksessa yleensä virka-aikaan. Tätä ongelmaa on helpotettu ottamalla yksi verkonomaisesta ilmaisusta kiinnostunut toimittaja kerrallaan Plus-deski-vuoroon, eli työskentelemään deskiin kolmeksi viikoksi. Salmisen mukaan kokemukset vuorosta ovat olleet hyviä, ja käytäntöä yritetään jatkaa tulevaisuudessa.

Yleensä projekteja johtaa toinen Plus-deskin tuottajista. Projekteja organisoidaan deskin omalla Google-kalenterilla ja pienissä aamupalavereissa, joihin osallistuvat kaikki projektiin liittyvät henkilöt. Ideat projekteihin tulevat yleensä toimittajilta. Kaikista ideoista, heikoistakin, pyritään keskustelemaan, eikä mitään suoraan torjuta. Toteutettavia ideoita tulee myös aluetoimituksilta. Plus-deskin juttuihin pyritään kuitenkin aina etsimään koko maan laajuinen näkökulma. Tebestin mukaan toimittajien ideat ovat harvoin varsinaista tutkivaa datajournalismia, vaan liittyvät yleensä erilaisiin visuaalisiin elementteihin, jotka lisätään jutun oheen. Salminen kokee, että toimittajat ovat kuitenkin sitten deskin alkuaikojen hieman kehittyneet sellaisen ideoinnin tasolta, jolla interaktiivisia elementtejä hyödynnetään vain datan esittämisessä. Toiveissa kuitenkin on, että tulevaisuudessa yhä useammassa projektissa korostuvat datan analysointi ja kontekstointi

sekä uusien havaintojen tarjoaminen. Tulevaisuudessa plusdeskiläiset toivovat voivansa tehdä nykyistä enemmän isoja ja yhteiskunnallisesti merkittäviä projekteja sekä tutkivampaa datajournalismia. He uskovat, että pienempien päivittäisten visualisointien tekemiseen ei tulevaisuudessa enää tarvita nykyisessä määrin heidän panostaan, vaan sitä voivat tehdä toimittajat itse.

Hyvien juttuideoiden saaminen edellyttää sitä, että toimittajat ymmärtävät, millaisia toteutuksia verkkoon on mahdollista tehdä ja miten dataa voi hyödyntää. Tätä pyritään edistämään Plus-deskin työtapoja ja työvälineitä esittelevillä lyhyillä kursseilla, joita pidetään sekä Pasilan toimituksissa että aluetoimituksissa. Kuitenkin erään Plus-deskin kanssa työskennelleen toimittajan mukaan esimerkiksi ulkomaantoimituksessa deski ja sen tarjoamat yhteistyömahdollisuudet ovat ensimmäisen toimintavuodenkin jälkeen melko vieraita. Deskin toiminnan aikana tietyt toimittajat ovat alkaneet käyttää heidän apuaan muita ahkerammin ja tulevat usein juttelemaan ideoistaan. Heidän määränsä myös kasvaa jatkuvasti. Salminen korostaa, että toimittajien on itsekin otettava vastuuta uusien työtapojen opettelusta ja mahdollisuuksiin tutustumisesta. Tebestin mukaan toimittajien ymmärrys työtavoista on lisääntynyt, eikä oikeassa muodossa olevan datan kanssa Plus-deskille kävelevä toimittajakaan ole enää harvinaisuus. Suurin osa teknistä osaamista vaativista työvaiheista tehdään kuitenkin edelleen deskissä. Plus-deskin työn tekninen ulottuvuus ei juurikaan vaikeuta kommunikointia toimittajien kanssa, koska juttuprojektien etenemisestä neuvottelevat toimittajataustaiset tuottajat. Myös tietotekniikka-alalle kouluttautunut Tebest kokee pystyvänsä keskustelemaan teknisistä asioista toimittajien kanssa ja sovittamaan teknisten yksityiskohtien määrän kunkin keskustelukumppanin osaamiseen.

Valmiille Plus-deskin jutuille annetaan paljon näkyvyyttä Yle Uutisten verkkosivuilla. Juttuja pyritään levittämään myös rakentamalla niihin sosiaalisessa mediassa jakamista edistäviä elementtejä, kuten houkutteleva ja omien valintojen mukaan personoitu jakoteksti. Jos kyseessä on testi, joitakin käyttäjien vastaustietoja saatetaan tallentaa jatkojuttuja varten. Vastaustietojen tallentamisesta ei ole erikseen ilmoitettu jutussa, mutta Salmisen mukaan siitä aiotaan jatkossa ilmoittaa. Jos datajuttujen aihe kantaa, niitä versioidaan myös

tv- ja radiouutisiin normaaleina uutisjuttuina, joissa kehoitetaan hakemaan lisätietoa Yle Uutisten verkkosivuilta. Toisinaan toimittaja esittelee uutissovelluksen plasmanäytöltä suorassa uutislähetysessä.

Yle Uutisissa on käytössä verkkojulkaisujärjestelmä Escenic, jossa luotun verkkoartikkeliin Plus-deskin tuottamat sovellukset ja muut sisällöt lisätään ulkoisina sisältöinä. Sovellusten ulkoasun ja toiminnallisuuden sisältävät HTML-, CSS- ja JavaScript-tiedostot sekä niiden mahdollisesti käyttämä data tallennetaan Ylen verkkolevylle ja luetaan sieltä, kun nettiartikkeli avataan selaimessa. Tämä erillään tallentaminen tekee sisällöistä kevyempiä ladata ja helpompia hallita. Tebestin mukaan Esceniciin halutuista muutoksista voi keskustella verkkokehitystiimin kanssa, ja toiveita pyritään kiireestä huolimatta huomioimaan. Tähän mennessä Plus-deskin aloitteesta sisällönhallintajärjestelmään on lisätty mahdollisuus julkaista ulkopuolisia toteutuksia leipätekstin lomassa. Ratkaisua odottavia Escenic-ongelmia ovat se, etteivät ulkoiset sisällöt toimi Ylen m.yle.fi-mobiilisivustolla ja se, että mikäli artikkeliin on liitetty ulkoisia sisältöjä, ja jutun leipäteksti sijaitsee ulkoisessa sisällössä, eli koodin joukossa, sen muokkaaminen vaatii tuon koodin sisältävän tiedoston muokkaamista, eikä sitä voi tehdä suoraan Escenic-artikkelissa. Tämä tarkoittaa käytännössä sitä, että Tebest myös korjaa Plus-deskin sisältöjä sisältävistä jutuista kirjoitusvirheet ja muut pikkuviat, koska osaamattoman muokkaajan koodiin jättämä virhe voisi häiritä jutun tai koko Yle Uutisten sivuston toimintaa. Toteutusten toiminta eri selaimilla ja alustoilla testataan deskin omilla tietokoneilla, älypuhelimilla ja tableteilla. Toteutukset tehdään toimimaan myös vanhalla Internet Explorer 8 -selaimella, mutta jos se ei onnistu, asia ilmoitetaan jutun yhteydessä, ja lukija ohjataan lataamaan uudempi selain. Tähän mennessä muutamaa sovellusta ei ole saatu toimimaan IE8:lla. Tulevaisuudessa on tarkoitus hankkia testilaitteisto, jonka avulla toimivuuden voi testata kätevästi samalta päätelaitteelta käsin.

Plus-deski saa itse valita käyttämänsä työkalut. Ainoa rajoite on se, että niillä tuotetun sisällön tulee toimia kaikilla Yle Uutisten sivujen tukemilla alustoilla ja selaimilla ja olla niin kevyttä, että se toimii myös mobiiliyhteyksien yli. Tebest tekee itse päätökset työkalujen sopivuudesta ja hankkii kuhunkin projektiin tarvittavat työkalut verkosta tai

rakentaa ne itse tarkoitukseen sopivalla ohjelmointikielellä. Hän pyrkii työssään automatisoimaan kaiken mahdollisen ja välttää aikaa vievää ja virhealtista käsityötä aina, kun se on mahdollista. Varsinaisia valmiita pohjia, templateja, Tebest ei kuitenkaan mielellään käytä, koska kokee ne rajoittaviksi. Vanhaa koodia ja toteutuksia kuitenkin käytetään merkittävässä määrin apuna uusien projektien pohjalla. Vanhojen toteutusten päälle on nopeampi rakentaa ja niiden toimivuus on todennettu, mikä vähentää yhteensopivuusongelmia ja muita jälkikäteen korjaamisen tarpeita. Toimintavuoden aikana deskin työn sujuvuus ja tekijöiden tekninen osaaminen ovat parantuneet, kun hyviä käytäntöjä ja erilaisten konseptien toteuttamista on opittu. Sovelluksissa ja visualisoinneissa käytettyjä koodeja myös pyritään jakamaan GitHub-sivustolla, mikä Tebestin mielestä sopii hyvin Ylen toimintaan. Toteutusten jakamisella on toinenkin tarkoitus: tehdä Plus-deskin työtapoja tutuksi ja mahdollistaa se, että tulevaisuudessa joku muukin voisi toteuttaa Ylen järjestelmässä toimivia sisältöjä.

Tähän mennessä Plus-deski on ostanut yhden datajournalistisen jutun ulkopuoliselta tekijältä, joka rakensi sen Haaga-Helia -ammattikorkeakoulun datajournalismikurssilla. Salmisen mukaan toteutuksen viimeisteleminen ja sovittaminen Ylen sivuille oli erittäin työlästä, ja tulevaisuudessa juttujen ostamista harkitaan tarkasti. Salminen pitää ulkopuolisten tahojen kiinnostusta hyvänä asiana ja merkinä siitä, että heidän työnsä tunnetaan. Ulkopuolista apua on ostettu esimerkiksi paikkatietoaineiston käsittelyssä. Heille on tarjottu yhteistyötä myös ukkosdatan käsittelyssä. Tulevaisuudessa deski voisi ostaa teknistä apua esimerkiksi koordinaattimuunnosten tekemiseen. Salminen suhtautuu työprosessin osien ulkoistamiseen kuitenkin melko skeptisesti: kommunikaatio on vaikeaa, jos ryhmä ei työskentele yhdessä. Tämä huomattiin esimerkiksi, kun Taloustutkimukselta tilattiin kyselydataa Arvokone-sovellusta varten. Toimituksen järjestelmän ja tavat tuntevien freelancereiden käyttöä voisi Salmisen mukaan harkita, mutta tällä hetkellä sellaisia ei ole.

7.3.1 Datan hankinta

Plus-deski hankkii projekteissaan käyttämänsä datan hyvin erilaisista lähteistä: viranomaisilta, verkkohauilla, skannatuista papereista, tai vaikka Wikipediasta, mikäli tietoa voi pitää luotettavana. Tebestin mukaan mikä tahansa voi olla datalähde. Yleensä datan saaminen viranomaisilta sujuu mutkattomasti. Syksyllä 2013 Plus-deski kuitenkin riitautti Verohallinnolta saamansa kielteisen päätöksen kaikkien suomalaisten verotietojen luovuttamisesta sähköisessä muodossa. Toimituksessa asia linjattiin periaatteelliseksi, ja datan saamisesta päätettiin taistella.

Data saadaan yleensä Excel-taulukkoina, kun se sellaisenaan pyydetään. Tebestin mukaan Excel-muotoisuuskaan ei kuitenkaan vielä tarkoita sitä, että data olisi yksikäsitteistä ja koneluettavaa, vaan taulukot voivat sisältää tietoa esimerkiksi värikoodauksen muodossa. Usein data muunnetaan lopullista analyysia ja visualisointia varten muuhun, kuin xls-muotoon, kuten csv-tiedostomuotoon. Excelistä data on kuitenkin suhteellisen helppo muuntaa haluamaansa formaattiin.

Joukkoistamista datan hankinnassa käytetään toisinaan. Joukkoistamisista tekee ongelmallisia etenkin se, ettei niissä ole minkäänlaista tilastollista otosta. Tebestin mukaan se toimii hyvänä datanhankintakeinona, kun kysytään mielipiteitä vaikkapa parhaista uimarannoista, mutta huonommin, kun joukkoistetun datan osoittamia asioita yritetään kertoa faktoina. Silloin väärinymmärrysten vaara on suuri, vaikka jutun yhteydessä esimerkiksi mainittaisiin, etteivät kyselyn tulokset ole yleistettäviä. Ruudun raavinnalla Plus-deski on hankkinut dataa muutaman kerran. Sen käyttöä ei rajoita niinkään tekninen monimutkaisuus, vaan se, ettei keinoa hyödyntäviä juttuideoita juurikaan keksitä.

7.3.2 Datan puhdistus ja analyysi

Datan siivous on datajournalistisen työprosessin aikaa vievin osuus. Usein esimerkiksi viranomaisilta hankitussa datassa on paljon ylimääräistä tietoa ja virheellisiä merkintöjä. Lisäksi eri kielialueilla syötetyssä datassa esimerkiksi desimaalipisteen ja -pilkun käyttö voi aiheuttaa ongelmia. Usein Tebest kirjoittaa datan puhdistusta varten algoritmin, joka etsii datasta tiettyjä ominaisuuksia tietyillä ehdoilla ja muuntaa ne haluttuun muotoon. Tällainen algoritmi saattaa muodostua vaikkapa muutamasta tarkoitukseen sopivalla ohjelmointikielellä kirjoitetusta säännöllisestä lausekkeesta. Joskus puhdistusta tehdään myös Excelillä tai OpenRefinellä. Satunnaisesti puhdistus vaatii myös käsityötä. Esimerkkitästä on eri automerkkien ja -mallien suosiota esittelevän laskurin 6 000-rivisen taustadatan puhdistus, joka vaatii tietämystä eri automallien ominaisuuksista.

Datan virheettömyydestä Plus-deskissä varmistutaan ensisijaisesti lähdekritiikin avulla. Esimerkiksi Tilastokeskuksen kaltaisesta viranomaislähteestä saadun datan voi olettaa olevan paikkansapitävää. Kaikissa dataprojekteissa kuitenkin pyritään ennen datan käyttämistä miettimään, onko siinä mitään järkeä (“sanity check”) ja kertooko se oikeasta asiasta. Valtavien data-aineistojen kokonaan tarkistaminen on mahdotonta, mutta Tebestin mukaan ainakin jutuissa mainitut luvut ja arvot tarkistetaan, eikä datasta esitetä väitteitä tarkistamatta niitä. Usein myös datan suurimman ja pienimmän arvon tarkistaminen näyttää, onko käytetty algoritmi toiminut oikein.

Eniten ongelmia datan puhdistuksessa aiheuttavat juuri kielten eroavat ominaisuudet, kuten erilaiset desimaalirajoittimet. Myös erilaiset suuren datamäärän aakkos- ja numerojärjestykseen järjestämiseen liittyvät ongelmat ovat aiheuttaneet vaikeuksia. Ratkaisuja kuitenkin keksitään sitä mukaa kun haasteita ilmenee. Tebest muistuttaa, että kaikki ongelmat on joku ratkaissut aiemminkin, ja tieto löytyy verkosta.

Useita datasettejä jutuissa yhdistetään jatkuvasti enemmän. Sitä Plus-deskissä pyritään tulevaisuudessa lisäämään entisestään, jotta päästäisiin pelkästä datan esittämisestä enemmän oikean analyysin suuntaan. Eri asioiden korrelaation ja kausaliteetin etsiminen tuottaisi oivaltavampaa journalismia.

7.3.3 Visualisointi tai muu lopputuote

Datan visualisointivaiheessa Plus-deskissä käytetään yleensä erilaisia valmiita HTML, CSS ja JavaScript -koodista muodostuvia komponentteja, kuten Googlen karttoja tai järjestettäviä listoja, joita voidaan yhdistellä ja muokata tarpeen mukaan. Joskus esimerkiksi erilaisissa laskureissa toteutus voidaan rakentaa alusta asti. Toteutuksissa pyritään käyttämään mahdollisimman paljon toiminnallisesti ja visuaalisesti samoja komponentteja, kuin aiemmissakin töissä. Tämän tarkoituksena on opettaa lukijat odottamaan tietyltä elementiltä tietynlaista toimintaa. Minkäänlaista visualisointiohjetta, kuten käytettyjen viivojen paksuudet ja värit sisältävää tyylikirjaa, Plus-deskillä ei ole, mutta Tebestin mukaan sellainen helpottaisi ja nopeuttaisi työtä.

Vaikka sovellukset ja visualisoinnit vetävät kävijöitä pidempään kuin tavalliset uutisjutut, niitä ei yleensä päivitetä, koska niiden lukijamäärä laskee hyvin nopeasti julkaisun jälkeen. Joskus sovellukset kuitenkin keräävät lukijoita pidemmänkin ajan kuluttua. Tästä esimerkki on painoindeksilaskuri, joka karusta toteutuksestaan huolimatta keräsi paljon lukijoita, ja sen ulkoasu päätettiin päivittää vielä pari viikkoa myöhemmin. Toisinaan samaa visualisointia, kuten sähkönsiirtoalueiden karttaa, kuitenkin käytetään uudestaan samaa aihepiiriä koskevassa jutussa. Myös Bostonin maratonin pommi-isku ja niiden jälkeisiä tapahtumia esittelevää visualisointia käytettiin uudestaan ja siihen lisättiin tapahtumia, kun uusia juttuja julkaistiin.

Reaaliaikaisesti päivittyvään, erilaisten ohjelmointirajapintojen kautta saatavaan dataan perustuvia sovelluksia Plus-deski ei ole tehnyt. Tebest muistuttaa, että päivittyvän sovelluksen kanssa julkaistu uutisjuttu vanhenisi nopeasti. Hän ei kuitenkaan sulje pois tällaisten palveluiden rakentamista tulevaisuudessa. Data-animaatioita deskissä ei ole tehty, mutta Salmisen mukaan animaation hyödyntämistä esityskeinona on kyllä suunniteltu.

Toteutuksessa käytetyt datasetit, varsinkin itse kerätty data, on pyritty julkaisemaan jutun mukana silloin kun se muistetaan ja siihen on aikaa. Plus-deskissä on kuitenkin pyrkimys liittää datan julkaisu kiinteämmin työprosessiin. Tebest tahtoisikin myös luoda paikan, josta lukija voisi löytää kaiken Ylen Uutisten jutuissa käytetyn datan. Toistaiseksi hän on

julkaissut datasettejä omassa blogissaan. Toimituksen hankkimien maksullisten datasettien jakamisesta Plus-deskillä ei ole ylemmän tahon linjausta, mutta Salmisen mukaan sellainen pitäisi tehdä. Tähän saakka maksullinen data on voitu jättää jakamatta, jos toimituksella on ollut tarkoitus julkaista sen pohjalta lisää uutisia.

7.3.4 Tuloerolaskuri

Vuoden 2012 verotietojen julkistamispäivää edeltävänä päivänä 31.10.2013 Yle Uutiset julkaisi laskurin, jonka avulla käyttäjä voi suhteuttaa omat tulonsa postinumeroalueensa tulojakaumaan. Otsikolla Kuinka moni naapurisi tienaa vähemmän kuin sinä? Kokeile itse laskurilla julkaistu verkkoartikkeli sisälsi laskurin lisäksi jutun, jossa käsiteltiin ja kontekstoitettiin mielenkiintoisimpia datasta tehtyjä havaintoja sekä siitä havaittuja ääripäitä. Veropäivän ennakkojulkaisusta päätettiin yhdessä taloustoimituksen kanssa jo syksyllä kaksi kuukautta ennen varsinaista veropäivää. Veropäivänä Yle Uutiset julkaisi myös listat eri alueiden eniten tienanneista ihmisistä. Näiden listojen nopeassa rakentamisessa Tebest käytti apuna Python-ohjelmointikielellä kirjoittamaansa ohjelmaa, joka järjesti ja siivosi verotoimistosta saadut verotietolistat haluttuun muotoon ja sijoitti ne järjestettäviin HTML-taulukoihin, jotka voitiin julkaista nopeasti (Tebest 2013).

Itse laskurin pohjadataa käytettiin Tilastokeskukselta saatuja vuoden 2011 tulotietoja postinumeroalueen mukaan. Kaikki tuloja saaneet täysi-ikäiset suomalaiset jaettiin kymmeneen luokkaan, joissa kaikissa on yhtä paljon ihmisiä. Tämän jälkeen joka postinumeroalueen tulonsaajat sijoitettiin näihin kymmeneen luokkaan. Joukosta rajattiin tietoturvan vuoksi pois postinumeroalueet, joilla asuu alle 100 täysi-ikäistä tulonsaajaa. Omat vuositulonsa ja postinumeronsa syöttämällä käyttäjä voi tarkastella alueensa tulojakaumaa ja nähdä, mihin noista Suomen kymmenestä tuloryhmästä kuuluu. Laskuriprojektin koordinoinnista ja datan hankinnasta vastasi tuottaja Salminen, toteutuksen rakensivat Tebest ja graafikko Juha Rissanen, ja aineiston hankki sekä jutun kirjoitti taloustoimittaja Johan Jaakkola. Valmista nettijuttua viimeisteltiin deskissä yhdessä, ja Jaakkola teki samasta aiheesta jutut myös tv- ja radiouutisiin.

Laskuri on hyvä esimerkki Plus-deski-vuorossa toteutetusta projektista, jossa toimittajalla oli riittävästi aikaa keskittyä tämänkaltaiseen laajaan kokonaisuuteen. Laskuria tehtiin kaikkine työvaiheineen yhteensä noin viikon ajan. Työläintä juttuprosessissa oli datan ymmärtäminen ja tarkistaminen, sekä oikeaan muotoon muokkaaminen. Plus-deskillä kuitenkin oli kirkas kuva siitä, millainen laskurista tahdottiin, eikä työprosessissa tapahtunut yllätyksiä.

Juttuun on linkitetty kaikki siitä tehdyt jatkojutut ja muut veropäivään liittyvät jutut. Laskurin Facebook- ja Twitter-jaettavuutta pyrittiin parantamaan tuomalla jakonapit selkeästi näkyviin kun laskurihaun tulos ilmoitetaan, sekä kertomalla jakotekstissä, minkä postinumeron tiedot jakaja on tarkistanut. Jakoteksti ei kuitenkaan ilmoita, mihin tuloluokkaan testin tekijä kuuluu. Salmisen mukaan vastaustietoja ei myöskään tallenneta. Juttu oli Yle Uutisten verkkosivuilla vuoden 2013 viidenneksi luetuin.

8. POHDINTA

Tutkimani datajournalismia tekevät ryhmät, Helsingin Sanomien datadeski, Svenska Ylen Grävnavet ja Ylen Plus-deski olivat kaikki erilaisia työtavoiltaan ja resursseiltaan, joten kaikkia toimituksia kuvaavaa mallia tai teoriaa ei mielestäni ole järkevää tavoitella. Joitakin yhteneviä piirteitä ja mielenkiintoisia havaintoja on kuitenkin syytä tuoda esille.

Ehkä perustavin havaintoni oli se, ettei missään toimituksessa tehdä pelkkää datajournalismia, ja kaikki luvussa 4 (s. 20) kuvaamani datajournalistisen työprosessin vaiheet sisältäviä juttuja julkaistaan itse asiassa melko vähän, mutta lähes kaikkien sisältöjen rakentamisessa käytetään jotakin datajournalismin työtappaa tai työkalua, kuten visualisointityökaluja. Myös datajournalistisen työprosessin osallistujien roolit sekoittuivat tutkittavissa toimituksissa ja niitä jaettiin projektin luonteesta riippuen. Lisäksi deskien jäsenet tekevät usein päällekkäin montaa eri projektia, ja uudet kiireisemmät toteutukset kiilaavat ajattomampien edelle.

Perimmäisenä tarkoituksena oli järjestelmällistää ja systematisoida verkkoekstrojen tekeminen, datajournalismi sanan varsinaisessa merkityksessä on osa sitä. Eli tämä ei oo datadeski sanan varsinaisessa merkityksessä. - - Jos puhutaan datajournalismista siinä merkityksessä että hankitaan tai kerätään datamassoja, analysoidaan ne ja visualisoidaan ne, niin se on yks osa. (Juho Salminen, tuottaja, Yle, Plus-deski.)

Helsingin Sanomien datadeski ja Ylen Plus-deski keskittyivät siis tuottamaan verkonomaista sisältöä, ja siitä vain yhden osan muodostaa datajournalismi. Suurin osa näiden deskien tuotannosta on muuta kuin datajournalismia. Grävnavet taas tekee tutkivaa journalismia, jossa data on vain yksi mahdollinen lähde, eikä sitä käytetä kaikissa projekteissa.

Toinen ennako-oletukseni oli se, että datajournalismiin läheisesti liittyvä teknologian kulttuuri tuo journalismin kulttuuriin erilaisia ajatuksia objektiivisuudesta, läpinäkyvyydestä, tiedon ja taustamateriaalin jakamisesta ja työtavoista (Parasie & Dagiral 2013; Royal 2010). Esittelen näistä asioista tekemiäni havaintoja seuraavissa työprosessia,

siihen liittyvää kommunikaatiota ja sen avoimuutta käsittelevissä luvuissa. Lopuksi luotaan kirjallisuuden ja aineistoni kautta datajournalismin tulevaisuudennäkymiä.

8.1 Työprosessi ja sen haasteet

Datajournalismitiimin perustaminen on ollut jokaiselle toimitukselle merkittävä panostus, joten sen suoriutumiselle on asetettu suuret odotukset. Kaikki haastatellut kertovat saavuttaneensa heille asetetut tavoitteet. Tämä tarkoitti esimerkiksi sitä, että tiimien luomat datajournalistiset sekä muut verkonomaiset jutut ovat toimitusten verkkosivuilla poikkeuksellisen suosittuja. Luvussa 7 (alkaen s. 78) esittelemäni projektit ovat hyviä esimerkkejä datajournalismiprojekteista, joiden eteen tiimi on nähnyt poikkeuksellisen paljon vaivaa, ja jutut ovat keränneet suuria näyttökertalukuja. Tutkimissani toimituksissa menestyneet jutut myös saavat lisänäkyvyyttä uutissivustoilla, mutta pelkkä datajournalismi ei kuitenkaan ole näkyvyyden jakamisessa itseisarvo.

Haastatellut kokevat, että heihin luotetaan ja että he ovat osoittaneet olevansa tuon luottamuksen arvoisia. Kaikissa toimituksissa korostuu esimiesten joustavuus paitsi työaikojen, myös työvälineiden ja tapojen suhteen. Työvälineinä haastatellut suosivat avoimuuden ihanteen mukaisesti avoimen lähdekoodin ohjelmia aina kun se vain on mahdollista. Datajournalismitiimit pystyvät myös melko hyvin vaikuttamaan toimitusten verkkojulkaisujärjestelmien kehittämiseen työhönsä sopivaksi, vaikkakaan muutoksia ei aina saada tehtyä halutulla vauhdilla. Joidenkin verkkojulkaisujärjestelmien ominaisuuksien lisäksi uutissovellusten kehittämistä hidastaa se, että ne täytyy saada toimimaan myös vanhoissa Internet Explorer -selaimissa. Tästä haastatellut toivovat pääsevänsä eroon. Merkittävä osa lukijoista kuitenkin edelleen käyttää vanhentuneita selaimia.

Kaikkien tutkittujen toimitusten datajournalismitiimit työskentelevät muita toimittajia joustavammin työajoin. Deadline-ajattelu yhdistettynä joskus yllättävänkin mutkikkaaseen tekniseen työprosessiin venyttää usein datajournalistin työpäivää, mutta toisaalta työajat joustavat myös toiseen suuntaan.

Koko prosessi, ollut yllätys paljonko se loppujen lopuksi vie aikaa. Kun on tullut ylityötä niin aika nopeesti pystytty ottamaan vapaata, työnjohto on ymmärtäväinen. Jos on ollut iso paketti ja ylityötä niin ei välttämättä seuraavalla viikolla niin iso paketti. (Peter Sjöholm, toimittaja, Svenska Yle, Grävnavet.)

Joskus haasteita aiheuttavat myös projektiin osallistuvien toimittajien erilaiset työajat ja useat päällekkäiset työtehtävät. Kaikki datajournalismiryhmät kokevat voivansa tehdä nykyisellä osaamisellaan ja miehityksellään laadukasta datajournalismia. Kuitenkin Helsingin Sanomat rekrytoi uutta koodarigraafikkoa ja Ylen Plus-deskissäkin odotetaan, että jossain vaiheessa tullaan palkkaamaan uusi ohjelmointitaitoinen työntekijä. Uudet rekrytoinnit tulevat nostamaan tuotanto-odotuksia entisestään. Teknisistä taidoista kaikki ryhmät kertoivat tarvitsevänsä lisää syvällistä tilastotieteen osaamista, ja myös paikkatieto-osaamisen tarve nousi esille. Tilasto-osaaminen auttaisi löytämään parempia tapoja datan tarkastamiseen ja analyysiin.

Ensimmäisinä vuosinaan datajournalismi on ollut erikoisuus ja sen julkaiseminen poikkeuksellista. Nyt kun datajournalismia ja interaktiivisuutta sisältäviä juttuja osataan julkaista kaikissa tutkituissa toimituksissa rutiininomaisesti, on toiveissa, että myös yleisö oppii odottamaan ja käyttämään niitä. Tämän vuoksi toimitukset pyrkivät käyttämään melko samankaltaisia elementtejä ja värejä kaikissa julkaisuissaan. Yleisön osallistamisen lisääminen oli kaikissa toimituksissa esillä: Helsingin Sanomissa pohdittiin verkkosivujen Mielipide-osaston vahvistamista ja nykyistä suurempaa käyttöä myös datan hankinnassa, Svenska Ylellä tahdottiin levittää onnistuneen Fixa högskolan -joukkoistusprojektin käytäntöjä myös muihin toimituksiin ja Ylen Plus-deski aloitteli Fixa högskolanin inspiroimaa Ajankohtaisen Kakkosen Peruskoulurempa-joukkoistusprojektia. Kaikki toimitukset pohtivat joukkoistamisen lisäämistä tulevaisuudessa, sen rajoitukset kuitenkin tiedostaen. Yleisöön tehtävien suorittajana ei voi täysin luottaa, eikä luotettavan määrällisen datan kerääminen joukkoistamalla ole helppoa. Laajan kyselyaineiston kerääminen on samalla markkinointia tulevalle jutulle. Esimerkiksi Helsingin Sanomien Luokkakonetta (ks. s. 55) varten kerätty 40 000 vastaajan aineisto varmasti loi sovellukselle valmiin innokkaan käyttäjäryhmän. Yleisön rooli sisältöjen levittämisessä yritetään maksimoida: suuremmissa datajournalistisissa projekteissa ja erilaisissa testeissä on panostettu

sosiaalisen median jaettavuuteen jo pitkään: käyttäjän vastausten mukaan personoituvat jakotekstit ja mahdollisimman houkuttelevasti asetellut jakolinkit ovat jo muodostuneet tärkeäksi osaksi projekteja.

Yleisön lisäksi oppimista toivotaan myös viranomaisilta, joilta dataa pyydetään. Viranomaisten puutteellinen julkisuuslain tuntemus vaikeuttaa edelleen yhteistyötä. Myös datan irrottamisesta ja luovuttamisesta pyydetävät maksut ovat joskus korkeita, mikä johtuu puutteellisista luovutuskäytännöistä, sekä siitä, että data voi olla paperimuodossa. Viranomaisten toiminnan yhtiöittäminen ja siirtäminen yksityisten tahojen hoidettavaksi vievät julkista dataa liikesalaisuuden piiriin. Näihin ongelmiin datajournalistit voivat yrittää varautua perusteellisella julkisuuslain tuntemuksella sekä suunnittelemalla tietopyynnöt huolella ja liittämällä niihin kohtia, joissa huomioidaan korkeiden maksujen kaltaiset erikoistapaukset.

Haastatellut datajournalismin tekijät eivät ole innostuneet datapohjaisten, päivitettävien palveluiden rakentamisesta. Missään toimituksessa myöskään vanhoja uutissovelluksia ja visualisointeja ei yleensä päivitetä, koska haastateltujen mukaan ne ovat usein päiväkohtaisia, eivätkä kerää merkittävästi näyttökertoja enää julkaisuajankohdan etääntyessä.

Datajuttu ilmestyy jonain päivänä ja jos sen tiedot vanhentuu, ne saa vanhentua. Joudun välillä sanomaan, että entä sitten jos tiedot on vanhentuneet. Se on julkaistu sinä ja sinä päivänä. Ei kukaan oletakaan. Mutta tällainen ajatus, että me tehtäisiin hirveä määrä palveluja ja joita jatkuvasti päivitetään, siitä mä olen yrittänyt ja päässytkin eroon. (Esa Mäkinen, uutispäällikkö, Helsingin Sanomat, datadeski.)

Joitakin poikkeuksia kuitenkin on tehty, kun vanhat jutut ovat jostain syystä alkaneet saada taas suosiota. Vanhat työt kuitenkin nousevat uudelleen esiin uusia kehitettäessä: osia vanhoista töistä ja vanhaa koodia käytetään projekteissa apuna mahdollisimman paljon.

8.2 Kommunikaatio

Kaikki haastatellut pitävät toimivan kommunikaation vuoksi tärkeänä, että datajournalismiryhmän työpiste on muun toimituksen yhteydessä. Helsingin Sanomien datadeski ja Ylen Plus-deski onkin sijoitettu uutiskoneiston ytimeen, Svenska Ylen Grävnavet taas työskenteli hiukan sivussa muusta toimituksesta, mutta heidänkin toiveissaan on tulevaisuudessa siirtyä keskeisemmälle paikalle.

Jokaisessa toimituksessa datajournalismia ja verkkokerrontaa suunnittelevat työntekijät joutuvat teknisinä osaajina neuvomaan muuta toimitusta paljon myös yksinkertaisissa asioissa, kuten verkkojulkaisujärjestelmän käytössä tai taulukkolaskennan perusteissa. Kukaan haastatelluista ei kuitenkaan koe sitä häiritsevänä, vaan se on kaikkien mielestä osa heidän työtään. Kysymyksiin vastaaminen ja auttaminen myös madaltaa muun toimituksen kynnystä tehdä yhteistyötä datajournalismiryhmän kanssa. Ryhmät ovat olemassaolonsa aikana myös tulleet tunnetummiksi omien toimitusorganisaatioidensa sisällä, ja haastateltujen mielestä toimittajat osaavat aiempaa paremmin lähestyä heitä datapohjaisten ideoiden kanssa. Kuitenkin edelleen on myös toimittajia, jotka eivät tiedä, mitä datajournalistit tekevät ja miten heidän apuaan voi käyttää.

Kaikki datajournalismiryhmät toivovat tulevaisuudessa voivansa tehdä enemmän suuria projekteja ja jättää rutiininomaisia töitä, kuten yksinkertaisia visualisointeja toimittajien itse tehtäväksi. Helsingin Sanomien Esa Mäkisen kehittämä visualisointityökalu Arkku (ks. s. 54) on ehkä suurin konkreettinen askel tätä kohti. Myös erilaisten ilmaistyyökalujen, kuten Googlen visualisointityökalujen, käyttötaito yleistyy toimittajien keskuudessa. Kaikki datajournalismiryhmät ovat myös järjestäneet toimituksissa työvälineitä ja työtapoja esitteleviä koulutuksia.

Toinen on se, että me pystytään tulevaisuudessa entistä enemmän keskittymään, että tavallaan ne hommat, jotka on ollut meidän duuneja alussa, joku karttojen värkkääminen ja sellainen, niin ne arkipäiväistyy pikku hiljaa. Tällaisen erikoisjoukon panosta ei välttämättä tarvita puolentoista vuoden päästä semmoiseen, johon vielä tarvittiin vuosi sitten. Ne perusvisualisointityökalut vaikkapa tai jotkut taulukot ja listat datalle niin ne tulee osaksi

arkipäivää enemmän, jolloin Plus-deskin tyyppinen operaatio voi keskittyä siihen, mikä silloin on uutta. (Juho Salminen, tuottaja, Yle, Plus-deski.)

Toimittajien yleinen tekninen osaaminen on vielä melko heikolla tasolla ja datajournalismiryhmien työskentelyä nopeuttaisi esimerkiksi se, että toimittajat osaisivat käyttää paremmin Exceliä. Tärkein toimittajilta toivottava ominaisuus on kuitenkin se, että he ymmärtäisivät paremmin datajournalismin ja verknonomaisten kerrontatapojen mahdollisuudet ja osaisivat siten suunnata ideointiaan, tuleehan suuri osa ideoista - Ylen Plus-deskissä jopa lähes kaikki - datajournalismiryhmän ulkopuoliselta toimitukselta. Moniin suomalaista datajournalismia tunnetuksi tehneisiin sovelluksiin on haettu mallia muiden maiden medioista, ja omien ideoiden määrää haluttaisi kasvattaa. Esimerkiksi ruudun raavinnan käyttö datan hankintakeinona ei ole tällä hetkellä harvinaista siksi, että se olisi teknisesti ylitsepääsemättömän vaikeaa, vaan siksi, että sen mahdollisuuksia ei osata ajatella juttuja ideoitaessa.

-- ihmisten toimintatavat tai siis semmoisten ihmisten, jotka ei oo tehnyt tämmöistä koskaan aikaisemmin, niin ne on ehkä kiinnostuneita, ja niiltä saatetaan ehkä vaatia, että mä näin Ylellä, että niillä on klikattava tsydeemi ja mekin haluttais, ja sitten joku graafikko tai mediatoimittaja, tän tyylinen hahmo, on ehkä vähän kiinnostunut, mutta ei oikein osaa, niin ei tommoisella päivän kurseilla ratkaista sitä, että susta tuleekin JavaScript-koodari. (Teemo Tebest, atk-toimittaja, Yle, Plus-deski.)

Huippukoodariksi muuttumisen sijasta datajournalistit toivovat yhteistyökykyä ja intoa tarttua uusiinkin työkaluihin. Vaikka datajournalismin tekijöiden työ on luonteeltaan teknistä, he eivät koe, että heidän olisi vaikeaa keskustella siitä muun toimituksen kanssa. Haastatellut korostavat toimittajataustaansa ja taitoaan pukea asiat ymmärrettävälle kielelle. Ainoa puhtaasti teknisen taustan omaava tekijä, Ylen Teemo Tebest, korostaa, että kuka tahansa ohjelmoija ei toimitustyöhön sovi, vaan taito kommunikoida selkeästi on välttämätön. Toimituksellista työtä tuntematon tekninen henkilö saattaisi törmätä väärinymmärryksiin. Uutistemmalli tehtävä sovellusten kehittäminen eroaa tavallisesta verkkokehittämisestä, mikä näkyy joskus myös yhteistyössä verkkokehitystiimien kanssa.

Että he esimerkiksi ajattelee että visualisointi ei oo koskaan valmis, koska he rakentaa näitä nettisivujen rakennetta, kun me tehdään juttuja, mut jos uutistoimittajan kannalta kattoo niin visualisoinnin on oltava valmis kun on deadline, että se ikä ei oo niin pitkä ja se datakin

vanhenee, et se ei oo niin että kahden vuoden kuluttua joku saa hirveesti kun lukee, ellei sitä arkistokontekstissa nosteta esille. Siinä on erilaisia ajattelutapoja, mutta niistä on hyvä keskustella, niin aina päästään vähän eteenpäin. (Eva-Maria Koskinen, tuottaja, Svenska Yle, Grävnavet.)

Keskustelu ja osaamisen jakaminen nähdään kaikissa välttämättöminä keinoina ammatillisten muurien ylittämiseksi ja datajournalismin kehittämiseksi.

8.3 Avoimuus

Kuten kirjallisuuden perusteella saatoinkin olettaa, suomalaistenkin datajournalismin tekijöiden kiinnostus avoimeen dataan näkyy myös haluna julkaista toimitustyössä käytettyjä data-aineistoja silloin, kun ne ovat itse hankittuja, muokattuja tai yhdisteltyjä. Suoraan alkuperäismuodossa käytetyn virastodatan julkaisua ei pidetä niinkään olennaisena, kunhan datan lähde on mainittu jutussa selkeästi.

Kaikki haastatellut pitävät datasettien julkaisua tärkeänä, mutta tuo ajatus ei kuitenkaan aina kannata käytäntöön asti, vaan kiire ja jonossa odottavat projektit syrjäyttävät avoimuuden periaatteen. Joskus data saattaa kaivata vielä siistimistä ennen julkaisua, tai siitä täytyy poistaa esimerkiksi henkilötietoja tai postinumerotietoja, eikä tälle ole aikaa. Esimerkiksi Helsingin Sanomien Luokkakoneessaan käyttämä data sisälsi postinumerotietoja.

Toisinaan julkaisu taas yksinkertaisesti unohtuu. Myös toimitukselliset syyt ajavat joskus avoimuusajattelun ohi.

Joskus me on tehty päätös, että jotain dataa ei oo haluttu julkaista siksi, että päätetty että me tehdään vielä jatkojuttuja siitä itse. Tällöiset on vähän harvinaisempia tapauksia. (Juho Salminen, tuottaja, Yle, Plus-deski.)

Haastatellut kuitenkin suunnittelevat tulevaisuudessa kiinnittävänsä enemmän huomiota datan julkaisemiseen. Ylen Teemo Tebest pohtii jopa, että tulevaisuudessa kaiken Ylen

uutisjutuissaan käyttämän datan voisi koota yhdelle sivustolle. Tuon sivuston pitäisi olla Yle Uutisten verkkosivun yhteydessä, eikä erillisenä siilona, koska muuten normaali lukija ei datan pariin löytäisi. Tällä hetkellä hän on koonnut joitakin käyttämiään datasettejä henkilökohtaiseen blogiinsa ja jakanut kirjoittamaansa koodia GitHub-sivustolla. Helsingin Sanomat on julkaissut joitakin viime vuosina käyttämiään datasettejä HS Next -blogissaan.

Avoimuuden kulttuuri näkyy myös tavassa, jolla kaikki datajournalistit käyttävät StackOverflown kaltaisia verkko yhteisöjä apunaan teknisten ongelmien ratkaisemisessa. Avoimuutta viestii myös halu muiden auttamiseen ja tiedon, mikä korostuu kaikissa haastatteluissa.

8.4 Tulevaisuus

Suomessa datajournalismia tekevät säännöllisesti vain suuret mediatalot, joilla on resurssien tuomaa rohkeutta kokeiluihin. Näissä toimituksissa myös uskotaan datajournalismin roolin kasvavan tulevaisuudessa. Helppokäyttöiset ilmaistyoökalut eivät ole ainakaan vielä demokratisoineet datajournalismia. CAR:in tulo toimituksiin noudatti aikoinaan teknologisen innovaation leviämisen s-käyrää (Garrison 1998). Sama kuvio näyttäisi toistuvan myös datajournalismissa: kun teknologian ja toteutustapojen hinta laskee ja osaamisen määrä kasvaa, myös taloudellisesti heikommat mediat pääsevät käsiksi uuteen teknologiaan, ja lopulta seuraa leviämisen taantuma. Datajournalismin kohdalla kyse ei ole niinkään valtavista laitteistohankinnoista, vaan riittävän osaavan työvoiman saamisesta toimituksiin. Säännöllinen ja tehokas datajournalismin tuottaminen vaatii erilaista osaamista kuin perinteinen uutistyö, eikä moni mediatalo ole heikossa taloustilanteessa valmis siihen investoimaan. Kun datajournalismin tuotantotavat kehittyvät ja toimittajat oppivat siihen vaadittuja taitoja, datajournalismi yleistyne myös pienemmissä toimituksissa.

8.4.1 Kannattaako datajournalismi?

Tulevaisuudessakin datajournalismin suurin haaste on taloudellinen kannattavuus. Erilaiset rahastot, kuten Knight Foundation voivat rahoittaa yksittäisiä projekteja ja innovaatiotyötä (Cohen, Hamilton & Turner 2011), mutta toimitukset tarvitsevat pitkäaikaisempia tulovirtoja. Verkkojournalismillekaan ei ole löydetty täysin toimivaa liiketoimintamallia, joten viime kädessä toimitusten datajournalismipanostuksiin vaikuttaa se, tuovatko ne lisätuloja vai ainoastaan uuden menoeran. Tällä hetkellä datajournalistiset jutut hyödyttävät mediataloja lähinnä tuomalla uutissivustoille kävijöitä, eikä sekään aina onnistu (Aitamurto 2011, 14). Tutkimissani suomalaistoimituksissa suuret datajournalismipanostukset kuitenkin useimmiten näkyvät myös suurina kävijämäärinä. On tietenkin otettava huomioon, että Yle ei julkisen palvelun toimijana ole kaupallisen median tavoin kytkeyty ansaintamallien etsimiseen.

Lorenz (2010) muistuttaa, että informaation vaihtaminen rahaan ei ole häviämässä ja mainitsee hyvinä esimerkkeinä uutistoimistot Thomson Reutersin ja Bloombergin. Molemmat tekevät voittoa tarjoamalla esimerkiksi laadukasta talousdataa eri alojen ammattilaisten käyttöön, mikä voisi olla osa datajournalisminkin tulevaisuutta. Aitamurron (2011) mukaan moni mediatalo suhtautuukin datajournalismin tulevaisuuteen toiveikkaasti: vaikka sitä tällä hetkellä tehdään enemmän journalistisen arvon vuoksi, sillä uskotaan vielä saatavan tulojakin.

Datajournalismi ei ole pelkkä menoera. Tulevaisuudessa datajournalismissa ja tietojärjestelmätieteessä käytetyt välineet ja algoritmit voivat tehostaa suurten datamassojen läpikäymistä ja siten tehdä toimituksilta odotetun yhteiskunnallisesti merkittävän raportoinnin edullisemmaksi. Myös esimerkiksi tilastotieteen, lääketieteen, oikeustieteen ja yhteiskuntatieteiden aloilla kehitetyt teknologiat, ohjelmistot ja algoritmit voivat sellaisinaan tai pienen muokkaamisen jälkeen sopia toimitustyön tehostamiseen. (Cohen, Hamilton & Turner 2011.)

Parhaimmillaan visualisoinnit ja uutissovellukset kasvattavat uutissivustojen kävijävirtoja. Suomalaistoimituksissa on myös havaittu, että lukijat viettävät jopa puolet pidempiä aikoja

datajournalistisen ja interaktiivisen sisällön, kuin muiden juttujen äärellä. Interaktiivisuutta lisäämällä tätä aikaa voitaisiin kasvattaa entisestään ja sitten veloittaa mainostajilta enemmän rahaa tällaisten sisällön yhteyteen myytävistä mainoksista. Myös premium-malli, eli laadukkaan datajournalistisen sisällön tekeminen maksulliseksi, voisi olla keino ansaita sillä. (Aitamurto 2011, 15.) Tulevaisuudessa ansaintamahdollisuuksia tarjoavat myös palvelut, joissa perustason toiminnot ovat ilmaisia, mutta lisämaksusta käyttäjä voi saada juuri hänelle räätälöityä tietoa. Tällaisten palveluiden käyttäjiä voisivat olla vaikkapa yritykset tai kodinostajat ja kiinteistövälittäjät. (Lorenz, Kayser-Bril & McGhee 2011.) Suomalaisoimitusten datajournalistit eivät ole vielä kuitenkaan innostuneet palveluiden kehittämisestä, koska he eivät usko pitkän tähtäimen kävijämäärien olevan palvelun ylläpitämisen aiheuttaman vaivan arvoisia.

Yhtenä tulevaisuuden ansaintakeinona on tarjottu myös mediatalojen muuntumista luotetuiksi datakeskuksiksi (Lorenz ym. 2011; Aitamurto 2011, 15). The Guardian-lehti toimi tässäkin edelläkävijänä ja ilmoitti vuonna 2011 tavoitteekseen kerätä Datablogiinsa omissa jutuissaan käyttämiensä datasettien lisäksi valtioiden avointa dataa kaikkialta maailmasta. Toimittaja Simon Rogers ilmoitti silloin, että The Guardian tahtoo tarjota ihmisille helpon ja luotettavan paikan löytää dataa avoimen tiedon paljoudesta (Vehkoo 2011, 109). Lehti on sittemmin kyllä kartuttanut datakatalogiaan, mutta hallitusten avaaman datan varastoksi tarkoitettu World government data -osio on hävinnyt sivulta. Datajournalistit Lorenz, Kayser-Bril ja McGhee uskovat, että medially on paljon saavutettavaa tällaisten datalla luottamusta luovien palveluiden saralla. Tällä hetkellä mediayritykset pyrkivät ansaitsemaan verkossa vetämällä ihmisten huomiota verkkomainontaan. Valtavan informaatiomäärän keskellä huomiomarkkinoilla kilpaileminen ei kuitenkaan enää ole kannattavaa kuin harvoille. Bloombergin ja Thomson Reutersin kaltaiset toimijat ovat siirtyneet kilpailemaan luottamusmarkkinoilla, jonne perinteisenkin median olisi siirryttävä selvitätseen. (Lorenz ym. 2011.) Tulevaisuudessa mediayhtiöt voisivat myös tuotteistaa kartuttamaansa dataosaamista ja myydä data-analyysipalveluita yrityksille ja organisaatioille (Aitamurto 2011, 15).

8.4.2 Datajournalismin ulkoistaminen ja ulkopuolinen apu

Koska datajournalismi tuo ainakin aloitusvaiheessa mediayhtiöille lisäkuluja, on sen työläimpien työvaiheiden, tai jopa koko prosessin, ulkoistamista pohdittu. Eräs näkyvä liikahdus ulkoistamisen suuntaan oli datajournalismia toimituksille myyvän Journalism++-yrityksen perustaminen vuonna 2011. Minna Rapeli (2013) ennustaa gradussaan, että tulevaisuudessa toimitusten omien datajournalismitiimien merkitys vähenee ja datajournalismin julkaiseminen tasa-arvoistuu, kun kaikki mediatalot voivat ostaa sen palveluna ulkopuolisilta yrityksiltä. Kun toimitukset jatkuvasti ulkoistavat osia toiminnastaan, näin onkin helppo ajatella. Rapelin haastattelema Ylen datajournalisti Teemo Tebest nostaa kuitenkin esiin työprosessin osien ulkoistamisen aiheuttamat kommunikaatiovaikeudet, jotka saattaisivat heikentää työn laatua. Varsinkin uutisrytmissä tuotettava nopea datajournalismi vaatii sujuvaa ja esteetöntä kommunikaatiota.

Yle Uutiset ja Helsingin Sanomat ovat jo ostaneet joitakin valmiita datajournalistisia tuotteita ulkopuolisilta. Kokemukset eivät ole kovin hyviä, sillä ulkopuolisen tekemiä toteutuksia on vaivalloista saada toimimaan toimituksen järjestelmissä.

Jos tekee etättyötä sellaisen tyyppin kanssa, joka ei tunne talon tapoja, niin se on vaikeeta tai siihen menee paljon aikaa siltä työn tilaajalta. Tää huomattiin siinä yhdessä proggiksessa, jossa ostettiin teksti ja visu friikkuryhmältä, joka teki opiskelijakurssilla tätä, niin sen ois tehnyt itse nopeemmin sen homman, kun siinä piti säätää ja neuvotella ja editoida ja tekninen toteutus piti kuitenkin Teemon vielä tehdä meidän standardeihin sopivaksi -- (Juho Salminen, tuottaja, Yle, Plus-deski.)

Tähän mennessä suomalaiset mediatalot ovatkin harvoja poikkeuksia lukuun ottamatta tuottaneet itse datajournalisminsa. Kokonaisten sovellusten ostamista ulkopuolisilta ei pidetä mahdollisena, mutta myyjän tulisi olla tietoinen siitä, missä muodossa tuote täytyy toimittaa, jotta ostaja ei joutuisi näkemään vaivaa muokkaamisessa. Helsingin Sanomien uutispäällikkö Esa Mäkinen pohtii mahdollisuutta ostaa uutissovellusten pohjia, joita toimitukset voisivat muokata ja käyttää omien sovellustensa luomiseen. Helsingin Sanomat on ostanut toimituksen ulkopuolelta myös erilaisten suurten projektien ohjelmointityötä sekä datan puhdistusta.

Yksi joka on toiminut on toi HS-raadin tekeminen. Se on toiminut hyvin, se on ostettu ulkopuolelta. Siinä on backendi, admini ja generoitavat uutissovellukset, joita tulee joka viikko ja rajapinnat ja muut. Sit tietenkun niinkun, ollaanhan me verokoneen koodaustyö ostettu ulkopuolelta. Nyt kun tehdään vaalikone ja tulospalvelu niin nekin ostetaan ulkopuolelta niiden koodaaminen ja sit mä suunnittelen konseptin ja Japi leiskat ja joku ulkopuolinen meille koodaa sitten. (Esa Mäkinen, uutispäällikkö, Helsingin Sanomat, datadeski.)

Ainakin journalistista havainnointia sisältävät työvaiheet, joissa jutun idea voi kehittyä, on syytä pitää ydinryhmän käsissä. Ulkoistamisajatus ei suinkaan ole uusi, siitä on puhuttu jo CAR:in kulta-aikoina. Paul Adrian (1999, 13) kritisoi aikanaan tietokoneavusteisen journalismin tiedonkäsittelyosien ulkoistamista:

Kuinka usein olet oikeasti päätenyt tekemään saman jutun, jonka aloitit? Aloitan tietyn juttuidean kanssa, mutta kun teen pohjatyötä, juttu usein muuttuu aivan toisenlaiseksi. Jos annat työn jollekulle muulle, hän tekee juuri sen, mitä pyysit häneltä, mutta hän ei tarkastele tietoja toimittajan silmin. Tuloksena voit jäädä paitsi jutusta.

Ulkopuolisen avun merkitystä ei kuitenkaan ole syytä väheksyä. Esimerkiksi yliopistoista löytyy tilastotieteen ja yhteiskuntatieteiden osaajia, joilta datajournalisti voi saada apua erikoisosaamista vaativiin työvaiheisiin (Bradshaw 2013, 40). Hyvä esimerkki tästä on Helsingin Sanomien Luokkakone, jonka rakentamiseen datadeski sai apua Turun yliopiston professori Jani Erolalta (ks. s. 55). Tämän kaltaisten yhteistyöverkostojen luominen myös mediamaailman ulkopuolelle voi tulevaisuudessa rikastaa ja syventää datajournalismia.

8.4.3 Millaista datajournalismia tulevaisuudessa tehdään?

Tuotantomallien muodostumisen lisäksi datajournalismin tulevaisuutta määrittävät teknologian ja työtapojen kehitys, sekä käyttökelpoisen datan entisestään paraneva saatavuus. Nämä muutokset ovat sysänneet eteenpäin myös rakenteettoman datan tulkittamisesta. Markkinointiyritykset, hallitukset, yhteiskuntatieteilijät ja muut tiedonkäsittelyn ammattilaiset ovat viime vuosina panostaneet innovatiivisiin algoritmeihin, jotka auttavat verkon tietovirtojen käsittelyssä (Cohen, Hamilton & Turner 2011). Myös toimittajien roolin ennustetaan muuttuvan, kun aiempaa suuremman osan tiedonhankinnasta ja yksinkertaisimpien juttujen kirjoittamisesta voi hoitaa tehtävään

suunniteltujen algoritmien avulla (Bell 2012). Algoritmeja käyttää esimerkiksi Narrative Science -yritys, joka tarjoaa palvelua, jossa datasta, kuten vaikkapa urheilutuloksista, muodostetaan tekstejä.

Perinteisessäkin mediassa on opittu hyödyntämään automaattista datan käsittelyä. Esimerkiksi Los Angeles Times on automatisoinut maanjäristysuutisten ja pikkurikosten käsittelyn sekä luonut arkistonsa ruokajutuista reseptitietokannan käyttäen oppivaa algoritmia, joka löytää ja merkitsee tekstistä tiettyjä ominaisuuksia. Median esittämät vaatimukset auttavat myös viranomaisia oppimaan datan luovuttamisen: LA Times sai rikoskarttansa päivittymään reaaliajassa vasta, kun lehti sai suostuteltua poliisin antamaan rikostiedotteensa tietyssä muodossa olevana syötteenä. (Welsh 2013.) Jatkuvasti päivittyvät uutisapplikaatiot varmasti kasvattavat suosiotaan, kun viranomaiset siirtävät dataansa saatavaksi helppokäyttöisten ohjelmointirajapintojen kautta. Joitakin esimerkkejä tästä on nähty Suomessakin. Helsingin Sanomat hyödynsi julkaisemassaan jatkuvasti päivittyvässä ruuhkakartassa kolmannen osapuolen luomaa Tieinfo API -palvelua, jonka kautta pääkaupunkiseudun liikennedata oli saatavissa helppokäyttöisemmässä muodossa kuin missä Liikennevirasto sitä tarjosi (Mäkinen 2012).

Tulevaisuudessa paitsi viranomaiset, myös järjestöt ja yksityishenkilöt tuottavat ja julkaisevat enemmän journalistiseen käyttöön soveltuvaa dataa. Sen vuoksi datajournalistien onkin hyvä suhtautua dataan kuin muihinkin uutisten raaka-aineisiin ja luoda järjestelmä, jonka avulla tärkeiden ja mielenkiintoisten datasettien julkaisusta saadaan tieto yhtä nopeasti kuin muistakin merkittävistä uutistapahtumista (Bradshaw 2013, 37 - 38). Esimerkki tulevaisuuden nopeista datalähteistä ovat sensoriverkot, jotka koostuvat kansalaisjärjestöjen käyttämistä ilmanlaatua valvovista laitteista (Bell 2012). Myös henkilökohtainen data, kuten etukorttiostoksista kerättävä tieto, tarjoaa tulevaisuudessa datajournalisteille uusia mahdollisuuksia.

Jos voitais tehdä sillä tavalla, että hae Keskolta omat ruokaostosdatasi meille, niin me näytämme kuinka terveellistä elämää sä elät. Sehän siis tarjoaisi aivan huikeita mahdollisuuksia visualisoinneille ja ettei tarvitsisi enää keskustella siitä, että syötkö sä terveellisesti vai. Kun sä ostat tällaista niin nähtävästi sä et syö kauheen terveellisesti. (Esa Mäkinen, uutispääällikkö, Helsingin Sanomat, datadeski.)

Datajournalismin tekijöiden on tulevaisuudessa kiinnitettävä lisää huomiota dataan liittyviin eettisiin näkökulmiin. Kahden julkisen datasetin yhdistämisestä voi syntyä uutta, vaikkapa liian syvälle yksityisyyden alueelle tunkeutuvaa dataa. Myös datan hallinnan ja omistajuuden kysymykset tulevat merkittävimmiksi datan määrän kasvaessa, eikä näitä asioita koskevaa lainsäädäntöä vielä välttämättä ole, joten toimittajien ja toimitusten oma harkinta korostuu (Bell 2012). Toimittajan tehtäviin kuuluu tulevaisuudessa myös näyttää yleisölle, millaista dataa erilaiset yritykset ja instituutiot heistä keräävät, ja kuinka arkaluontoista se voi olla (Belaega 2013, 26).

Kun datajournalismi yleistyy, myös heikkolaatuisia ja eettisesti ongelmallisia juttuja tullaan julkaisemaan yhä enemmän (Howard 2013). Kaikkea saatavissa olevaa dataa ei suinkaan kannata julkaista. Hyvä esimerkki tästä on tapaus, jossa newyorkilainen *The Journal News* julkaisi kartalla kaikkien Westchesterin ja Rocklandin piirikunnissa asuvien pistooliluvan haltijoiden nimet ja osoitteet. Erityisen tulenaraksi julkistamisen teki se, että se tehtiin Sandy Hookin kouluampumisten jälkeen, ja julkaisu herättikin vastustusta, joka johti aselupatietojen avoimuuden heikentämiseen useissa osavaltioissa (Bartzen Culver 2013). Bartzen Culverin mukaan lehden olisi pitänyt pelkän datan julkaisun sijasta tarkastella sitä kriittisesti, yhdistellä sitä muuhun dataan ja käyttää sitä tarkemman raportoinnin lähteenä. Datan laadun tarkastamiseenkaan lehti ei ollut panostanut, sillä aselupadata todettiin myöhemmin myös monelta osin virheelliseksi (Howard 2013). Tämä on hyvä esimerkki siitä, että toimittajan rooli ei suinkaan ole häviämässä tai muuttumassa merkityksettömäksi. Pelkän datan julkaisu saattaa jopa luoda hämmennystä sen sijaan, että se auttaisi yleisöä ymmärtämään ympäröivää maailmaa. Lorenz (2011) nimeää tämän yhdeksi datajournalismin tulevaisuuden suurimmaksi haasteeksi ja huomauttaa, että sensaatiohakuisesti ja ilman perusteellista taustoitusta julkaistu data ja luvut toimivat, jos media kilpailee huomiomarkkinoilla, mutta kun kilpaillaan luottamuksesta, huono datajournalismi kääntyy tekijöitään vastaan. Osa luottamuksen tavoittelua on siirtynyt kohti juttuja, jotka eivät ole pelkkää datan esittämistä, vaan porautuvat syvemmälle yhteiskunnan rakenteisiin antaen lukijalle oikeasti uutta tietoa

ja oivalluksia. Toimituksissa on tahtoa aiempaa kunnianhimoisemman datajournalismin tekemiseen.

Ennen se oli skuuppi, että meillä oli kartta -- sit se muuttu jossain vaiheessa et me havainnoitiin itsekin kartasta jotain. Sit seuraava taso ois varmaan se, että meillä ois jotain tutkivaa. (Teemo Tebest, atk-toimittaja, Yle, Plus-deski.)

Datajournalismi on syntynyt tietojärjestelmätieteen, tilastotieteen, yhteiskuntatieteen ja journalismin kohdatessa, ja tämän monialaisuuden tunnustaminen ja vahvistaminen on tärkeää tulevaisuudessakin. Cohenin, Hamiltonin ja Turnerin (2011) mukaan syvällisemmän ja tehokkaamman datajournalismin kehittämiseksi on teknisten haasteiden lisäksi voitettava kulttuuriset ristiriidat, joita eri alojen osaajien kohdatessa väistämättä syntyy. Eri alojen osaamisen lisäksi toimitukset voisivat hyötyä muiden alojen käyttöön kehitetyistä ohjelmistoista (Cohen ym. 2011). Tällä hetkellä toimituksissa käytetyt työkalut ja julkaisujärjestelmät toimivat usein tekstiartikkeleiden ehdoilla ja sopivat huonosti datajournalismin sisältöjen julkaisuun ja käyttökelpoisten tietokantojen muodostamiseen omista jutuista (Lorenz 2010). Suomalaiset datajournalistit kuitenkin kokevat voivansa ainakin jonkin verran vaikuttaa julkaisujärjestelmien kehitykseen ja voivansa täysin itsenäisesti päättää käyttämistään työkaluista. Tulevaisuudessa toimitusjärjestelmätkin siis muokkautunevat enemmän datajournalismin julkaisemista ja ehkä jopa tietokantojen luomista tukeviksi.

Yleisön merkittävä rooli datajournalismin tuottajana ja tuotteiden levittäjänä tunnustetaan sekä Suomessa että maailmalla. Tulevaisuudessa datajournalismin tekijöiden on silti tärkeää tarkentaa juttujensa lähestymistapoja sen mukaan, millaista yleisöä he palvelevat (Hurrel & Walton 2013, 24). Svenska Ylen Grävnavet näytti mallia kohdeyleisön tavoittamisesta Fixa högskolan -projektillaan (ks. s. 64), jossa onnistunutta joukkoistamista edelsivät huolellinen suunnittelu ja kontaktien luominen kohderyhmän avainhenkilöihin. Datajournalistit näkevät tärkeänä myös uutissovellusten personoitavuuden, mahdollisuuden tuoda vaikeatkin uutisaiheet yksittäisen lukijan tasolle. Siksi laskuri tai testi, joka näyttää suurten poliittisten päätösten tai muiden uutistapahtumien vaikutukset yksilötasolla, onkin usein parempi tapa esittää tietoa kuin perinteinen visualisointi (Hurrel & Walton 2013, 22).

Datajournalismia on tähän saakka esitetty lähinnä verkon mahdollistamin ilmaisukeinoin, mutta mikäli monimediallisesti toimivat mediatalot jatkavat dataan panostamista, myös muiden välineiden ilmaisua on kehitettävä. Datajournalismin esittämistä televisiossa ja radiossa on Suomessa pohdittu ainakin Yle Uutisissa ja Svenska Ylen uutisissa. Tv:n ja radion uutis- ja ajankohtaisformaatteihin datan esittämistä on vaikea sovittaa tyylikkäästi ja dataa onkin käytetty lähinnä juttujen tai ajankohtaiskeskusteluiden innoittajana ja taustatietona sekä esitetty tapausesimerkkien kautta. Ylen toimittajat ovat myös esitelleet uutissovelluksia suorissa uutislähetyksissä. Tämä voi tulevaisuudessa auttaa tuomaan datajournalismin pariin myös yleisöä, joka ei juurikaan käytä verkon uutissivustoja (Hurrel & Walton 2013, 23). Tulevaisuudessa tv-uutisten datailmaisun keinoihin voisi lisätä myös esimerkiksi hollantilaisessa Netherlands from Above -ohjelmassa käytetyt data-animaatiot. Lehden ja verkon yhteistyöstä hyvä esimerkki on Helsingin Sanomien Luokkakone (ks. s. 55), jonka näyttävä markkinointi Kuukausiliitteen samaa teemaa käsitelleessä jutussa varmasti toi sovellukselle uutta yleisöä.

8.4.4 Datajournalismikoulutus

Tutkittuaan New York Timesin datajournalisteja vuonna 2010 Cindy Royal totesi, että datajournalismi tulee asettamaan haasteita myös media-ammattilaisia tuottaville koulutusaloille. Koulut reagoivat alan muutokseen hitaasti, koska on vaikeaa löytää osaavaa ja kiinnostunutta opetushenkilökuntaa (Royal 2010).

Vuonna 2011 erilaisia tietotekniikkaosaamista ja journalismia yhdisteleviä maisteriohjelmiä järjestivät jo ainakin Columbian ja Northwesternin yliopistojen journalismikoulut sekä Georgia Institute of Technology (Cohen, Hamilton & Turner 2011). Sen jälkeen datajournalismin maisteriohjelmiä on käynnistetty ainakin Tilburgin yliopistossa Hollannissa ja Rey Juan Carlosin yliopistossa Madridissa. Viime vuosina datajournalismiopetuksessa merkittäviksi ovat nousseet myös erilaiset avoimet verkkokurssit, joita järjestävät esimerkiksi Teksasin yliopiston Knight Center for Journalism in the Americas ja European Journalism Centre.

Datajournalismikouluttaja ja avoin data -aktiivi Antti Poikolan (2013) mukaan ensimmäiset datajournalismista kiinnostuneet suomalaiset olivat opiskelleet taitoja itsenäisesti jo kauan ennen alan verkostojen syntyä. Sosiaalisen median kautta datajournalismiverkostoja kuitenkin alkoi syntyä, ja tällä hetkellä aiheesta kiinnostuneet keskustelevat esimerkiksi useissa Facebook-ryhmissä. Alussa aiheesta kiinnostuneet hakivat tietoa verkosta lähinnä datajournalismin edelläkävijöiden pääosin englanninkielisiltä sivuilta, mutta vuoden 2011 tienoilla suomalaisiakin datajournalismia käsitteleviä blogeja ja verkkosivuja alkoi ilmaantua. (Poikola 2013) Oppilaitoksista ensimmäisen datajournalismikurssin järjesti Otavan Opisto syksyllä 2011. Perässä seurasivat Tampereen yliopisto keväällä 2012 ja Jyväskylän yliopisto sekä Haaga-Helia -ammattikorkeakoulu keväällä 2013. (Wikiopisto 2013.) Tampereen yliopiston ja Haaga-Helian kursseilla toimittajaopiskelijat työskentelivät yhteistyössä ohjelmoijien kanssa.

European Journalism Centre (EJC) teki vuonna 2011 toimittajille kyselyn datajournalismikoulutuksen tarpeesta. Toimittaja ja datajournalismikouluttaja Mirko Lorenz (2011) tulkitsee kyselyn tulosten perusteella, että koulutusta tarvitaan, ja että kouluttajien täytyisi panostaa perusasioiden, kuten tilastotieteen perusteiden opettamiseen. Hän myös korostaa, ettei jokaisen toimittajan välttämättä kannata opetella datajournalismin teknisten prosessien osaajaksi, vaan tärkeämpää on oppia toimimaan tiimeissä ja soveltamaan toimittajan taitoja ja ominaisuuksia datajournalismiajatteluun. Toimittajan on tärkeää myös oppia aggregoimaan eri lähteistä saatavaa tietoa ja organisoimaan muiden työtä esimerkiksi joukkoistamisprojekteissa (Lorenz, Kayser-Bril ja McGhee 2011). Haastattelemani datajournalismin tekijät ovat sitä mieltä, että datajournalismi on liian monimutkainen asia opittavaksi lyhyillä kursseilla, vaan sen oppiminen vaatii oikeaa kiinnostusta ja halua kokeilla asioita käytännössä.

Jos haluaa ns. oikeaa ymmärrystä niin pitäisi melkein olla joku sivuainekokonaisuus että voi keskittyä siihen. En mä luule, että kahdeksan opintopisteen datajournalismikurssi tulee antamaan paljoa vaatimuksia tehdä sitä oikeesti. Saa nähdä, jos koulut sattuu löytämään tarpeeksi innostuneita opiskelijoita, että voi tehdä kokonaisen sivuainekokonaisuuden datasta -- (Peter Sjöholm, toimittaja, Svenska Yle, Grävnavet.)

Myös Royalin (2010) haastattelemat koodaritoimittajat korostavat, että tärkeimmät datajournalismin parissa työskentelevältä toimittajalta vaaditut ominaisuudet ovat uteliaisuus, oppimiskyky ja ongelmanratkaisukyky. Samaa toistavat myös suomalaiset datajournalistit.

Mun mielestä tärkein taito mikä datajournalistilla on, (on se, että) osaa itse hankkia työkaluja ja opetella niiden työkalujen käyttöä ja tarvittaessa myös tehdä niitä työkaluja itse. En tiedä miten sellaista voisi opettaa -- (Esa Mäkinen, uutispäällikkö, Helsingin Sanomat, datadeski)

Ajatusta tukee myös Los Angeles Timesin datajournalisti Ben Welsh (2013), jonka toimituksessa tietotekniikkaosaajat on hajautettu eri osastoille ja he muodostavat muuttuvia tiimejä datajournalismista kiinnostuneiden ja oppimishaluisten toimittajien kanssa. Kaikki eivät silti voi olla datajournalismitiimissä pelkässä toimittajan roolissa, vaan ohjelmointitaitoisille, journalismista kiinnostuneille osaajille on kysyntää, eikä heitä ainakaan Suomessa toistaiseksi juuri ole.

Ala haluisi liikkua nopeemmin, kun mitä on tekijöitä. Esimerkiksi se puhe, että täällä haluttaisi rekrytää ihmistä, mutta meillä ei oo mitään hajua, kuka se voisi olla, eikä kauheen hyvää luottoakaan, että semmosta tyyppiä voisi löytyykään. -- Ei semmoisia ainakaan tuolta koulusta yhtäkään rupee putkahtele tuollee, et ”nyt niitä on kuin sieniä sateella” -tyyppisesti. (Teemo Tebest, atk-toimittaja, Yle, Plus-deski.)

9 LOPUKSI

Tutkimukseni tarkoitus oli selvittää, millainen on datajournalismin työprosessi Helsingin Sanomissa ja Ylen suomen- ja ruotsinkielisissä toimituksissa, ja millaisia haasteita datajournalistit työssään kohtaavat. Esittelen havaintojani luvussa 7 (s. 48) ja syvennän niitä luvussa 8 (s. 77). Tutkimuksessani selvitin myös, kuinka datajournalismia voidaan tulevaisuudessa toimituksissa edistää ja millaisia kehityssuuntia on nähtävissä. Tämä pohdinta löytyy luvusta 8.4 (s. 84). Tarkoitukseni ei ollut luoda teoriaa, vaan tuottaa uutta tietoa juuri näiden toimitusten käytännöistä juuri tässä ajassa, ja mielestäni onnistuin tehtävässäni.

Pyrin tekemään tutkimuksestani mahdollisimman läpinäkyvää kirjoittamalla auki tieteenfilosofiset lähtökohdat ja ajatukseni tiedon rakentumisesta vuorovaikutuksessa, sekä laadullisen tutkimuksen ominaisuudet tiedon tuottamisessa (s. 38). Koska tutkijan ominaisuudet vaikuttavat merkittävästi tiedon muodostumiseen, esittelin myös tutkimukseni motiivit ja lähtökohdat (s. 35) Perustelin mahdollisimman tarkasti etnografisen lähestymistapani sekä metodivalintani, joihin liittyy metodisen triangulaation rooli haastattelun ja havainnoinnin heikkouksien voittamisessa (s. 38). Olen myös pyrkinyt tiedostamaan ja kuvaamaan muut tiedonkeruuseen ja analyysiin vaikuttavat tekijät, joita ovat haastateltavan ja haastattelijan ammatillisen statuksen erot, sekä työstään asiantuntijana haastateltaville ihmisille ominainen tapa kuvata oman organisaationsa toiminta mahdollisimman järkevänä ja loogisena. Tiedonkeruuta rajoitti myös lyhyt kahden päivän havainnointiaika ja se, että tein havainnointijaksot juuri ennen joululoman alkua, eikä toimitusten työ ole silloin tyypillisimmillään. Tuloksia koostaessani ja analysoidessani olen pyrkinyt varmistumaan havaintojeni oikeellisuudesta hakemalla niille vahvistusta alan kirjallisuudesta ja kokeneiden datajournalistien teksteistä, sekä antamalla haastateltavieni lukea tutkielmani ennen sen julkaisua.

Datapaljous ympärillämme kasvaa ja mediayritykset siirtävät resurssejaan verkkosisältöjen tuotantoon, joten lisätutkimus suomalaisesta datajournalismista on ehdottomasti tarpeen.

Koska tutkimukseni lähestymistapa datajournalismiin ja sen tulevaisuuteen oli laadullinen ja laaja, uskon sen auttavan jatkotutkimuksen tekijöitä valitsemaan syvennettäviä aiheita. Tälläkin hetkellä Suomessa on tekeillä muutamia muitakin datajournalismia käsitteleviä pro gradu -tutkielmia, joiden tekijöiden käyttöön olen tarjonnut materiaalini, jotta he voisivat pureutua valitsemiinsa aihepiireihin mahdollisimman tarkasti. Datajournalismin alueella perusteellisempaa tutkimusta pitäisi tehdä ainakin teknisten osaajien ja toimittajien kommunikaatiosta. Esimerkiksi Helsingin Sanomien Esa Mäkisen luomaa Arkku-työkalua ja sen toimittajien keskuudessa saamaa vastaanottoa voisi tutkia datajournalismin työkäytäntöjen leviämisen selvittämiseksi. Myös onnistuneen joukkoistuksen salaisuus sekä televisiossa ja radiossa toimiva datajournalismi kaipaavat vielä perusteellista katsausta. Ehkä tärkein olisi kuitenkin tutkimus, joka määrittelisi kattavasti, mitä kaikkea datan käyttö journalismissa oikeasti on.

KIRJALLISUUS

- Aalto, T. 2005. Suomalaismedian ensimmäiset 22 vuotta verkossa. Yle. Saatavilla: <http://blogit.yle.fi/tuija-taalla-hei/suomalaismedian-ensimmaiset-22-vuotta-verkossa>. [Viitattu 21.1.2014].
- Aarnos, E. 2001. Kouluun lapsia tutkimaan: havainnointi, haastattelu ja dokumentit. Teoksessa: Aaltola, J. Valli, R (toim.). Ikkunoita tutkimusmetodeihin I. Jyväskylä: Gummerrus Kirjapaino Oy.
- Adrian, P. 1999. Put Your Career in the Fast Lane with CAR. Teoksessa Poynter Institute for Media Studies. When Nerds and Words Collide: Reflections on the Development of Computer Assisted Reporting. Saatavilla <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf> [Viitattu 10.9.2013].
- Aisch, G. 2012. Using Data Visualization to Find Insights in Data. Teoksessa. Gray, J. Chambers, L. Bounegru, L.(toim.). The Data Journalism Handbook. O'Reilly Media. Saatavilla: <http://datajournalismhandbook.org/1.0/en/index.html>. [Viitattu 10.9.2013].
- Aitamurto, T. Lewis, S. C. 2012. Open Innovation in Digital Journalism: Examining the Impact of Open APIs at Four News Organizations. New Media & Society. Saatavilla: <http://nms.sagepub.com/content/15/2/314>. [Viitattu 5.2.2014].
- Aitamurto, T. 2011. Data journalism strategies in the US and elsewhere. Teoksessa Aitamurto, T. Sirkkunen, E. Lehtonen, P. Trends In Data Journalism. Next Media. Saatavilla: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf. [Viitattu 9.10.2013].
- Alasuutari, P. 2011. Laadullinen tutkimus 2.0. Tampere: Vastapaino.
- Bartzen Culver, K. 2013. Where the Journal News Went Wrong in Mapping Gun Owners. PBS. Saatavilla: <http://www.pbs.org/mediashift/2013/02/where-the-journal-news-went-wrong-in-mapping-gun-owners053/>. [Viitattu 2.12.2013].

- Belaega, T. 2013. In a decade or less, could not all reporters be required to be data literate? Teoksessa Mair, John; Keeble, Richard Lance; Bradshaw, Paul; Belega, Teodora. Data journalism - Mapping the future. Suffolk: Abramis academic publishing.
- Bell, E. 2012. Journalism by numbers. Columbia Journalism Review, Sep/Oct2012, Vol. 51 Issue 3, 48-49.
- Bellinger, G. Castro, D. Mills, A. 2004. Data, information, knowledge, and wisdom. Systems Thinking. Saatavilla: <http://www.systems-thinking.org/dikw/dikw.htm> [Viitattu 3.2.2014].
- Beta.yle.fi. 2014. Saatavilla: <http://beta.yle.fi/>. [Viitattu 4.2.2014].
- Bradshaw, P. 2011a. The inverted pyramid of data journalism. Online Journalism Blog. Saatavilla: <http://onlinejournalismblog.wordpress.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>. [Viitattu 30.10.2013].
- Bradshaw, P. 2011b. 6 ways of communicating data journalism (The inverted pyramid of data journalism part 2). Online Journalism Blog. Saatavilla: <http://onlinejournalismblog.com/2011/07/13/the-inverted-pyramid-of-data-journalism-part-2-6-ways-of-communicating-data-journalism/>. [Viitattu 30.10.2013].
- Bradshaw, P. 2012a. Model for the 21st Century Newsroom – Redux. How digitisation has changed news organisations in a multiplatform world. Leanpub.
- Bradshaw, P. 2012b. What Is Data Journalism? Teoksessa. Gray, J. Chambers, L. Bounegru, L.(toim.). The Data Journalism Handbook. O'Reilly Media. Saatavilla: <http://datajournalismhandbook.org/>. [Viitattu 28.10.2013].
- Bradshaw, P. 2013. Data journalism workflow: Confronting the myths. Teoksessa Mair, J. Keeble, R. L. Bradshaw, P. Belega, T. (toim.). Data journalism - Mapping the future. Suffolk: Abramis academic publishing.
- Bradshaw, P. 2013. Data Journalism Heist. Leanpub.
- Jensen, K. B. 2002. The qualitative research process. Teoksessa Bruhn Jensen, K (toim.). A handbook of media and communication research: qualitative and quantitative methodologies. London: Routledge.

- Bounegru, L. 2012. Data Journalism in Perspective. Teoksessa. Gray, J. Chambers, L. Bounegru, L.(toim.). The Data Journalism Handbook. O'Reilly Media. Saatavilla: <http://datajournalismhandbook.org/>. [Viitattu 28.10.2013].
- Chesbrough, H. 2010. Open Services Innovation. Rethinking Your Business to Grow and Compete in a New Era. 2nd Edition. Hoboken: Jossey-Bass.
- Creswell, J. W. 2013. Research design: qualitative, quantitative, and mixed methods approaches. Los Angeles: Sage.
- Cohen, S. Hamilton, J. T. Turner, F. 2011. Computational journalism. Communications of the ACM, 54 (10). 66-71. Saatavilla: <http://cacm.acm.org/magazines/2011/10/131400-computational-journalism/fulltext>. [Viitattu 28.10.2013].
- Cohen, S. Li, C. Yang, J. Yu, C. 2011. Computational Journalism: A Call to Arms to Database Researchers. Saatavilla: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.7178&rep=rep1&type=pdf> [Viitattu 28.10.2013].
- Cox, M. 2000. The Development of Computer Assisted Reporting. A paper presented to the Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium, March 17 - 18, 2000, University of North Carolina, Chapel Hill. Saattavilla: <http://pdf.thepdfportal.com/?id=105496&nocache#>. [Viitattu 29.10.2013].
- Daniel, A. Flew, T. 2010. The Guardian reportage of the UK MP expenses scandal: a case study of computational journalism. In Record of the Communications Policy and Research Forum 2010, Network Insight Pty. Ltd., Sydney, pp. 186-194. Saatavilla: <http://eprints.qut.edu.au/39358/1/39358.pdf>. [Viitattu 7.11.2013].
- Davenport, L. D. Fico, F. DeFleur, M. H. 2002. Computer-assisted Reporting in Classrooms: A Decade of Diffusion and a Comparison to Newsrooms. Journalism & Mass Communication Educator, Spring 2002, Vol. 57 Issue 1. 6-22.
- Domingo, D. 2003. Ethnography for new media studies: A field report of its weaknesses and benefits. Paper presented at the annual LAMCR Conference, Porto Allegre,

Brazil. Saatavilla: http://makingonlinenews.net/docs/domingo_nrn2003.pdf. [Viitattu 19.11.2013].

Domingo, D. 2005. The difficult shift from utopia to realism in the Internet era. A decade of online journalism research: theories, methodologies, results and challenges. Paper presented at the First European Communication Conference, Amsterdam, the Netherlands. http://makingonlinenews.net/docs/domingo_amsterdam2005.pdf. [Viitattu 19.11.2013].

Domingo, D. 2008. *Inventing Online Journalism: A Constructivist Approach to the Development of Online News*. Teoksessa Domingo, D. Paterson C. (toim.) *Making Online News: The Ethnography of Media Production*. New York: Peter Lang Publishing.

Erola, J. 2014. Näin yhteiskuntaluokat luotiin. Jani Erolan blogi. Saatavilla: <http://janierola.net/2014/01/03/miten-yhteiskuntaluokat-luotiin/>. [Viitattu 17.1.2014].

Eskola, J. Vastamäki, J. 2001. *Teemahaastattelu: opit ja opetukset*. Teoksessa Aaltola, J. Valli, R. 2001. *Ikkunoita tutkimusmetodeihin I*. Jyväskylä: Gummerrus Kirjapaino Oy.

Filloux, F. 2013. Data Journalism is improving – fast. *The Guardian*. Saatavilla: <http://www.theguardian.com/technology/blog/2013/jul/01/data-journalism-awards-improving>. [Viitattu 30.10.2014].

Flew, T. Spurgeon, C. L. Daniel, A. Swift, A. 2010. The promise of computational journalism. *Journalism Practice*, 6:2, 157-171. Saatavilla: <http://www.tandfonline.com/doi/pdf/10.1080/17512786.2011.616655>. [Viitattu 28.10.2013].

Garrison, B. 1998. Newspaper size as a factor in use of computer-assisted reporting. Paper presented to the communication technology and policy division of the Association for Education in Journalism and Mass Communication, Baltimore, MD, 5–8 August. Saatavilla: <http://com.miami.edu/car/baltimore1.htm>. [Viitattu 2.12.2013].

Gray, J. Chambers, L. Bounegru, L.(toim.). *The Data Journalism Handbook*. O'Reilly Media. Saatavilla: <http://datajournalismhandbook.org/1.0/en/index.html>.

- Grönfors, Martti. 2001. Havaintojen teko aineistonkeräyksen menetelmänä. Teoksessa Aaltola, J. Valli, R. 2001. Ikkunoita tutkimusmetodeihin I. Jyväskylä: Gummerrus Kirjapaino Oy.
- Heinonen, A. 1999. Journalism In The Age of The Net: Changing Society, Changing Profession. Tampere: Tampereen yliopisto. Väitöskirja.
- Hirsjärvi, S. Hurme, H. 2011. Tutkimushaastattelu. Teemahaastattelun teoria ja käytäntö. Helsinki: Gaudeamus.
- Hirst, T. 2013. Several Takes on Defining Data Journalism. School of data. Saatavilla: <http://schoolofdata.org/2013/06/11/several-takes-on-defining-data-journalism/>. [Viitattu 9.10.2013].
- Hurrel, B. Walton J. 2013: How does that affect me? Making data personally relevant for your audience. Teoksessa Mair, J. Keeble, R. L. Bradshaw, P. Belega, T. (toim.). Data journalism - Mapping the future. Suffolk: Abramis academic publishing.
- Houston, B. (1996). Computer-assisted reporting. New York: St. Martin's Press.
- Howard, A. 2013. On the Ethics of Data-Driven Journalism: of fact, friction and public records in a more transparent age. Tow Center for Digital Journalism. Saatavilla: <http://towcenter.org/blog/on-the-ethics-of-data-driven-journalism-of-fact-friction-and-public-records-in-a-more-transparent-age/>. [Viitattu 3.12.2013].
- Huttunen, L. 2010. Tiheä kontekstointi: Haastattelu osana etnografista tutkimusta. Teoksessa Ruusuvuori, J. Nikander, P. Hyvärinen, M (toim.) Haastattelun analyysi. Tampere : Vastapaino.
- Jaakkola, J. Tebest, T. Rissanen, J. 2013: Kuinka moni naapurisi tienaa vähemmän kuin sinä? Kokeile itse laskurilla. Yle Uutiset. Saatavilla: http://yle.fi/uutiset/kuinka_moni_naapurisi_tienaa_vahemman_kuin_sina_kokeile_its_e_laskurilla/6908390. [Viitattu 23.1.2014].
- Kinnari, T. 2013. Open Data Business Models for Media Industry: Finnish Case Study. Aalto University School of Business. Pro gradu -tutkielma. Saatavilla: https://aaltodoc.aalto.fi/bitstream/handle/123456789/10140/hse_thesis_13166.pdf?sequence=1. [Viitattu 3.2.2014].

- Kirk, A. 2012. The 8 hats of data visualisation design. Visualising Data. Saatavilla: <http://www.visualisingdata.com/index.php/2012/06/article-the-8-hats-of-data-visualisation-design/>. [Viitattu 30.10.2013]. Koskinen, E. Tebest, T. 2012. Här säljer alkoholdrycker bäst. Svenska Yle. Saatavilla: <http://svenska.yle.fi/artikel/2012/06/01/har-saljer-alkoholdrycker-bast>. [Viitattu 20.1.2014].
- Kundra, V. 2011. Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect. Joan Shorenstein Center on the Press, Politics and Public Policy. Discussion Paper Series. Saatavilla: http://shorensteincenter.org/wp-content/uploads/2012/03/d70_kundra.pdf. [Viitattu 3.2.2014].
- Kunelius, R. 2009. Viestinnän vallassa. Helsinki: WSOYpro Oy.
- Kuusisto, P. Sirkkunen, E. 1999. Journalismi uuden kynnyksellä. Tampereen yliopisto, tiedotusopin laitos, journalismin tutkimusyksikkö. Tampere: Tampereen yliopistopaino.
- Kuutti, H. 2011. Julkisuusjournalismi. Jyväskylä: Bookwell Oy.
- Lehtonen, Pauliina. 2011. Open data in Finland. Public sector perspectives on open data. Next Media. Saatavilla: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2._Hyperlocal_Open%20data%20in%20Finland.pdf. [Viitattu 3.2.2014].
- Lorenz, M. 2010. Status and Outlook for data-driven journalism. Data-driven journalism: What is there to learn? A paper on the data-driven journalism roundtable held in Amsterdam on 24 August 2010. Saatavilla: http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf. [Viitattu 30.10.2013].
- Lorenz, Mirko (2011): Training data driven journalism: Mind the gaps. Datadrivenjournalism.net. Saatavilla: http://datadrivenjournalism.net/news_and_analysis/training_data_driven_journalism_mind_the_gaps. [Viitattu 30.10.2013].
- Lorenz, M. Kayser-Brill, N. McGhee G. 2011. Media Companies Must Become Trusted Data Hubs. Owni.eu. Saatavilla: <http://owni.eu/2011/02/28/media-companies-must-become-trusted-data-hubs-catering-to-the-trust-market/>. [Viitattu 4.10.2013].

- Lorenz, M. 2012. Why Journalists Should Use Data. Teoksessa Teoksessa. Gray, J. Chambers, L. Bounegru, L.(toim.). The Data Journalism Handbook. O'Reilly Media. Saatavilla: <http://datajournalismhandbook.org/1.0/en/index.html>. [Viitattu 10.9.2013].
- McGregor, S. E. (2013): A Brief History of Computer-Assisted Reporting. Columbia Journalism School, Tow Center For Digital Journalism. Saatavilla: <http://towcenter.org/blog/a-brief-history-of-computer-assisted-reporting/>. [Viitattu: 28.10.2013].
- Merrin, W. 2009. Media Studies 2.0: upgrading and open-sourcing the discipline. Studies in Communication and Culture Volume 1 Number 1. Intellect Ltd. Saatavilla: <http://clairemacme4.files.wordpress.com/2011/05/w-merrin-we-need-media-2-01.pdf>. [Viitattu 7.11.2013].
- Meyer, P. 1999. The Future of CAR: Declare Victory and Get Out! Teoksessa Poynter Institute for Media Studies. When Nerds and Words Collide: Reflections on the Development of Computer Assisted Reporting. Saatavilla <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf> [Viitattu 10.9.2013].
- Meyer, P. 1991. The New Precision Journalism. The University of North Carolina Chapel Hill. Saatavana: <http://www.unc.edu/~pmeyer/book/Chapter1.htm>. [Viitattu 10.9.2013].
- Miller, T. 1988. The data-base revolution. Columbia Journalism Review, Sep/Oct88, Vol. 27 Issue 3. 35-38.
- Morris, D. 1999. Driving Defensively Through Data's Dangers. Teoksessa Poynter Institute for Media Studies. When Nerds and Words Collide: Reflections on the Development of Computer Assisted Reporting. Saatavilla <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf> [Viitattu 10.9.2013].
- Mäkinen, E. 2014. Selvitä Kuukausiliitteen Luokkakoneella, mihin yhteiskuntaluokkaan kuulut. Helsingin Sanomat. Saatavilla: <http://www.hs.fi/datajournalismi/a1305759784361>. [Viitattu 17.1.2014].
- Nygren, G. Appelgren, E. Hüttenrauch, H. 2012. Datajournalistik – ett växande område. Nordicom-Information 34 (2012) 3-4, 81-88.

- Open Knowledge Foundation. 2014. Open Definition. Saatavilla: <http://opendefinition.org/>. [Viitattu 3.2.2014].
- Paterson. C. 2008. Introduction: Why Ethnography? Teoksessa Domingo, D. Paterson C. (toim.) Making Online News: The Ethnography of Media Production. New York: Peter Lang Publishing.
- Parasie, S. Dagiral, E. 2013. Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media & Society*, Sep2013, Vol. 15 Issue 6. 853-871. Saatavilla: <http://nms.sagepub.com/content/15/6/853.full>. [Viitattu 9.10.2013].
- Paul, N. 1999. Assessing 10 Years of Computer Assisted Reporting. Teoksessa Poynter Institute for Media Studies. When Nerds and Words Collide: Reflections on the Development of Computer Assisted Reporting. Saatavilla <https://s3.amazonaws.com/s3.documentcloud.org/documents/757701/nerds-and-words.pdf> [Viitattu 10.9.2013].
- Pietilä, V. Ridell, S. 2008. Verkkomedia toimijuuden alustana: yleisö, yhteisö, julkiso ja YouTube. *Lähikuva* 21(2): 27-43. Saatavilla: <http://elektra.helsinki.fi.ezproxy.jyu.fi/se/1/0782-3053/21/2/verkkome.pdf>. [Viitattu 7.11.2013].
- Poikola, A. Kola, P. Hintikka, K. A. 2010. Julkinen data. Helsinki: Edita Prima. Saatavilla: www.scribd.com/doc/28845102/Julkinen-data. [Viitattu 3.2.2014].
- Poikola, A. 2013. Miten datajournalismin opetus tuli Suomeen. *Datajournalismi*. Saatavilla: <http://datajournalismi.fi/>. [Viitattu 3.12.2013].
- Puijok, R. 2008. Ethnographic Media Production Research in a Digital Environment. Teoksessa Domingo, D. Paterson C. (toim.) Making Online News: The Ethnography of Media Production. New York: Peter Lang Publishing.
- Rapeli, M. 2013. Data journalism: an outlook for the future processes. Aalto-yliopisto. Kauppakorkeakoulu. Tietojärjestelmätiede. Pro gradu -tutkielma. Saatavilla: <http://epub.lib.aalto.fi/fi/ethesis/id/13341>. [Viitattu 30.10.2013].

- Reavy, M. 1996. How the media learned computer-assisted reporting. Paper presented to the Southeast Colloquium, Newspaper Division, Association for Education in Journalism and Mass Communication. Roanoke, VA.
- Rogers, S. 2011. Data journalism broken down: what we do to the data before you see it. The Guardian Datablog. Saatavilla: <http://www.theguardian.com/news/datablog/2011/apr/07/data-journalism-workflow>. [Viitattu 30.10.2013].
- Rogers, S. 2011. The first Guardian data journalism: May 5, 1821. Guardian Datablog. Saatavilla: <http://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian>. [Viitattu 18.9.2013].
- Royal, C. 2010. The journalist as programmer: A case study of the New York Times interactive news technology department. In: Paper presented at the international symposium in online journalism, University of Texas, Austin, TX, 23 April. Saatavilla: <https://online.journalism.utexas.edu/2010/papers/Royal10.pdf>. [Viitattu 9.10.2013].
- Ruusuvuori, J. Nikander, P. Hyvärinen, M. 2010. Haastattelun analyysin vaiheet. Teoksessa Ruusuvuori, J. Nikander, P. Hyvärinen, M (toim.) Haastattelun analyysi. Tampere : Vastapaino.
- Sirkkunen, E. 2011. Overview: Public data and data journalism in some countries. Teoksessa Aitamurto, T. Sirkkunen, E. Lehtonen, P. Trends In Data Journalism. Next Media. Saatavilla: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf. [Viitattu 9.10.2013].
- Svenska Yle. 2013. Fixa högskolan. Saatavilla: <http://svenska.yle.fi/fixa-hogskolan>. [Viitattu 20.1.2013].
- Tebest, T. 2013. Datajournalistin veropäivä. Datajournalismi-blogi. Saatavilla: <http://datajournalismi.blogspot.fi/2013/12/datajournalistin-veropaiva.html/>. [Viitattu 24.1.2014].
- Tuomi, J. Sarajärvi, A. 2009. Laadullinen tutkimus ja sisällönanalyysi. Helsinki: Kustannusosakeyhtiö Tammi.

Varto, J. 1992. Laadullisen tutkimuksen metodologia. Tampere: Tammer-Paino Oy 1996.

Vehkoo, J. 2011. Painokoneet seis! Kertomuksia uuden journalismin ajasta. Vaajakoski: Bookwell Oy 2012.

Väliverronen, E (toim.). 2011. Journalismi murroksessa. Helsinki: Gaudeamus.

Warren, C. A. B. 2001. Qualitative interviewing. Teoksessa Gubrium J. F. Holstein J. A. (toim.) Handbook of interview research. Thousand Oaks: Sage.

Welsh, B. 2013. What is a Data Desk? Konferenssiesitys 22.11.2013. Ruotsi, Huddinge, Södertörns högskola.

Wikiopisto. 2013. Datajournalismi. Saatavilla:
<http://fi.wikiversity.org/wiki/Datajournalismi>. [Viitattu 3.12.2013].

LIITE

1 Tausta

Työnimike

Mitä työnkuvaasi kuuluu?

Kauan työskennellyt datajournalismin parissa?

Koulutus?

2 Datajournalismi toimituksessanne

Mistä datajournalismi teidän toimituksessanne alkoi?

Onko toimituksessanne käytetty aiemmin ns. CAR-työtapoja?

Miksi teillä lähdettiin tekemään datajournalismia? kilpailu, laatu, tehokkuus?

Miten suosittuja datajournalistiset jutut sivuillanne ovat?

Viettävätkö ihmiset niiden parissa enemmän aikaa kuin tavallisten juttujen?

Kuinka deskinne toiminta on sen olemassaoloaikana kehittynyt?

Onko työhön syntynyt deskinne toiminnan aikana nopeuttavia rutiineja? Millaisia esimerkiksi?

Työskentelettekö toimituksen ulkopuolisten ihmisten kanssa? (freet etc)

Keitä deskiinne kuuluu? Mitkä heidän roolinsa ovat?

Keitä toimituksessanne datajournalismin parissa työskentelee?

Montako ihmistä yhden jutun tekoon usein osallistuu?

Kuka johtaa juttuprojektia?

Käytetäänkö projektin hallitsemisen apuna erityisiä työkaluja? (Keskustelutiloja, versionhallintaa jne)

Mitä tyypilliseen työpäiväänne kuuluu?

Millaisia datajournalismiprojekteja yleensä teette? (Tutkivia/päivänpäällisiä/isoja/pieniä)

Millaisia aiheita yleensä?

3 Työprosessi

Alkusysäys

Mistä aloite datajournalistisiin juttuihin tulee?

Mistä hankitte dataa

Millaisia työkaluja/ohjelmointikieliä käytätte työprosessin alkuvaiheessa (screippaus jne)

Datan hankinnan suurimmat haasteet?

Datan puhdistus ja käsittely

Osallistuuko deskin ulkopuolinen toimitus tähän työvaiheeseen?

Kuinka usein yhdistätte useampia datasettejä?

Millaisilla keinoilla varmistatte, että data on mahd. virheetöntä?

Työkalut?

Haasteet?

Visualisointi/muu lopputuote

Työkalut?

Haasteet?

Käytättekö valmiita templateja visualisoinneissa?

Liitättekö juttuihinne niissä käytetyt datasetit? Miksi/miksi ei?

Kuinka valmis datajournalistinen juttu pyritään saamaan yleisön luettavaksi? (Jakaminen, muissa välineissä promoaminen)

Päivitetäänkö appeja/visuja? Palataanko niiden pariin millään tavalla?

4 Haasteet

Mitkä ovat suurimmat haasteet työprosessissa?

Löytyykö deskistänne riittävät tekniset taidot haluamannelaisen datajournalismin toteuttamiseen?

Millaisia teknisiä taitoja tiiminne mielestäsi tarvitsisi?

Paljonko voitte vaikuttaa työkaluihin itse?

Onko työkalujen käytössä rajoituksia?

Joudutteko käyttämään toimitustyökaluja, jotka eivät suoraan sovellu työhönne?

Onko teillä mielestäsi riittävästi aikaa datajournalismin tekemiseen?

Tuleeko teille teknisinä osaajina “ylimääräistä” työtä?

Onko teillä mielestäsi riittävästi väkeä datajournalismin tekemiseen?

Tukeeko muu toimitus deskiänne riittävästi?

Aiheuttaako työnne tekninen puoli vaikeuksia kommunikoida muun toimituksen kanssa?

Millaisissa tilanteissa?

Millaisia odotuksia muu toimitus työllenne asettaa?

Teettekö datajournalismin työtapoja tutuksi muulle toimitukselle/miten?

Millaisen toimittajan kanssa on helpointa tehdä datajournalismia?

Miten hyvin toimittajat ovat oppineet käyttämään teidän osaamsitanne?

5. Tulevaisuus

Millaisena näet datajournalismin tulevaisuuden teidän toimituksessanne?

Miten kehittäisit deskinne toimintaa, jos voisit?

Voisiko työprosessin osia mielestäsi ulkoistaa?

Millaisena näet datajournalismin tulevaisuuden toimituksenne ulkopuolella?

Miten datajournalismi tulisi huomioida journalistikoulutuksessa? Tarvitaanko koodaritoimittajia?

6. Muuta

Jäikö mielestäsi keskeisiä asioita kysymättä?

Onko kommentoitavaa haastattelusta?