

**This is an electronic reprint of the original article.  
This reprint *may differ* from the original in pagination and typographic detail.**

**Author(s):** Uskali, Turo

**Title:** Big Data ja mediasuhteet : miten suuret data-aineistot mullistavat taas

**Year:** 2015

**Version:**

**Please cite the original version:**

Uskali, T. (2015). Big Data ja mediasuhteet : miten suuret data-aineistot mullistavat taas. ProComma Academic, 2015, 76-92.

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# Big Data ja mediasuhteet: Miten suuret data-aineistot mullistavat taas kaiken?

Turo Uskali

*Voidaan perustellusti väittää, että digitaalisten data-aineistojen merkitys yhteiskunnassa kasvaa koko ajan. Jos sosiaalisen median nopea nousu uudeksi keskeiseksi viestinnän muodoksi tuli monille yllätyksenä viime vuosikymmenen aikana, sama muutos on toistumassa nyt laajojen data-aineistojen kanssa. Datan määrän lisääntyessä koko ajan tarvitaan eri aloilla uudenlaista viestinnällistä osaamista, sillä datajournalismi on tiimityötä. Tarvitaan työkäytäntöjen kehittämistä muun muassa data-aineistojen visualisoinnin suuntaan. Uudessa ympäristössä viestinnän ammattilaiset kohtaavat sekä datajournalisteja, jotka hyödyntävät suuria aineistoja, että data-aktivisteja, jotka vaativat suurempaa läpinäkyvyyttä. Viestinnän ammattilaisten tulisi kehittyä dataviestijöiksi, jotka hallitsevat niin organisaation sisäistä kuin ulkoistakin datavirtaa. Suurten data-aineistojen jemmaaminen vain organisaation omaan käyttöön vaikuttaa hieman vanhanaikaiselta strategialta avoimen innovaatioiden ja joukkoistamisen maailmassa. Vaikka uutisarvoista dataa on usein vain yksinoikeudella avattu materiaali, tulevaisuuden dataviestijän tehtäviin kuuluu keskeisenä läpinäkyvyystyö, jossa putsataan suurista raakadata-aineistosta ymmärrettäviä esityksiä.*

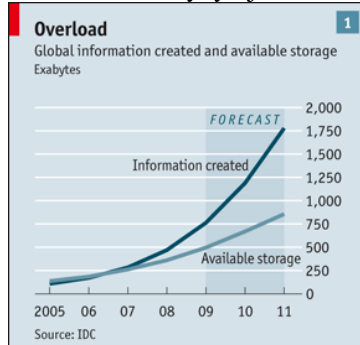
**Asiasanat:** Data, avoin data, big data, datajournalismi, dataviestintä, läpinäkyvyys

## Johdanto

Journalismin taloudellinen ja teknologinen murros on jatkunut jo pitkään. Viime vuosien digitaalinen viestintäkehitys on vienyt kohti reaaliaikaisuutta. Etenkin sosiaalisen median ja verkkoanalytiikan myötä on ollut mahdollista seurata varsin yksityiskohtaisesti eri journalististen juttujen ja sisältöjen saamaa suosiota. Reaaliaikaisuus on tehostanut myös palautteen saantia etenkin mediayhtiöissä. (Stone 2014.) Viime vuosien uudet haasteet ovat koskeneet esimerkiksi natiivimainontaa, droonikameroita ja ennen kaikkea suuria data-aineistoja. Kaiken takana on lisääntynyt tieto siitä, miten lukija tai median kuluttaja käyttäytyy ja mitä hän mediasisällöiltä hakee.

Kun internet of things (esineiden internet) sensoreineen mahdollistaa aikaisempaa suuremmat reaaliaikaisen datan virrat elinympäristöstämme ja droonikamerat voivat lentää vaarallisillekin kuvausalueille etsimään totuutta, on datan asema ja hallinta nousemassa viestinnässä keskiöön. Havaintopsykologinen fakta on, että ymmärrämme monimutkaisiakin asioita paremmin visualisointien kuin esimerkiksi pitkien tekstikatkelmien avulla, jolloin viestinnän ammattilaisten tehtäväksi nousee ottaa haltuun suuriakin organisaation datamääriä ja niiden avaaminen. Keskeisenä kysymyksenä on, miten parantaa dataviestintää digitaalisessa uutisekosysteemissä?

Data miningista eli datan louhinnasta on puhuttu esimerkiksi tietojenkäsittelytieteessä jo vuosikymmeniä (Fayyad et al. 1996), mutta digitaalisen datan määrän nopea kasvu ja sen laajamittainen arkistointi on noussut laajemmin julkisuuteen vasta viitisen vuotta sitten. (The Economist 2010; Hilbert & López 2011.) Kuvio 1 voidaan nähdä, miten valtavasti tiedon määrä on lisääntynyt, ja miten esimerkiksi sen varastointikapasiteetti seuraa perässä.



Kuvio 1. Tiedon lisääntynyt määrä suhteessa varastointikapasiteettiin (The Economist 25.2.2010)

Näyttää siltä, että big data on viime vuosina alkanut syrjäyttää data mining -termiä, vaikka itse asiassa kyse on paljolti samasta ilmiöstä, suurten data-aineistojen haltuunotosta ja jalostamisesta tiedoksi (Stone 2014). Muun muassa amerikkalainen talouslehti Forbes on ennustanut laajojen data-aineistojen hyödyntämisen synnyttävän uuden kultaryntäyksen (Peters 2012) tai tarjoavan öljyn kaltaista uutta vaurauden lähdeä (Rotella 2012). Voidaan jopa väittää, että uudet käsitteet, kuten ”big data” ja ”open data” ovat jo hypetyksen kohteena.

Datan, informaation sekä tiedon (knowledge) välinen käsitteellinen ero on usein kuvattu hierarkkisenä kolmiona. Alimmalla tasolla on jäsentymätön raaka data. Mitä lähempänä kolmion huippua ollaan, sitä laadukkaammasta ja jalostuneemmasta informaatiosta on kysymys, jopa tiedosta. **Viestinnän ammattilaisten tehtävänä onkin toimia kaikilla kolmion tasoilla, sekä avata raakadataa että jalostaa dataa tiedoksi ja tarjota sitä eri tarpeisiin.**

Journalismin tehtävänä on jo pitkään ollut eri lähteistä saatavan ajankohtaisen tiedon analysointi, tiivistäminen ja kärjekäs esittäminen. Journalismissa on perinteisesti arvostettu myös niin sanottuja asiantuntijalähteitä, kuten viranomaisia, poliitikkoja ja yritysjohtajia. Näiden henkilölähteiden lisäksi toimittajat ovat aina olleet kiinnostuneita myös dokumenttilähteistä, joihin datajournalismin voidaan katsoa lukeutuvan. Etenkin tutkivan journalismin työmetodeihin on kuulunut aika ajoin laajojen paperisten asiakirjapintojen läpikäyminen. Digitaalisen asiakirjatalennuksen aikakaudella suurtenkin data-aineistojen läpikäynti on tullut helpommaksi ja nopeammaksi.

#### Datajournalismin muodot

Datajournalismi voidaan jakaa kolmeen tyyppiin:

- 1) päivittäinen, nopea datajournalismi
- 2) hidas, tutkiva datajournalismi
- 3) algoritmeihin perustuva reaaliaikainen datajournalismi, kutsutaan myös ”robotjournalismiksi” ja ”sensorjournalismiksi”.

Nopea datajournalismi on mahdollista vain, jos toimituksiin saatava organisaatioiden avaama data-aineisto on valmiiksi hyvin käsiteltyä. Se ei saisi sisältää lyöntivirheitä, ja sen olisi oltava koneluettavassa muodossa. Dataviestinnän kannalta nopea datajournalismi on paras ja tehokkain yhteistyömuoto viestinnän ammattilaisten kanssa, sillä sitä kautta sekä organisaatio että media saavat nopeasti data-aineistoja julkisuuteen. Esimerkiksi kun Aller Media julkaisi keväällä 2015 Suomi24-verkkopalvelun keskustelut avoimena datana, jo pelkkä uutinen data-aineiston julkaisemisesta levisi laajasti uutismediassa (Kauppalehti 11.4.2015).

Hidas, tutkiva datajournalismi toimii pidemmällä aikajänteellä ja tarvitsee yleensä data-aineistoja, jotka eivät ole helposti saatavilla. **Viestinnän ammattilaisten kannalta tutkiva datajournalismi on haastavaa, sillä se käsittelee yleensä niin sanottuja arkoja aiheita, joita organisaatio ei välttämättä halua julkaista.** Tutkiva datajournalismi perustuikin usein datavuotoihin, jotka joku organisaatiossa työskentelevistä on halunnut tehdä ja saada siten julkisuutta havaitsemalleen epäkohdalle.

Reaaliaikainen koneellinen datajournalismi vaatii saumatonta teknologista yhteistyötä dataa tuottavan organisaation ja uutistoimituksen välillä. Käytännössä dataa tuottavan organisaation välittämä tieto julkaistaan uutismedian verkkopalvelussa muutamassa sekunnissa ilman yhdenkään ihmisen väliintuloa. Uutisointi on täysin automatisoitu. Reaaliaikaiselle datajournalismille on tarvetta etenkin kriisiuutisoinnissa, esimerkiksi maanjäristysten yhteydessä, jolloin muutaman minuutin viiveetkin saattavat merkitä ihmishenkien menetyksiä pelastustöissä. Robottijournalismia on käytetty jo ainakin Yhdysvalloissa myös yritysten osavuositarkastusten ja urheilutulosten uutisoinnissa.

Alkeellisimmillaan datajournalistinen tuotos voi olla vaikka digitaalisella karttapohjalla esitettyä tietoa paikkakunnan rikoksista, parkkisakoista, liikenneonnettomuuksista tai vaikkapa eduskuntavaalien puoluejakautumista kaupunginosittain. Vaativimmillaan datajournalistinen juttu voi kertoa vaikkapa eduskunnan valiokuntien lobbareista, Eurooppaan salaa yrittävien maahanmuuttajien kuolemista tai veroparatiisiyhtiöiden globaalista toiminnasta. Nämä kaikki edellä mainitut datajuttuaiheet palkittiin vuosien 2013–2014 aikana.

On huomattavaa, että data ja informaatio voivat olla myös visuaalisessa muodossa. Käytännössä tämä on tarkoittanut myös suurten kuva- ja videoaineistoarkistojen syntyä ja niiden käyttöä datajournalistisesti. Esimerkiksi Washington Postin jutussa (6.4.2015) tehtiin yhteenvedoita noin 95 000 taidemaalauksesta vuodesta 1800 lähtien. Yksi tuloksista oli, että sininen väri oli alkanut lisääntyä moderneissa taidemaalauksissa Picasson niin sanotun sinisen kauden (1901–1904) jälkeen. Tämä tieto voitiin myös visualisoida. **Voidaankin väittää, että juuri suurten data-aineistojen visualisointi on dataviestinnässä keskeisessä asemassa.** Yksi hyvä visualisointi auttaa viestinnän perille menossa paljon tehokkaammin kuin pitkä teksti. Tämä korostaa tietysti graafikoiden työn merkitystä viestinnässä. (Thierauf 1999; Davenport & Prusak 2000; Frost 2010.)

On kiinnostavaa pohtia, missä sitten on esimerkiksi big datan ja muun data-aineiston välinen raja. Columbian yliopiston datajournalismin konferenssissa lokakuussa 2014 asiantuntijat päätyivät kollektiivisesti määrittelemään rajan miljoonaan yksikköön – eli kun data-aineistossa on miljoona riviä informaatiota, se olisi big dataa. Loppujen lopuksi näin suuria data-aineistoja on ainakin toistaiseksi vielä vain harvoin saatavilla avoimena datana, eikä tällaisten aineistojen avaaminen välttämättä onnistu viestinnän ammattilaisilta vasta kuin esimerkiksi tilastokeskuksessa.

## Tietovuodoista datavuotoihin

Toimituksille on aina vuodettu tietoja organisaatioista käsin. Yleensä kysymyksessä ovat olleet niin sanotut arat poliittiset aiheet. Tietovuotojen merkitys on tutkivassa journalismissa suuri, ja datajournalismillakin voidaan nähdä yhteyksiä tietovuotoihin. Jo Watergate-paljastukset, jotka johtivat Yhdysvaltain presidentin Richard Nixonin eroon vuonna 1974, perustuivat tietovuotoon. (Ettema & Glasser, 1998.) Tietovuotajaksi osoittautui myöhemmin FBI:n varajohtaja William Mark Felt. Maailmanmaineeseen kohosivat kuitenkin the Washington Postin tutkivat toimittajat Bob Woodward sekä Carl Bernstein. **Kiinnostavaa kyllä, 2010-luvulla nimenomaan tietovuotajat ovat yleensä saaneet suuremman mediahuomion kuin tietovuodoista kirjoittaneet toimittajat.** Tämä on toisaalta uhka nykyorganisaatioiden suurien datamäärien suojelulle, mutta myös mahdollisuus yhteistyöhön. Bradley (nyk. Chelsea) Manning on tullut tunnetuksi monien keskeisten WikiLeaks-vuotojen tekijänä ja Edward Snowden puolestaan Yhdysvaltain NSA:n ja brittien vastaavan vakoilulaitoksen GCHQ tietojen sekä toimintatapojen paljastajana.

Suurten data-aineistojen avaamisella on myös juridiset perusteet. Maailman ensimmäinen tiedonvapaus- tai julkisuuslaki (Freedom of Information Act, FOIA) otettiin käyttöön Ruotsissa 1766. Tiedonvapauslain 250-vuotisjuhla vietetään siten ensi vuonna, 2016. Kiinnostavasti suomalaisilla oli vahva osuutensa lain synnyssä. Kokkolan kirkkoherra Anders Chydenius oli sen keskeinen ideoija, kirjoittaja ja eteenpäin viejä. (Mustonen et al. 2006.) Ruotsin esimerkin mukaan läpinäkyvyyttä tukeva julkisuuslaki on otettu käyttöön yhteensä 95 maassa. Se rinnastetaan usein jo perusihmisoikeuksiin. Esimerkiksi Yhdysvalloissa presidentti Lyndon Johnson hyväksyi lain vuonna 1966, ja Iso-Britanniassa se astui voimaan vasta kymmenen vuotta sitten, vuonna 2005. (Riekkinen & Suksi 2014; Schudson 2014; Worthy & Hazell 2014.)

Ei olekaan mikään ihme, että nimenomaan yhdysvaltalaiset toimittajat ovat olleet julkisuuslain innokkaimpia käyttäjiä jo vuosikymmenien ajan. Yhdysvalloista tietopyyntöjen tehtäily on vähitellen levinnyt eri puolille maailmaa, myös Suomeen, jossa Tietopyyntö-verkkosivusto avautui vuonna 2015. Sen kautta oli maaliskuuhun 2015 mennessä tehty jo lähes 100 tietopyyntöä. Pääsääntönä näyttää olleen, että eri maiden hallitukset ovat avanneet ensin lähinnä harmittomia ja hyvin usein journalistisesti käyttökeltottomia tietokantoja ja vasta journalistien tietopyyntöjen ja oikeuden päätösten jälkeen myös hieman arkaluonteisempia aineistoja, kuten esimerkiksi Isossa-Britanniassa parlamentin jäsenten kulukorvauksia. (Hewett 2013; Worthy & Hazell 2014.) Vanhat journalistiset periaatteet pätevät osin myös datajournalismin aikakaudellakin. **Aivan olennaista journalistisen julkisuuden kannalta on, että tieto tai data-aineisto on saatu yksinoikeudella eli samaa aineistoa ei ole muilla uutisorganisaatioilla.**

Tässä artikkelissa pääpaino on suurissa data-aineistoissa, jotka voivat olla strukturoimattomia tai strukturoituja. Strukturoiduilla data-aineistoilla tarkoitetaan hyvin jäsenneityjä ja järjestettyjä tietokokonaisuuksia, ns. "big dataa". (Manyika et al. 2011; Lewis and Westlund 2014, 2.) Avoimen datan voi puolestaan määritellä dataksi, "jota voi vapaasti käyttää ja jakaa kenelle tahansa" (Open Knowledge Foundation 2012). Avoimella datalla on läheinen historiallinen yhteytensä myös avoimen lähdekoodin sovellutusten kehitystyöhön sekä tieteelliseen vertaisarviointiin, jonka tarkoituksena on varmistaa työn laatu ja luotettavuus (Lesage & Hackett 2013, 42).

## Onko datan avaamisesta hyötyä?

Avoimen datan on uskottu lisäävän taloudellista tehokkuutta ja toimeliaisuutta. Esimerkiksi Yhdysvalloissa ja Euroopan unionissa avoimen datan on toivottu synnyttävän uusia yrityksiä ja jopa teollisuudenaloja. EU:n avoimen datan politiikka lanseerattiin vuonna 2011. Euroopan komissio arvioi avoimen datan politiikan luovan vuosittain noin 40 miljardin euron arvosta uusia palveluja. (European Commission 2015; The Verge 10.4.2013.)

Taloudellisen toimeliaisuuden lisäksi globaalina tavoitteena on ollut lisätä etenkin julkishallinnon läpinäkyvyyttä avoimen data avulla. Käytännössä tämä on ollut yhä varsin takkuista. Pontevista poliittisista julkilausumista huolimatta eri maiden välillä sekä myös maiden sisällä on yhä suuria eroja avoimen datan käytännöissä. Esimerkiksi Isossa-Britanniassa on jopa pelätty, että avoimen datan politiikka on jo saattanut vähentää muun muassa asiakirjoihin tehtäviä merkintöjä. (Bowles et al 2014; Schudson 2014; Riddell 2014; Worthy & Hazell 2014.) Käytännössä jalo tavoite hallinnon avoimuuden lisäämiseksi voi itse asiassa johtaa myös täysin päinvastaiseen vaikutukseen ylimääräisenä varovaisuutena ja tiedon pimityksenä. Kun kirjoittaja tietää, että hän voi joutua joskus julkistamaan esimerkiksi kokousasiakirjan tiedonvapauslain nojalla, pöytäkirjaan ei ehkä merkitäkään niin paljon tietoja kuin aiemmin oli tapana, ja esimerkiksi keskustelupöytäkirjat saattavat muuttua päätöspöytäkirjoiksi. **Näin tavoiteltu läpinäkyvyys tiedossa saattaakin käantyä epämääräisyydeksi.**

Vaikka monet organisaatiot ovat jo Suomessakin pitkään tallentaneet tietoja digitaalisessa muodossa, niillä ei ole ollut riittävästi ammattitaitoa avoimen datan julkaisemiseen. Käytännössä julkaiseminen vaatii data-aineistojen siirtämisen koneluettavaan muotoon, kuten esimerkiksi Excel-tiedostoihin. Suomessa avoimen datan julkaisemisen edelläkävijöitä ovat olleet uutistietojen mukaan esimerkiksi Maanmittauslaitos ja Ilmatieteenlaitos sekä pääkaupunkiseudun suuret kaupungit (Helsinki Region Infoshare). Maailmalla edelläkävijänä on pidetty esimerkiksi Yhdysvaltain keskushallinnon ylläpitämää avoimen datan portaalia, data.gov.

Länsimaisen journalismin perustehtävänä on vallankäytön valvonta ja myös yhteiskunnan kannalta tärkeiden epäkohtien julkistaminen. Juridisesti salaisten tietojen vuotajat saattavat toiminnallaan rikkoa lakeja ja joutua siitä vastuuseen, kuten esimerkiksi Afganistanin ja Irakin sotaan liittyneiden tietojen vuotaja Manning, joka sai 35 vuoden vankeustuomion. Toimitukset, jotka julkaisivat salaista tietoa, eivät sen sijaan tehneet laittomuuksia, eikä syytteitä niitä vastaan nostettu. Toimittajien lähdesuojan tarkoituksena on varmistaa, että yhteiskunnan kannalta tärkeitä tietoja vuodetaan myös jatkossa toimituksiin. Vuotojournalismiin saattaa liittyä myös vuotajien henkilökohtaisia motiiveja, kuten pettymyksiä tai kostonhimoa. Toisinaan vuodot saattavat olla myös hämäyksiä tai suoranaisia väärennöksiä. Lähdekritiikki ja eettiset ohjeet pätevät myös datavuotojournalismiin.

## Data ja tekijät

Tässä uudessa avoimen datan ympäristössä viestinnän ammattilaiset kohtaavat sekä datajournalisteja, jotka hyödyntävät suuria aineistoja, että data-aktivisteja, jotka käyttävät dataa omiin tavoitteisiinsa. Taulukko 1 summaa eri dataviestinnän kentän pelaajat. Datajournalistinen työ on sen verran teknisesti vaativaa, että käytännössä vain harvoin yksi työntekijä hallitsee kaikki sen tarvitsemat osa-alueet ja vaiheet. Yleensä median sisällä datatiimi koostuu toimittajasta, koodarista ja graafikosta, ja datajournalistisen työn aikaa

vievin osuus on ”datan puhdistaminen” eli lähinnä data-aineiston kirjoitusvirheiden löytäminen ja korjaaminen siten, että aineisto on yhdenmukaisessa kielellisessä asussa. Vasta datan puhdistamisen jälkeen voidaan tehdä luotettavaa analysointia.

Taulukko 1. Dataviestintäkentän pelaajat.

Datajournalistit	Suuria data-aineistoja journalistisessa työssään hyödyntävät toimittajat, graafikot tai koodarit. Kiinnostuksen kohteena ovat suuret data-aineistot, jotka ovat uutisarvoisia tai muuten kiinnostavia. Uutisarvoa lisää se, että data-aineistoa ei ole kenenkään muun uutisorganisaation hallussa. Suuria data-aineistoja voidaan myös niin sanotusti vuotaa toimituksiin.
Data-aktivistit	Avoimen datan lisäämistä yhteiskunnassa yleisesti tai tietyssä kontekstissa vaativat henkilöt tai järjestöt. Tarkoituksena on lisätä etenkin läpinäkyvyyttä poliittisessa ja taloudellisessa päätöksenteossa, mutta myös tarjota uusia liiketoimintamahdollisuuksia. Kiinnostuksen kohteena ovat kaikki suuria data-aineistoja keräävät organisaatiot, yhteisöt ja yritykset.
Dataviestijät	Suurten data-aineistojen käsittelyyn erikoistuneita organisaation viestinnän ammattilaisia. Kiinnostuksen kohteena on saada organisaation tuottamat data-aineistot leviämään ja hyödynnettäväksi mahdollisimman tehokkaasti sekä organisaation sisällä että sen ulkopuolella.
Datajoukkoi- taminen	Suurten data-aineistojen jakamista verkon kautta. Tavoitteena on saada suuri määrä ihmisiä käymään vapaaehtoisesti läpi data-aineistoja sekä löytämään niistä merkittäviä tietoja tai kehittämään uusia hyötysovellutuksia.

Dataviestintä on datajournalismia laajempi käsite, ja sitä voidaan käyttää myös muissa kuin toimitusorganisaatioissa. Keskeinen ero datajournalismiin verrattuna on, ettei dataviestinnän tarvitse olla uutisellista. Hyvä dataviestintä voi kuitenkin käyttää hyväkseen journalistisen verkkokerronnan tapoja, esimerkiksi visualisointeja ja tarinallistamista.

Datajournalismin ja dataviestinnän yhteydessä voidaan puhua myös huomiotyöstä (”attention work”), jolla tarkoitetaan viestinnän ammattilaisten kykyä välittää kiinnostavia sisältöjä eri yleisöille. Perinteisesti journalistien ja PR:n ammattilaisten on nähty toimivan ikään kuin samoilla markkinoilla, mutta eri lähtökohdista, jopa vihamielisinä toisilleen. On kuitenkin mahdollista löytää myös yhteisiä päämääriä ja tavoitteita näiden ammattiryhmien välillä, kuten esimerkiksi innovaatioviestinnässä ja -journalismissa on havaittu. Näyttää myös siltä, että etenkin talouden laskusuhdanteen aikoina ammattiryhmät ovat lähentyneet toisiaan (Luoma-aho et al. 2009), kaikkien toimiessa oman maineensa ja verkostojensa varassa.

Lesage ja Hackett (2013, 41–42) ovat käyttäneet termiä ”transparency work” eli läpinäkyvyystyö, joka kuvaa hyvin myös dataviestintää. Tarkoituksena on tehdä suuristakin raakadata-aineistosta viestinnän avulla ymmärrettäviä tiedollisia esityksiä. Yksi kenties helposti lähestyttävä esimerkki on infografiikka, josta löytyy jo tuhansittain esimerkkejä internetistä. Infografiikassa suuriakin data-aineistoja voidaan tiivistää numeroin, tekstein, erilaisin graafisin esityksin sekä kuvin. **Havaintopsykologinen fakta on, että ymmärrämme monimutkaisiakin asioita paremmin visualisointien kuin esimerkiksi pitkien tekstikatkelmien avulla.**

## Datajournalismin ensiaskeleita?

Datajournalismi ja suuriin data-aineistoihin perustuvat uudet journalistiset juttumuodot voidaan myös nähdä yhdeksi taloushaasteiden kanssa painivan journalismin ”pelastajista” (Rogers 2011; Uskali 2011; Gray et al. 2012.) Perustellusti voidaankin todeta, että viestinnän alalla journalistit näyttävät toistaiseksi reagoineen ammattiryhmistä nopeimmin suurten digitaalisten data-aineistojen avaamiin uusiin mahdollisuuksiin.

Modernin datajournalismin synty voitaneen ajoittaa vuosiin 2005–2006. Esimerkiksi Jonathan Hewett (2013, 4) on arvioinut, että ensimmäisiä merkkejä uudeltaisesta datajournalismista oli näkyvissä jo vuoden 2005 aikana, jolloin etenkin taloustoimittajat kiinnostuivat hyödyntämään internetin avulla digitaalisia asiakirjoja. Toisaalta the Washington Postin entisen toimittajan Adrian Holovaty'n Chicagocrime-verkkosivusto (2005), tietovuotosivusto WikiLeaksin perustaminen (2006), ruotsalaisen väestötieteilijän Hans Roslingin TED-luento Kalifornian Montereyssä (2006) sekä Yhdysvaltain varapresidentin Al Gore'n Oscar-palkittu ilmastodokumentti *An Inconvenient Truth* (2006) osoittivat kaikki, miten suuria datamääriä voidaan tehokkaasti ja ymmärrettävästi uutisoida, tarinallistaa ja visualisoida.

Datajournalismi, tietovuodot ja WikiLeaks

Varsinainen datajournalismin kansainvälinen läpimurto tapahtui kuitenkin vasta 2010, WikiLeaksin suurimpien, Afganistanin ja Irakin sotaan sekä Yhdysvaltain diplomaattipostiin liittyvien tietovuotojen uutisointien yhteydessä. Suurten tietoaineistojen hallinnointiin, käsittelyyn ja analysointiin tarvittiin systemaattista otetta ja ammattitaitoa – datajournalistista toimitustyötä. Esimerkiksi the Guardianin datajournalisti Simon Rogers on väittänyt, että datajournalismi syntyi viimeistään vuonna 2010. Tuolloin WikiLeaksin välittämiä, Afganistanin sotaan liittyviä 92 201 riviä dataa yritettiin viikkokausia turhaan mahduttaa eri puolilla maailmaa yleisessä käytössä olevaan Excel-taulukko-ohjelmaan. Senaikaiseen Excel-versioon pystyi syöttämään vain noin 60 000 riviä. Tämän jälkeen taulukko-ohjelma meni aina jumiin. (Rogers 2011.)

Ongelman ratkaisemiseen tarvittiin koodaustaitoa, ja esimerkiksi the Guardianissa perustettiin WikiLeaksille vuodettuja datasettejä varten omia datavarastoja. Vasta tämän jälkeen datan käsittely oli mahdollista ja juttujen teko alkoi luonnistua. Esimerkiksi sotatiedostot olivat hyvin strukturoituja ja siten nopeasti käyttökelpoisia journalistiseen tiedon etsintään. Ajan myötä kävi ilmi esimerkiksi, miten paljon siviiliuhreja Irakin sodassa oli ollut: yli 66 000. (Rogers 2011.) **Datajournalismin kehitys on siis ollut myös yhteiskunnallisen läpinäkyvyyden kehitystä tukeva asia.**

Suurten data-aineistojen on toivottu auttavan myös esimerkiksi vakavien globaalien ympäristöongelmien ratkaisemisessa. Toive perustuu ajatukseen, että mitä enemmän, tarkemmin ja nopeammin – toisin sanoen läpinäkyvämmiin – saamme dataa ympäristöstämme, sitä enemmän tiedämme ja voimme myös tehdä parempia päätöksiä. Esimerkiksi Kööpenhaminan ilmastokokous vuonna 2009 synnytti yhteistyöverkoston ilmastonmuutoksesta kiinnostuneiden tutkijoiden ja toimittajien kesken. Sen tarkoituksena on tuottaa ja uutisoida ”geodataa”. (Geojournalism.org 2015)

Suurimpien tietovuotojen joukkoon voidaan lukea myös kansainväliselle tutkivien toimittajien järjestölle (ICIJ) viime vuosina vuodetut journalistisesti äärimmäisen



kiinnostavat data-aineistot yritysten ja yksityishenkilöiden veroparatiisijoituksista. Nämä tunnetaan Offshore- ja SwissLeaks-nimillä. Myös nämä tietovuodot ovat olleet data-aineiston suhteen suuria, big dataa. Avoimeen dataan verrattuna tietovuodot tarjoavat datatoimittajille ja datatoimitukselle yleensä eräänlaisen yksinoikeuden data-aineistoihin ja samalla mahdollisuuden skuppeihin, eli jymyjuuttuihin. Esimerkiksi amerikkalaisessa talousuutismediassa avointa dataa ei pidetä läheskään yhtä arvokkaana uutislähteenä kuin esimerkiksi itse hankittuja, julkisuuslain mukaisiin tietopyyntöihin perustuvia data-aineistoja tai tietovuotoja. **Jos sama data-aineisto on samaan aikaan kaikkien toimitusten käytössä, sen uutisarvo saattaa vähetä, etenkin jos joku uutistoimitus on jo ehtinyt julkaista ensimmäiset jutut aineistosta.** Tietysti, jos avoimen data-aineiston tiedot ovat kiinnostavia ja tai yleisesti merkittäviä, myös muut toimitukset saattavat perehtyä aineistoon, mutta hyvin usein ne käytännössä tyytyvät vain referoimaan skuupin tehnyttä uutismediaa.

**Vuotojournalismin saaman julkisuuden yhtenä lähtökohtana voidaan pitää, että mitä salaisemmasta ja usein myös arkaluonteisemmasta tietoaineistosta on kysymys, sitä enemmän se kiinnostaa tutkivia datajournalisteja.** Veroparatiisiyhtiöt, eri maiden tiedustelulaitokset tai sotavoimat eivät perinteisesti ole pitäneet kovin suurta ääntä omasta toiminnastaan, joten jos näitä organisaatioita koskevia tietoja yhtäkkiä tulee toimituksiin, on selvää, että uutisointi on näyttävää ja pitkäaikaista.

Data-aktivistit ja joukkoistaminen

Datajournalistit ovat useissa maissa tehneet yhteistyötä avoimen datan aktivistien kanssa. Yhteisenä pyrkimyksenä on ollut yhdistää toimittajien ja aktivistien voimat ja saada ennen kaikkea uusia julkishallinnon data-aineistoja julkisuuteen. (Lesage & Hackett 2013, 43.) Open Knowledge Foundation ja Sunlight Foundation ovat kansainvälisesti tunnetuimmat avoimen datan aktivistiorganisaatiot. Myös Suomeen datajournalismi rantautui ainakin osaksi avoimen datan aktivistien kautta vuosien 2011–2012 aikana. Samaa tapaan kuin avoin data korostaa läpinäkyvyyden merkitystä yhteiskunnassa, korostuu läpinäkyvyys myös aktivistien vaatimuksissa ja datajournalismissa.

Läpinäkyvyysvaatimusten mukaan journalistisessa jutussa olisi kerrottava datan hankintatapa, datan käsittelyn metodit sekä analyysimenetelmät. Ehkäpä ensimmäistä kertaa journalismin historiassa myös toimittajilta edellytetään omasta työstään näin suurta työstöprosessin läpinäkyvyyttä ja avoimuutta. (Lesage & Hackett 2013.) Idealin datajutun yhteydessä julkaistaan myös koko käytetty data-aineisto koneluettavassa muodossa, jolloin muut toimittajat ja toimitukset voivat tarpeen tullen tarkistaa ja myös jatkojalostaa data-aineistoa omiin juttuihinsa. (Rogers 2011; Lesage & Hackett 2013.) Käytännössä kokonaisten datasettien avoin jakaminen datajuttujen yhteydessä on ollut yhä varsin harvinaista, vaikka ajoin pyrkimystä siihen suuntaan onkin.

Myös joukkoistamista eli yleisön saamista talkootöinä mukaan journalististen juttujen tekoon on kokeiltu datajournalismissa. Tähän mennessä tunnetuimmat tapaukset liittyvät jo aiemmin mainittuun brittiläisten parlamentin jäsenten kulukorvauksiin (MP's Expenses) sekä Lontoon mellakoihin (Reading the Riots). (Lesage & Hackett 2013, 40; Vehkoo 2013, 6–7.) Molemmissa tapauksissa tuhannet ihmiset avustivat the Guardianin toimittajia. Jälkimmäisessä juttuprojektissa oli mukana myös akateemisia tutkijoita, The London School of Economics and Politicsista.

Tähän asti tunnetuin kotimainen datajournalistinen joukkoistaminen perustui pankkien omaisuudenhoitajien ja meklarien tekemiin kauppa-aineistoihin. Helsingin Sanomat kävi vuonna 2011 ensin itse läpi aineistoa ja teki sen pohjalta uutisia. Lopulta lehti päätti kuitenkin joukkoistaa laajan data-aineiston lukijoilleen, jotka tekivätkin uusia uutislöytöjä, yhteensä noin 50. Lukijoiden avulla selvisi muun muassa, että monet meklarit ja omaisuudenhoitajat olivat tehneet laittomia ja epäeettisiä kauppvoja sekä saaneet ylimääräisiä bonuksia verottomasti. (Vehkoo 2013, 15–21.)

### Datapelit

Suurin osa datajournalistisista jutuista tarjoaa jonkinasteista interaktiivisuutta, vaikka kokonaisia data-aineistoja onkin yhä vain harvoin tarjolla. Verkkojutun lukija voi usein itse klikata esiin häntä itseään eniten kiinnostavia teemoja ja yksityiskohtia. **Useat uutistoimitukset ovatkin kertoneet, että datajournalistiset verkkojutut ovat erittäin luettuja ja lukijat viettävät niiden parissa keskimääräistä enemmän aikaa.**

Monissa luetuimmissa datajutuissa on myös pelillisiä piirteitä, kuten esimerkiksi The New York Timesin vuoden 2013 suosituimmassa jutussa, joka oli murretesti. Samankaltainen murretesti julkaistiin myöhemmin myös Helsingin Sanomien verkkosivuilla. **Hyvin usein datajournalistit saavatkin juttuideoitaan muiden tekemistä jutuista.** Näyttää vahvasti myös siltä, että 2014 oli nimenomaan verkkotestien läpimurtovuosi. Ne toivat kaivattua klikkivirtaa uutismedian, etenkin Yleisradion ja Helsingin Sanomien verkkosivuille. Myös eduskuntavaalit keväällä 2015 inspiroivat monia vaaliteemaisia pelisovellutuksia perinteisistä vaalikoneista puoluejohtajien ajatustenlukupeliin. (<https://vaalikoneet.zef.fi/eduskuntavaalit/>.)

Kiistatta yksi datajournalismin keskeisistä vahvuuksista on, että ainakin teoriassa sen avulla voidaan saada haltuun kokonaisia lähdeaineistoja, esimerkiksi koko maan kattavia tietoja eri alueilta. Yleensä journalismissa käytetään vain varsin suppeaa lähdepohjaa, esimerkiksi joitakin asiantuntijoiden haastatteluja. Voidaan kriittisesti kysyä, onko tietopohja tällöin liian sattumanvaraista, vain anekdootin tasoista. **Muun muassa Lesagen & Hackettin (2013, 43) mukaan datajournalismi saattaa olla ainakin jollain tasoilla täydempää ja myös objektiivisempää kuin esimerkiksi yksittäisen journalistin keräämä aineisto ja tulkinnat.**

Datajournalismiin, kuten myös kaikkeen muuhunkin dataviestintään, pätevät yleiset eettiset säännöt. **Dataviestinnässä on otettava erityisesti huomioon, että vaikka data-aineistossa olisi yksityiskohtaisia henkilö- ja osoitetietoja, niitä ei automaattisesti sellaisenaan saa julkaista.** (Bradshaw 2013.) Etenkin yksityisyyden suoja on kunnioitettava. Yhdysvalloissa käynnistyi ensimmäinen laaja eettinen keskustelu sen jälkeen, kun eräs itärannikon paikallislehti julkaisi paikkakunnan kaikkien aseluvan haltijoiden tiedot karttapohjalla. Datajutun julkaiseminen ajoittui Sandy Hookin kouluampumistragedian jälkeiseen aikaan vuonna 2012. (Culver 2013.) Selkeä yksityisyyden loukkaus olisi esimerkiksi, jos julkaistaisiin tietoja kaikista suomalaisista, joilla on aseenkantolupa tai vaikkapa yksityiskohtaisia sairaustietoja.

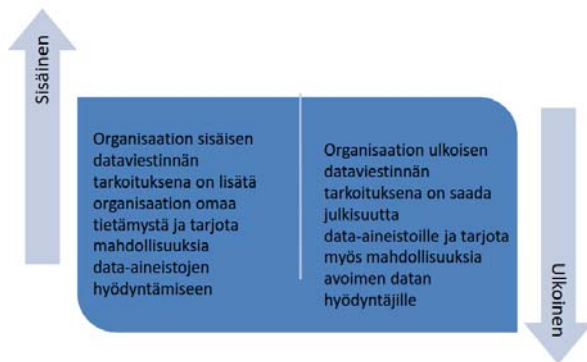
### Dataviestintää käytännössä

**Voidaan perustellusti väittää, että digitaalisten data-aineistojen merkitys kasvaa koko ajan yhteiskunnassa.** Digitaalisen datajournalismin lyhyt historia osoittaa, että uusia datakerronnan tapoja kehitellään koko ajan. Kuvaavaa on, että datajournalismin kansainvälisessä palkinnossa (GEN) harkitaan mobiilikategorian mukaanottoa. Dataviestintä

onnistuu vain, jos tekijöiden tilastotieteellinen tietämys data-aineistojen analysoimisessa on kohdallaan. Dataviestintä ja datajournalismi mahdollistavat myös tulevaisuuden ennakoinnin: Maailman parhaiten palkattu datajournalisti, amerikkalainen Nate Silver, on taustaltaan – tilastotieteilijä. Hän nousi maailmanmaineeseen ennustamalla the New York Timesin blogissaan Yhdysvaltain presidentinvaalien lopputulokset useaan kertaan lähes täysin oikein, ja paljon paremmin kuin uutismedian perinteisesti suosimat vaaligallupit tai vaaliasiantuntijat. Nykyään Nate Silver työskentelee amerikkalaisen urheilukanavan ESPN:n datajournalismin erikoistuneen blogin FiveThirtyEightin päätoimittajana. Nate Silverin esimerkin siivittämänä Yhdysvaltoihin on perustettu useita datajournalismin erikoistuneita blogeja ja muita verkkopalveluja, kuten New York Timesin the Upshot ja Vox.

Suomessa johtavia datajournalismin tekijöitä ovat Helsingin Sanomien datadeski ja Yleisradion plusdeski, joka jakaantui kahtia vuoden 2015 alun organisaatiouudistuksessa. Osa plusdeskin toimittajista, koodareista ja graafikoista jäi Yhteiskunta-toimitukseen ja osa siirtyi tukemaan ulkomaanuutisiin keskittyvän Maailma-toimituksen työtä. Lisäksi Yleisradion ruotsinkielisellä puolella on pieni datajournalismin yksikkönsä. Myös Suomen Kuvalehdessä on tehty aika ajoon näyttävää datajournalismia.

Miten paljon laajoja data-aineistoja eri yrityksissä ja muissa organisaatioissa on kyetty aidosti hyödyntämään, on jossain määrin jäänyt hämärän peittoon. Jotain kuitenkin kertoo se, että aika ajoin suuret yritykset, kuten Nokia, ovat avanneet patentejaan ja tarjonneet niitä kenen tahansa hyödynnettäviksi. Suurten data-aineistojen jemmaaminen vain omaan käyttöön vaikuttaa hieman vanhanaikaiselta strategialta avoimen innovaatioiden (Chesbrough 2003) ja joukkoistamisen maailmassa (Aitamurto 2014). Dataviestinnällä onkin organisaatioissa kaksi tehtävää, sekä sisäinen että ulkoinen. Kuvio 3 summaa nämä.



Kuvio 3. Dataviestinnän tehtävät organisaatiossa.

Dataviestinnässä pitäisikin korostua avoimuus ja ulospäin suuntautuneisuus, jopa läpinäkyvyys. Dataviestinnässä käytössä ovat luonnollisesti kaikki yleiset viestintäkanavat perinteisestä uutismediasta uusiin sosiaalisen ja mobiiliviestinnän sovellutuksiin. Viestintästrategiassa olisi hyvä miettiä, mitkä data-aineistot voisivat kiinnostaa eri sidosryhmiä ja mitä kautta nämä ryhmät olisivat parhaiten tavoitettavissa. Käytännössä dataviestinnässä ollaan kuitenkin vasta alkuvaiheessa, ja hyvän dataviestinnän käytännöt sekä pelisäännöt vaativat vielä kehitystyötä.

Organisaatioiden dataviestintä voi olla joko oma-aloitteista tai tietopyyntöihin reagoivaa. **Avoimuus- ja läpinäkyvyysperiaate toteutuvat luonnollisesti paremmin proaktiivisessa dataviestinnässä.** Molemmissa tapauksissa tietoaineistojen julkistaminen voi kuitenkin lopulta johtaa organisaation kannalta tärkeisiin ja kiinnostaviin lopputuloksiin. On oletettavaa, ettei kaikilla organisaatioilla välttämättä ole itsellään riittävästi omia resursseja kaikkien data-aineistojensa analyysiin, saati sitten tehokkaaseen dataviestintään. Hieman lisää julkaisuutta voi saada, jos tarjoaa (tai vuotaa) kiinnostavaa data-aineistoa ensin vain yhdelle joukkoviestimelle. Perinteisesti esimerkiksi toimitusten johtavat tutkivat journalistit ovat saattaneet saada aika ajoin erikoispostilähetyksiä ja vinkkejä. Heihin ja heidän työnsä jälkeen luotetaan.

Data-aineiston julkaiseminen avoimena datana ja sen analyysin joukkoistaminen voi olla kaikkien kannalta paras ratkaisu. Etenkin julkishallinnossa hallinnon läpinäkyvyyden periaate toteutuu varsin tehokkaasti nimenomaan suuria data-aineistoja julkaisemalla.

Datajournalistien ja muiden avoimesta datasta kiinnostuneiden henkilöiden tekemät tekstiyhteenvedot ja visualisoinnit saattavat paljastaa mielenkiintoisia asioita, kuten esimerkiksi taloudellista väärinkäyttöä. Hyvin usein organisaatioiden edustajat ovatkin perustelleet avoimen datan julkaisemista sillä, että myös he itse tarvitsevat jalostuneempaa tietoa ja palautetta omasta toiminnastaan.

Paul Bradshaw on aiheellisesti muistuttanut The Data Journalism Handbookin (2012) johdannossa: ”Kuten mikä tahansa lähde, sitä (data-aineistoa) pitäisi käsitellä skeptisesti; ja kuten mitä tahansa työkalua käytettäessä, meidän tulisi olla tietoisia sen rajoituksista.” Laajataan data-aineistot eivät välttämättä ole täydellisiä tai kerro kaikkea. Niiden kokoamiseen ja tallentamiseen voi liittyä monia inhimillisiä virhemahdollisuuksia. Niitä voidaan myös tarkoituksellisesti manipuloida ja väärentää. Peruskriittisyys on hyvä muistaa aina, kun käyttää lähteenään suuria data-aineistoja. On kuitenkin varmaa, että esimerkiksi datajournalismin nykyisistä parhaista työkäytännöistä löytyy runsaasti käyttökelpoisia malleja ja ideoita myös organisaatioviestinnän ammattilaisille.

Alla on listattuna viisi datajournalismin työkäytännöistä juontuvaa perusneuvoa, joista voisi lähteä liikkeelle oman dataviestintäkoulutuksen ja datastrategian kehittämiseksi.

- 1) Etsi ja tunne organisaatiosi data-aineistot. Yleensä tietokannoista ja -järjestelmistä on olemassa niin sanotut rekisteriselosteet, joiden pitäisi kertoa data-aineistojen sisällöistä. Julkisuuslain mukaan nämä rekisteritiedot olisi kerrottava aina myös tietopyynnön tekijälle.
- 2) Opettele taulukkolaskentaohjelmien käyttöä. Datan käsittelyn perustaitoihin kuuluu taulukkolaskentaohjelmien hallinta (esimerkiksi Excel tai Googlen vastaavat ohjelmat). Monet edistyneemmätkin datajournalistit ovat kertoneet, että heidän työnsä perustuu edelleen pitkälti taulukkolaskelmien käytölle.
- 3) Muokkaa data-aineisto koneluettavaan muotoon. Dataviestinnän näkökulmasta on erittäin tärkeää, että suuret data-aineistot pystytään muokkaamaan koneluettavaan muotoon eli esimerkiksi Excel-tiedostoiksi. Käytännössä etenkin julkishallinnossa on paljon data-aineistoa tallennettuna pdf-muodossa, joka ei sellaisenaan kelpaa dataviestinnän sisällöksi. Pdf-muotoinen data-aineisto on ensin muunnettava johonkin taulukkolaskentaohjelmaan sopivaksi. Myös verkkosivuilla julkaistua data-aineistoa kannattaa muuntaa koneluettavaan muotoon. Tätä kutsutaan datan skreippaamiseksi (scraping). Skreippaamista varten on olemassa useita verkkotyökaluja.

- 4) Puhdista data-aineistoa. Valtaosa raakadata-aineistoista on aluksi niin sanottua likaista dataa. Datan puhdistaminen eli käytännössä datan tallentamisessa tapahtuneiden lyöntivirheiden korjaaminen ja data-aineiston kriittinen sisällön laadun arviointi on olennaisen tärkeää laadukkaassa dataviestinnässä. Likaisen datan puhdistaminen saattaa viedä jopa yli 80 prosenttia dataviestintään käytetystä ajasta. Valmiiksi puhdistettujen ja kiinnostavien datasettien tarjoaminen toimituksiin voisi olla uusi, tehokas viestinnän muoto – etenkin, koska pienet toimitukset eivät useinkaan pysty itse tekemään datan puhdistusta ja samalla datajournalismia resurssien vähäisyyden vuoksi. Suomessa etenkin Helsinki Region Infoshare ja Open Knowledge Foundation voivat antaa hyviä neuvoja dataviestinnän aloittamisessa.
- 5) Visualisoi data-aineistoja. Datan visualisointi ja tarinallistaminen lisäävät data-aineistojen kiinnostavuutta. Data-aineistojen visualisointeihin on olemassa monia ilmaisia ja helppokäyttöisiä tietokoneohjelmia.

### Näistä voit aloittaa

Kuten yllä olevasta käy ilmi, dataviestinnässä ei välttämättä tarvita aluksi lainkaan koodaus- eli ohjelmointitaitoa. On kuitenkin eduksi, jos dataviestijä tuntee koodareita ja saa tarpeen tullen myös koodausapua. Kuten aiemmin jo on mainittu, dataviestintä vaatii useimmiten tiimityötä ja eri erikoisalueiden hyvää hallintaa. Lisäksi kannattaa liittyä sosiaalisen median data-aiheisiin ryhmiin. Esimerkiksi monet Facebookin aktiiviset datajournalismiin ja avoimeen dataan erikoistuneet kansainväliset ja kansalliset ryhmät toimivat tehokkaina, jopa reaaliaikaisina, uuden tiedon jakajina sekä helpdeskeinä. Sosiaalisen median ryhmät toimivat myös verkostoitumisen sekä oman työn markkinoinnin apuna.

- 1) Gray, J., Bounegru, L. & Chambers, L. (toim.) (2012). *The Data Journalism Handbook*. EJC: Brussels. <http://datajournalismhandbook.org/1.0/en/>.
- 2) Bradshaw, P. (2013a). *Scraping for Journalists*. Lean Publishing. <https://leanpub.com/scrapingforjournalists>.
- 3) Bradshaw, P. (2013b). *Data Journalism Heist*. Lean Publishing. <https://leanpub.com/DataJournalismHeist>.

### Lähteet

- Aitamurto, T. (2014). *Collective Intelligence in Open Journalism: Power, Knowledge and Value*. Tampere: University of Tampere.
- Bowles, N., Hamilton, J. T. & Levy, D.A.L. (2014). Introduction. Bowles, N., Hamilton, J. T. & Levy, D.A.L.: *Transparency in Politics and the Media*, xi–xxiii. Reuters Institute for the Study of Journalism; London: I.B.Tauris.
- Bradshaw, P. (2012). Introduction: What is Data Journalism? Gray, J., Bounegru, L. & Chambers, L. (toim.). *The Data Journalism Handbook*. Brussels: EJC. <http://datajournalismhandbook.org/1.0/en/>.
- Bradshaw, P. (2013). Ethics in Data Journalism: accuracy. Online Journalism Blog. <http://onlinejournalismblog.com/2013/09/13/ethics-in-data-journalism-accuracy/>, 16.3.2015.

Kentän koodi muuttunut

Kentän koodi muuttunut

Kentän koodi muuttunut

- Campbell, D. (2013). *How ICIJ's Project Team Analyzed the Offshore Files*. ICIJ.org, <http://www.icij.org/offshore/how-icijs-project-team-analyzed-offshore-files>, 19.3.2015. Kentän koodi muuttunut
- Chesbrough, H. W. (2003). *Open Innovation: The new imperative for creating and profiting from technology*. Boston: Harvard Business School Press.
- Culver, K. (2013). *Where the Journal News Went Wrong in Mapping Gun Owners*. Mediashift. Pbs.org, <http://www.pbs.org/mediashift/2013/02/where-the-journal-news-went-wrong-in-mapping-gun-owners053/>, 17.3.2015. Kentän koodi muuttunut
- Ettema, J. S. & Glasser, T.L. (1998). *Custodians of conscience: investigative journalism and public virtue*. Columbia University Press, New York.
- European Commission (2015). *Digital Agenda: Turning government data into gold*, [http://europa.eu/rapid/press-release\\_IP-11-1524\\_en.htm?locale=en](http://europa.eu/rapid/press-release_IP-11-1524_en.htm?locale=en), 14.3.2015. Kentän koodi muuttunut
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17(3), 37–54.
- Friedersdorf, C. (2014). The Latest Snowden Leak Is Devastating to NSA Defenders. *The Atlantic*. <http://www.theatlantic.com/politics/archive/2014/07/a-devastating-leak-for-edward-snowdens-critics/373991/>, 17.3.2015. Kentän koodi muuttunut
- Geojournalism.org. 2015. <http://geojournalism.org/>, haettu 14.3.2015. Kentän koodi muuttunut
- Gray, J. Bounegru, L. & Chambers, L. (toim.) (2012). *The Data Journalism Handbook*. EJC: Brussels. <http://datajournalismhandbook.org/1.0/en/> Kentän koodi muuttunut
- Hilbert, M & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* 332(6025), 60–65. Muotoiltu: suomi
- ICIJ.org. 2015. <http://www.icij.org/>
- Kuutti, Heikki. (2011). *Julkisuuslaki*. Jyväskylä: Jyväskylän yliopisto.
- Lewis, S. C. & Westlund, O. (2014). Big Data and Journalism. *Digital Journalism*, DOI: 10.1080/21670811.2014.976418.
- Lesage, F. & Hackett, R. A. (2013). Between Objectivity and Openness – The Mediality of Data for Journalism. *Media and Communication* 1(1), 29–50.
- Luoma-aho, V., Uskali, T. & Weinstein, A. (2009). 14. Pitfalls of Attention Work in the Innovation Ecosystem. Stanford University 2009. *Innovation Journalism* 6(5). <http://www.innovationjournalism.org/archive/injo-6-4.pdf>. Muotoiltu: suomi
- Manyika, J., Chui, M., Brown, B., Bughin, J., Roxburgh, C. & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Quarterly, [http://www.mckinsey.com/insights/mgi/research/technology\\_and\\_innovation/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation), 16.3.2015. Kentän koodi muuttunut
- Morozov, E. (2011). *The Net Delusion. The Dark Side of Internet Freedom*. New York: Public Affairs. Kentän koodi muuttunut
- Mustonen, J. (toim.) (2006). *The World's First Freedom of Information Act*. Anders Chydenius Foundation: Kokkola. Kentän koodi muuttunut
- Oxford English Dictionary. 2015. <http://www.oed.com/>. Kentän koodi muuttunut
- Peters, B. (2012). *The Big Data Gold Rush*. Forbes.com, <http://www.forbes.com/sites/bradpeters/2012/06/21/the-big-data-gold-rush/>, 16.3.2015. Kentän koodi muuttunut
- Rogers, S. (2011). *Facts are sacred: The power of data*. *Guardian shorts*. London: The Guardian. Kentän koodi muuttunut

Rotella, Perry. (2012). *In Data the New Oil*. Forbes.com,

<http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/>,  
16.3.2015.

Riddell, P. (2014). Impact of Transparency on Accountability. Bowles, N., Hamilton, J. T. & Levy, D.A.L. (toim.): *Transparency in Politics and the Media*, 19–30. Reuters Institute for the Study of Journalism; London: I.B.Tauris.

Schudson, M. (2014). Origins of the Freedom of Information Act in the United States. Bowles, N., Hamilton, J. T. & Levy, D.A.L. (toim.): *Transparency in Politics and the Media* 1–18. Reuters Institute for the Study of Journalism; London: I.B.Tauris.

Stone, M. L. (2014). *Big Data for Media*. Oxford: Reuters Institute for the Study of Journalism.

The Verge. 10.4.2013. EU Open Data law set to speed the development of Europe-wide services; Could generate an extra \$52 billion for the EU each year.  
[http://www.theverge.com/2013/4/10/4208196/eu-open-data-law-passed-by-council,14.3.2015.](http://www.theverge.com/2013/4/10/4208196/eu-open-data-law-passed-by-council,14.3.2015)

~~The Verge. 10.4.2013. EU Open Data law set to speed the development of Europe-wide services:~~

~~Could generate an extra \$52 billion for the EU each year, 14.3.2015.~~

Uskali, T. (2011). *Innovaatiot ja journalismi*. Helsinki: Infor.

Worthy, B. & Hazell, R. (2014). The Impact of the Freedom of Information Act in the UK. Bowles, N., Hamilton, J. T. & Levy, D.A.L. (toim.): *Transparency in Politics and the Media*, 31–48. Reuters Institute for the Study of Journalism; London: I.B.Tauris.

Kentän koodi muuttunut

Muotoiltu: Fontti: 12 pt

Kommentti [EJ1]: Tästä hävinnyt jotain väliä? En uskaltanut muokata ollenkaan.

Muotoiltu: Alaviitteen teksti, Ei mitään, Sisennys: Vasen: 0 cm, Riippuva: 1 cm, Kuvio Läpinäkyvä

## Kirjoittajan lyhyt esittely

Turo Uskali (synt. 1968) on journalistiikan dosentti Jyväskylän yliopiston viestintätieteiden laitoksella. Ennen yliopistouraansa hän oli kymmenisen vuotta toimittajan tehtävissä alueellisista sanomalehdistä valtakunnallisiin uutismedioihin. Uskali on viime vuosina tutkinut ja opettanut etenkin journalistisia innovaatioita. Hän on johtanut useita monitieteisiä Tekes-hankkeita sekä Helsingin Sanomien säätiön rahoittamia tutkimusprojekteja vuodesta 2008 alkaen. Uskali on toiminut myös vieraillevana tutkijana Stanfordin ja Oxfordin yliopistoissa vuosina 2006–2008.