



D.3.2.1.2.B

## Trends In Data Journalism

Editor: Esa Sirkkunen  
Author(s): Tanja Aitamurto, Esa Sirkkunen, Pauliina Lehtonen  
Confidentiality: Public  
Date and status: 08.09.2011 1.0

This work was supported by TEKES as part of the next Media programme of TIVIT (Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT)

## TRENDS IN DATA JOURNALISM

{Participants = all research organisations and companies involved in the making of the deliverable}

Participants	Name	Organisation
Researcher	Tanja Aitamurto	<b>UTA/Stanford</b>
Researcher	Pauliina Lehtonen	<b>UTA/JRDC</b>
Researcher	Esa Sirkkunen	<b>UTA/JRDC</b>

next Media  
www.nextmedia.fi  
www.tivit.fi

## Executive Summary

The majority of European countries have opened their data resources quite slowly—even in spite of the directive to do so, given by the European Commission in 2003. When regarded from a global perspective, it becomes apparent that the three largest categories of government that are actively opening their data to the public, currently, are: countries fighting corruption, post-authoritarian regimes that erect legislative barriers to promote open data, and countries who publish data due to tradition or conviction.

Data journalism is finding its way into newsrooms. Reporters at US daily newspapers routinely turn to local, state and federal government websites to hunt for data that they can use in their stories, a recent survey by the Reynolds Journalism Institute found. And a survey (n=108) done by the non-profit European Journalism Centre, shows that most respondents are already engaged in data projects. The main barrier preventing journalists from starting data projects, then, appears to be a lack of knowledge about how to work with data. The most urgent training is needed in the area of developing stories from data, but it's also called for in the areas of visualising and analysing data. More than that, journalists registered an interest in wanting to learn better ways to find, verify and clean data.

Data journalism has become an integral part of several news organisations' online strategies. They see data journalism as a part of a transition that's moving them from being a news-and-information site toward a more interactive news-and-information platform. In their new role, news organisations want to provide their readers with ways to interact with the content, and refine it further, and data journalism provides an opportunity for this.

According to the interviews conducted for this study, data journalism is perceived to be a tool for producing better journalism with the help of statistical methods, visualisation and interactive means of presenting information. Journalists see data journalism as a way to find hidden stories—stories that wouldn't otherwise get told. News organisations hope to serve public interest better with data journalism and, thus, have a positive impact on society. Furthermore, data journalism is perceived as a way to increase transparency in journalistic processes. This transparency potentially increases trust in journalism and the media, given that the sources of information in stories are open for anybody to access.

Data journalism adds new requirements to the journalist's profession. Now these individuals need skills for data analysis, web development and community management. However, the most important factor for successful data journalism is a journalistic sense of what is relevant and interesting, and of what questions need to be asked.

News organisations are looking for sustainable ways to support data journalism, but viable business models have yet to be discovered. News organisations have visions of becoming number-one data stores and of adding the provision of data analysis to their oeuvre of paid service offerings. The most likely future scenario will see news organisations existing in a state of symbiosis with other data service providers, booming in the new ecosystem evolving around this area.

## Table of Contents

Executive Summary.....	<u>2</u>
1 Introduction.....	<u>4</u>
2 Overview: Public data and data journalism in some countries (Esa Sirkkunen) ....	<u>5</u>
2.1 The state of open, publicly produced data .....	<u>5</u>
2.2. The state of data journalism.....	<u>6</u>
2.3. Some new trends: trusted data hubs and automatic news.....	<u>7</u>
3 Data journalism strategies in the US and elsewhere (Tanja Aitamurto).....	<u>9</u>
3.1 Role of data journalism in news organisations .....	<u>9</u>
3.2 Sources of data.....	<u>10</u>
3.3. Workflow in data journalism .....	<u>10</u>
3.4. Best practices in the workflow.....	<u>11</u>
3.5. Data journalism teams, skills and tools .....	<u>12</u>
3.6. Data hackathons, hack days and codeathons.....	<u>13</u>
3.7. Challenges in data journalism.....	<u>13</u>
3.8. Business models and the new data ecosystem .....	<u>14</u>
3.9. Vision: The future of data journalism.....	<u>16</u>
4 Conclusions.....	<u>17</u>

## Appendixes

Appendix 1. A selection of data journalism cases, especially of local journalism .....	<u>18</u>
(Pauliina Lehtonen)	
Appendix 2. Subjects interviewed by Tanja Aitamurto .....	<u>26</u>
References .....	<u>26</u>

## 1 Introduction

The aim of this report is to produce a short overview of what's happening in the field of open data and data journalism in the US and Europe. This is a continuation on the Next Media report, *Open Data in Finland* (Lehtonen, 2011), which dealt with the perspectives of some of the key actors on the future of open data in Finland.

The structure of this report will be as follows: first, a brief summary of the current state of public data and data-driven journalism in different countries; next, a more detailed assessment of data-journalism innovations, workflows and business models. The report will finish up with conclusions and a collection of links—close to 60—that show examples of how different types of data can be visualised and connected to stories and journalistic storytelling. Some of the links are chosen with the Next Media WP 3's emphasis on hyperlocal content creation in mind.

In compiling this report, the authors have used relevant research literature and conducted interviews with journalists from the US, UK and Argentina who hail from such journalistic institutions as *The Guardian*, iWatch News, Pro Publica, *Oakland Local*, *Bay Citizen*, *La Nacion*, *The Chicago Tribune* and *The New Scientist*.

The news organisations chosen for the study represent a wide variety of outlets in terms of size, reach and innovation. The interviewees range from traditional large newsrooms—such as *The Chicago Tribune*; to the world's leading innovators in online journalism—such as *The Guardian*; to new, innovative local and national newsrooms—such as *The Oakland Local*, *The Bay Citizen* and ProPublica. All of this was supplemented with input from some of the older innovators in investigative reporting, such as the Center for Public Integrity, one of the oldest and largest nonpartisan and nonprofit investigative news organisations in the US. The same open-ended, semi-structured questions were asked of all of the interviewees. Furthermore, several interviews with open data experts have been conducted and used as background information for the report.

## 2 Overview: Public data and data journalism in some countries (Esa Sirkkunen)

### 2.1. The state of open, publicly produced data

The term “open data” shares some roots and an ethos with such concepts as “open source,” “open content” and “open access.” All of these ideals want to free data—and information in general—from all kinds of restrictions, be they copyrights, patents or other kinds of control.

There are several definitions of open data, all of which refer more or less to the same features. For example, Schellong & Stepanets (2011) present eight principles according to which it is possible to define whether data is open or not (see also Lehtonen, 2011). Open data principles according to Schellong and Stepanets abide by the following eight principles.

First, *completeness*. All public data are made available. Public data are not subject to valid privacy, security or privilege limitations. Second, *primacy*. Data are collected at the source, with the finest possible level of granularity, not in aggregated or modified forms. Third, *timeliness*. Data are made available as quickly as possible to preserve their value. Fourth, *accessibility*. Data are available to the widest range of users for the widest range of purposes. Fifth, *machine readability*. Data are reasonably structured to allow for automated processing. Sixth, *absence of discrimination*. Data are available to anyone, with no requirement of registration. Seventh, *absence of propriety*. Data are available in a format over which no entity has exclusive control. Finally, *an absence of license requirements*. Data are not subject to any copyright, patent, trademark or trade-secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

Open government has become an important term underlining the significance of public access to governmental data and information, and the transparency of governmental actions. Among the origins of the recent discussions on this subject is that the US and Great Britain have started the opening of governmental data with projects like [www.data.gov](http://www.data.gov) and [www.data.gov.uk](http://www.data.gov.uk). The US and Great Britain are quite far when it comes to implementing open data ideas into practices. Also Canada and Australia seem to be well on their way ([Guardian, 2010](#)).

A web journal, [Owni.eu](http://Owni.eu), recently published an interesting summary by Pierre Alonso (Alonso, 2011) of the legal frameworks around open data in different countries based on [OECD information](#). Through a comparison of the different types of open data, Alonso was able to grade the freedom of data in different countries into five clusters.

In the first cluster, the majority of data must be published under penalty of law. Examples: *Russia, Estonia, Finland, Iceland, Ukraine, Turkey, Hungary, Slovakia, Italy, Spain, Portugal, Israel, Canada, Mexico, Chile, Korea*. In the second cluster, the majority of data are published by custom, without a set legal framework. Examples: *Brazil, Sweden, Denmark, the Netherlands, UK, Switzerland*. In the third cluster, as much data are published by law as is published customarily. Example: *Australia*. In the fourth cluster, as much data are published by custom as are not published at all. Examples: *United States, Egypt, Norway*,

*Ireland, Austria, Czech Republic.* In the fifth cluster, the majority of data are not published. Examples: *France, Belgium, Poland, New Zealand.*

According to Alonso, the strict legislation is typical of “young countries that are in transition and who value transparency in government as a way to fight corruption.” The three largest governmental categories that are actively opening their data to the public are countries fighting corruption, post-authoritarian regimes who erect legislative barriers to promote open data, and countries who publish data due to tradition or conviction.

According to Alonso, there are substantial gaps between different countries’ published data. “Only 5.5 percent do not publish reports related to their budget, institutions and respective procedural documents, while 64 percent of governments do not publish a list of civil servants and their salaries.” Those who do offer this kind of data are fighting corruption, including countries like Mexico and Italy, Alonso states.

It is important to understand that those countries lacking strict legislation governing their public data are not necessarily engaged in secretive activity. On the contrary, in countries like Great Britain and Sweden, the tradition of common law and customary unwritten law stress the importance of easily accessed public data.

The majority of European countries have opened their data resources slowly. A report (Schellong & Stepanets, 2011) dealing with the open data policy ecosystem in Germany, France, Denmark, Sweden, Norway, Finland, Italy and Spain, shows that none of these states had a national open data portal by the end of 2010. The analysis also demonstrates that the national statistics offices’ online data already fulfill a number of open data principles, but that there is a lot of room for improvement.

To conclude, one could say in these eight European countries, the process of opening official data for public use is still very much ongoing. One can say that the situation differs from country to country - not only is legislation different, but also the way a country’s political culture shapes its open data policies. And there are other important sources of open data than the governmental one, including various NGOs, the EU and the OECD, which produce useful open data for journalists’ use. In addition, data-mining and -analysing companies have started to offer their services to media houses (more about data companies in chapter 3.2).

## 2.2. The state of data journalism

Ten years ago, using large data sets in reporting was difficult and required skills that went beyond what was common among journalists. Various studies on journalistic practices verify that journalists most often rely on data provided by other sources—governments, city officials, research companies, etc. Only in some cases of investigative journalism have journalists been able to find resources to gather and analyse their own data and publish their results. But the situation is changing rapidly. More and more free data are available online and there are efficient tools with which anyone can analyse, visualise and publish large amounts of data.

The awareness of new possibilities has grown among journalists, especially after the publication of the Afghanistan war logs by *The New York Times*, *Der Spiegel* and *The Guardian*. In particular, the way *The Guardian* was able to reveal the expenses of British MPs has been celebrated among journalists. Data journalists of the European Journalism Centre find that developing the know-how to use available data more effectively could represent a huge opportunity for breathing new life into journalism. “Reporters can find new roles as ‘sense-makers’ by digging deep into data, in turn making journalism more socially relevant,” the EJC report on data-driven journalism (EJC 2010) states.

Data hunting is already an important part of the daily routines of many journalists. Reporters at US daily newspapers routinely turn to local, state and federal government websites to hunt for data that they can use in their stories, a [recent survey](#) by the Reynolds Journalism Institute found (Hertzog, 2011). The reporters ranked state data as the most accurate, and local as the least. They said that their local governments were the most transparent and accountable, while the federal government ranked the lowest on those counts. According to the results of the survey, campaign finance and lobbying data are the most available online; workplace safety inspection records are the least.

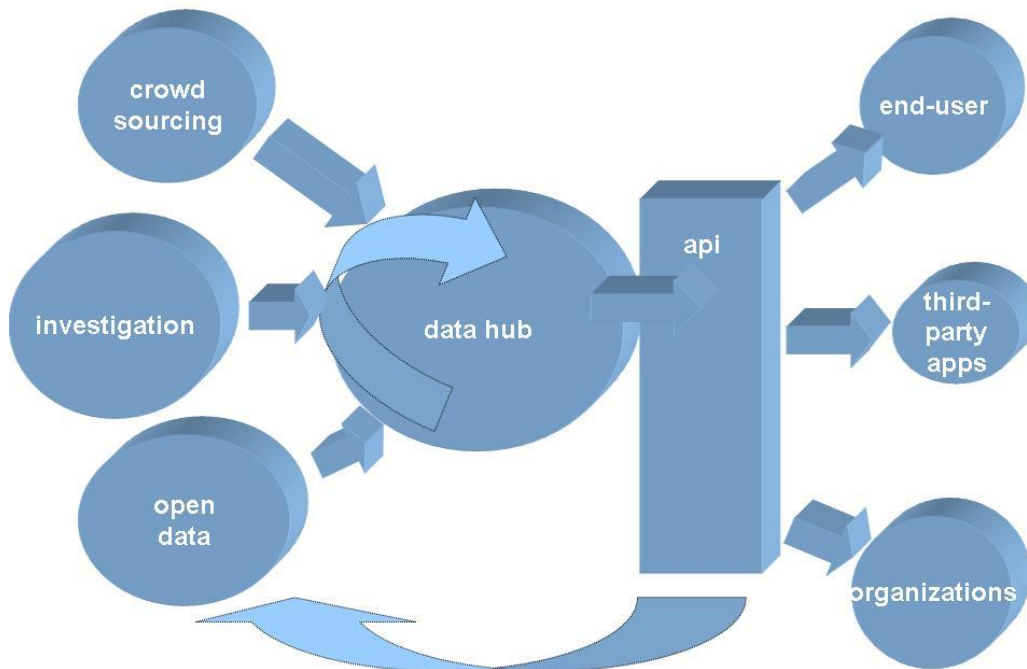
Preliminary results of an ongoing survey (n=108) done by the EJC among journalists interested in data journalism showed that most of the respondents were already involved in data projects (Bounegru, 2011). According to the results, the main barrier preventing journalists to start data projects is a lack of knowledge about how to work with data. The responses indicate that the most urgent need for training is in the development of journalistic stories from data material. Also, visualising and analysing data were mentioned as crucial to the learning process. And journalists expressed an interest in developing better ways to find, verify and clean data.

### 2.3. Some new trends: trusted data hubs and automatic news

Three data-oriented journalists—Mirko Lorenz from Germany, Nicholas Kayser-Bril from France and Geoff McGhee from the US—published a polemic article about the future of journalism ([Lorenz, M.&Kayser-Bril N.&McGhee G. 2011](#)) in the spring of 2011. Therein, the journalists state that media companies should leave the old business model, based on selling the attention of their audience to advertisers, alone. They should adapt to new technologies and new opportunities and become data hubs, they say, building infrastructures that turn them into centres of reliable data, able to analyse even very large and complex data sets internally and build stories on their insights.

How should these data hubs be built? First of all, the data should be open for use. The data could then be reused through APIs (application programming interfaces) that let developers use the bits of information to provide “information-as-a-service” interfaces (see the following chart). There might even be new earning possibilities. Beyond the ad-sponsored model, media organisations could implement several tiers, on a freemium model. A visualisation for the end user could be free. But tailored information crafted in specific ways

to serve businesses, public safety officials, social service providers, realtors and home-buyers would provide immense value,” the authors believe.



**Chart 1. In the data hub model, the media house collects data from different sources, and makes them accessible to outside end-users, developers and organisations interested in data. In this model, some earning can be enjoyed by tailoring the information for different users and their needs. Source: Nicholas Kayser-Bril (2011) on Prezi.**

The authors also point out that trust, not information, is the scarcest resource in today’s world. Trust is something that journalism very much depends on. “The trust market is still up for grabs. Most media players are still competing in the ‘attention market.’” Please read also a more gloomy discussion about business models of data journalism in chapter 3.8. Heinonen (2011) and Nippard (2011) discuss these themes, too.

There is also a strengthening trend of making data journalism without journalists. Some companies have already developed automatic news-creation systems based on open data. For example, the company [StatSheets](#) creates “machine-made” sports news. New companies are popping up in this field, including a startup company called [Narrative Science](#) that promises to deliver stories without human authoring or editing. Narrative Science offers not only sports stories but financial analyses, real estate analyses and local community content. It is evident that this kind of algorithm-based journalism is going to become more common in coming years.

### 3 Data journalism strategies in the US and elsewhere (Tanja Aitamurto)

#### 3.1. Role of data journalism in news organisations

For news organisations, data journalism is a tool for better journalism—journalism that follows normal news criteria. News organisations see data journalism as a tool to fulfill their basic tasks: providing relevant information to readers in meaningful ways. By using data in their stories, news organisations want to find hidden stories, as well as discover more facts to support more evident narratives.

News organisations perceive data journalism as a part of a transition from being rather passive news-and-information sites to more interactive news-and-information platforms. On these platforms, readers can not only consume journalism, but interact with data visualisations and raw data sets.

Furthermore, news organisations interviewed for this study express an interest in giving readers more ways to engage with the content, and receive as much information as possible from the data sets used in the story. As **Scott Klein**, the editor of news applications in ProPublica, puts it:

*“We want to make sure that our data tell not only narrative stories, but tell thousands of narrative stories so that anybody can come into a news application and find relevance to the story about their own experience and their own communities. So it’s not just dialysis facilities that are failing and succeeding in these ways nationwide, but I can look up a dialysis facility very near me and see how it’s doing and be able to make conscious decisions for myself and my loved ones about where I should go get dialysis if I need dialysis.”*

News organisations see value in transparency which data journalism adds to the journalistic process. Data sets used in reporting are often published online, and are therefore accessible for anybody to use. **Angelica Peralta**, editor of news applications at *La Nacion* in Argentina, describes this thinking thus:

*We believe that data show a lot of transparency, and if we can put our audience in contact with our data sets, that will give them more power to investigate themselves. So, we will do both things - release some articles and also put the database available to them.*

Some news organisations use the readers’ interactivity as one of the metrics with which they measure the success of their data journalism. They count the number of tweets and retweets about their stories, and the amount of activity on social media sites such as Flickr and Facebook, as distinct from the traditional metrics for measuring online journalism (namely, the number of readers and the time readers spend with stories).

### 3.2. Sources of data

News organisations have two major ways to obtain the data they use in data journalism. They obtain data from organisations, like government agencies, either by downloading publicly released data from websites, issuing specific requests for data from or scraping data from websites. In addition, news organisations use data sets which they have collected from several sources and then compiled themselves.

In the US, news organisations obtain their data from federal government agencies such as the Department of Health and Human Services and the Environmental Protection Agency, as well as from local city- and county-based agencies. News organisations also use data sets from less orthodox sources, including the World Bank, all manner of non-profit organisations and Wikileaks.

Even though there is a growing phenomenon of federal and local governance opening up data sets, news organisations have yet to use them very actively, because they often don't contain newsworthy content, and are therefore not of value to journalistic work.

The news organisations interviewed for this study hadn't used data markets such as [InfoChimps](#) for three primary reasons. First, the data journalism stories often originate in the newsroom so that a reporter first receives a data set which she or he wants to use in the story, thus eliminating the need for seeking further data in data marketplaces. Second, the news organisations want to get their data directly from sources without any intermediaries. Third, they don't want to pay for the data. As data editor of the Center for Public Integrity, **David Donald**, says:

*Our point of view if it's federal government data is that people have already paid for it through their taxes in the collection of the data, so we shouldn't have to pay again to get what we've already paid for.*

### 3.3. Workflow in data journalism

There are two primary ways to start a data journalism story. In the first option, a data set provides more information for a story topic that was already discovered by the newsroom. In the second, a data set serves as a starting point for the whole story. The former seems to be more common in the news organisations interviewed for this study.

Regardless of how the story gets started, the most important thing is the “journalistic sense of what is interesting,” as **Simon Rogers**, editor of the *Guardian* data team said in describing his organisation's principles in their approach to data journalism.

After obtaining the data, the data team typically cleans the data and analyses it based on the questions that need to be addressed in the story. Some news organisations use crowdsourcing through [Mechanical Turk](#) in processing the data. If needed, the data set is crossed with other data sets, to create new data from existing data sets. After the analysis,

visualisations and graphics are created, either by the data team or in a joint effort with the news organisation's other departments.

The data-analysis process can result in stories with visualisations, or in more permanent news applications on the Web, such as the [Bay Citizen's Bike Accident Tracker](#) or the [Chicago Tribune Nursing Home Application](#) (see the list of examples in Appendix 1)). Further, news organisations are looking into the possibility of developing mobile applications, as well, but the emphasis is still heavily on the Web.

The results of data analysis are being published alongside stories which use the data, either simultaneously or after the actual storyline begins. News organisations emphasize the storyline here: data visualisations and analyses are often a part of story, embedded in the storyline, rather than the story's main point.

Data sets used in analysis are usually published on the news organisations' website. Publishing the raw data is perceived as an integral part of the story process. This is primarily for two reasons. First, because of transparency. When the data sets are published online, anybody can run the same analysis on them and compare the results with the ones the news organisation got. Second, by publishing the data, news organisations invite readers—basically anybody—to work on the data, and create new information around them. Therefore, the simple act of publishing the data sets can benefit society in a sense broader than just publishing a news story. Others can find relevant information that goes unnoticed by either journalists or the news stories they produce.

In summary, the workflow can be roughly divided into following stages:

1. Identifying the gist for the story and the potential role for the data in the story.
2. Identifying and obtaining the right data sets to respond to journalists' questions.
3. Modifying the data to make them ready for analysis—e.g., correcting errors in the data sets.
4. Analysing the data with the right tools, and mashing the data with other data sets if relevant.
5. Producing the story: text, visualisations, interactive elements.
6. Publishing the data sets that were used in the analysis.
7. Inviting readers to participate by reusing the data, commenting on and sharing the story through applications in social media, and submitting more content through applications like Flickr, Facebook and Twitter.

### 3.4. Best practices in the workflow

News organisations have identified some best practices in the data journalism workflow.

1. Be agile and do the analysis quickly.
2. Provide an easy access to the data used in the stories. For example, create an API or other ways for efficient sharing. Always publish the raw data, because

everything you do with data has to be replicable, and you need to show that to people.

3. Reuse the data if possible. Same data can be used in a variety of ways, for several visualisations and analysis.

4. Add metadata. Code the data with institution and country codes, for example. This way, you'll make the data easier to find.

5. Create many channels for readers to participate—ways that don't require a time investment from the newsroom. For example, *The Guardian* has 500 submissions on its data Flickr site.

### 3.5. Data journalism teams, skills and tools

Data journalism teams in news organisations tend to be small, usually only two to five people, even in big companies. Just three people work in the *Guardian*'s data-journalism team, for example, and this is a news organisation that boasts a daily print readership of about 260,000, and oversees an online version that's said to be the second-largest English-language online newspaper after *The New York Times*. The teams' boundaries are flexible though, and oftentimes reporters from other departments participate in data-processing and analysis. Also, the graphics department may play a role by creating the graphics for the story, or helping out with visualisations. Typically, the data-journalism team works closely with the newsroom and reporters. Indeed, in many cases, data-journalism teams are physically located in the newsroom, rather than in IT departments.

Some news organisations, such as *The Guardian* and [Center for Public Integrity](#), use in-demand developers or third-party companies in cases where data processing requires a lot of time.

The teams generally consist of a combination of skills in journalism, Web development, data analysis, visualisation and statistics. Typically, the leader of a data team is a journalist with data-analysis skills. The rest of the team are alike—either journalists or non-journalists with skills in statistical analysis, data processing, visualisations, infographics, Web development and programming. In the future, news organisations see a need for mathematicians and statisticians on their data-journalism teams, or at least in an auxiliary role, providing support to the data-analysis process. Whatever the background and compilation of the team, it emphasizes the need for constant learning and experimenting with new tools. Nowadays, news organisations are using the following tools: Google spreadsheets for publishing data sets, Timetric, Graphinder, Google Fusion Tables, Google Refine and the R Program.

### 3.6. Data hackathons, hack days and codeathons

Several news organisations have organised events called hackathons, hack days and codeathons, designed to push forward projects around data journalism. In these events, news organisations typically provide data sets for participants to work on. They typically last from 12 to 24 hours, and the participants develop story projects and web applications based on the data sets provided. For example, *The New York Times* had a [TimesOpen Hack Day](#) which collected students and developers to invent new possibilities to work, share and display data. And the *Bay Citizen* organised codeathons with open data and [HSOpen in](#) a similar attempt to build networks between developers and media in Finland. The events are open for anybody to participate, and the participants are often programmers, web developers, designers, freelance journalists and open society activists. News organisations obtain the data used in hackathons from the local or federal government, organisations or they have compiled the data sets themselves.

These events have two primary goals. First, news organisations want to kickstart their data journalism projects. By inviting a diverse group to ideate and work together, data projects can take major steps forward. Second, news organisations want to develop innovation networks that feature the skills needed in a modern news organisation—web development, programming, creative thinking, etc.—around them. By investing in this innovation network, news organisations apply the principles of open innovation: by embracing the flow of innovation and ideas—both inbound to the organisation and outbound from the organisation—they increase innovation within the company (see Aitamurto & Lewis, 2011). Furthermore, news organisations attract a pool of talent into which they can tap as needed (e.g., for data-journalism projects).

News organisations perceive their needs for skills to be constantly changing throughout the agile development process. It's partly why they so closely follow the myriad application contests organised around open data. In the United States, government agencies are organising these application challenges, as are organisations and foundations supporting open society, such as the Sunlight Foundation.

### 3.7. Challenges in data journalism

Data journalists and data teams in news organisations are facing many challenges. One of the biggest is obtaining the data: to actually get the data and the format in which the data are provided.

Most of the news organisations interviewed for this study obtained their data from federal governmental agencies. Sometimes, the process of doing so gets complicated, and there is governmental resistance to releasing the data. In those cases, news organisations may use the Freedom of Information Act (FOIA), a federal law that allows for the full or partial disclosure of information and documents controlled by the US government.

News organisations either appeal to the FOIA, or try to obtain the data by using other methods, such as scraping, as has been the case at *The Chicago Tribune*, as described by Joe Germuska, a senior developer at the paper:

*“But for the nursing homes, to some extent, we actually scraped some of that data. So, we wrote programs that went to web pages and extracted data from them systematically. Because, the Department of Public Health, in that case, didn’t make that data available much more readily and we, we didn’t actually make a FOIA request for that data, because we just thought that they would be unresponsive and we were afraid that it would be a delay that we couldn’t bear.”*

Another big challenge for the news organisations is the format of the data. Even in cases where data were publicly available and obtaining them wouldn’t cause problems, the data are still often in a non-machine-readable format, e.g., in a .pdf file, or even as a .pedia file, which stalls the workflow to process the data. Therefore, processing and cleaning the data can take a lot of resources from the news organisation. The Center for Public Integrity has solved this problem so that sometimes they buy services from third parties to process the data. For example, when The Center for Public Integrity investigated the Department of Defense officials' travels, they hired a data-entry firm to process 20,000 travel documents that were in a .pdf format.

Furthermore, data sets almost always include errors, which journalists have to catch. Particularly when scraping data, the errors can occur in data sets.

Other problems that news organisations face in their data-journalism projects is the lack of metadata in data sets. This adds complications to the workflow when processing the data.

### 3.8. Business models and the new data ecosystem

There is no proven business model for data journalism. Overall, news organisations conclude that data journalism is so far a matter of journalistic value, and hold potential for impact rather than for revenue. Data journalism often takes more time than regular journalistic processes. Just the same, the news organisations interviewed for this study see future potential for data journalism as a revenue source, a potential that’s elaborated below.

Nowadays, the main way data journalism brings revenue to news organisations is with traffic. Thus, data journalism holds the potential to attract more advertising revenue. Data-journalism stories have, in some cases, become the most read stories in the publications in which they appear. However, this is not always the case. Often, there’s a big investment in time and resources associated with data-journalism stories, and the stories never become very popular.

Another way to increase the appeal of data journalism for the public is to use interactivity in visualisations. For example, *The Guardian*, among a couple of other publications, has discovered that readers spend more time with data-journalism stories than

they do with regular stories. If data journalism could attract a large number of readers, and make them stay longer on the site with the data stories and visualisations than with average news stories, the advertisers might be willing to pay a bigger prize for ads that appear with the data-journalism section.

Apart from advertising, news organisations envision some other future possibilities for business models in data journalism, such as the following:

- Premium model: Making people to pay for better visualisations.
- Becoming the go-to store for data for organisations as well as for individual citizens. At the store, extra value would be created by compiling *mega databases* of, for example, government data, looking for common identifications in data sets and making these databases able to talk to each other. The data store could sell access to these databases and provide analysis services and visualisations.
- Providing services for data analysing for organisations and companies.

Overall, news organisations see the value in data journalism as being about impact rather than traffic—impact that belongs to the role of journalism serving the public interest. **David Donald**, data editor from the Center for Public Integrity, describes the goals of data journalism as they relate to any other kind of journalism:

*One of the things that we may try to do when we investigate something that is either wrong □ doing, fraud or corruption, or that shows that the system isn't working, we want to see Congress take action. Change the law, close a loophole or pass some piece of legislation that would take care of the problem.*

Along with business models for sustainable data journalism, news organisations are looking for their place in a new ecosystem that is taking shape around data. In this ecosystem, there are new actors that provide similar content to news organisations, and thus, can be seen as competitors for news organisations' data operations. These players are often specialized in certain niches. For example, in the UK, a site called *Where does my money go* provides data and visualisations about tax spendings, and the Maplight Organization in the US provides data about politics. These organisations are typically non-profits, and are supported by foundations that promote transparency in society.

Furthermore, there is a new marketplace forming around data brokering and analysis. Players such as InfoChimps function as marketplaces for data sets. Currently, the company provides access to over 14,000 data sets, of which some are free and some are charged for. The company offers the first 100,000 API calls for data sets for free; after that, subscriptions are required.

The question news organisations are currently asking is: What is our role in this new emerging ecosystem for data? Could and should news organisations compete with data marketplaces and start to provide data analysis and visualisation services? Or should they focus on using data in their reporting in the most efficient and innovative ways, and thus provide better journalism for readers?

The strategies which respond to the question posed above are in making in news organisations all over the world. Compared to other data-analysis providers, news organisations perceive that they have an advantage in their ability to provide well refined information in a timely manner—the core competency of good journalism. As described by **Joe Germuska**, a senior developer at *The Chicago Tribune*:

*“We felt that simply replicating services that already exist without much story, people don’t have time for that. You know, it’s some of us who are kind of nerdier who might spend 10 or 15 minutes poking around with some data, but most people, they’re going to ask, you know, ask themselves what does this mean to me? And, so most people won’t take that long before they answer that question and then move on, because they have other stuff they want to do.”*

News organisations feel that the best way to connect their readers with data is to create data journalism that is easy to digest, and has high relevancy and many entry points for the readers. However, if news organisations decide not to build and maintain extensive data sets and create useful services on top of those data, the other players on the field will fill in the gap quickly. Readers will learn to find the data about relevant issues such as government spending in other websites beyond those affiliated with news outlets, and spend their minutes there.

### 3.9. Vision: The future of data journalism

Based on the analysis done for this study, the following vision of the future of data journalism can be postulated. As the cost of data analysis continues to decrease, the skill requirements of data analysis keep growing, and open data is made more available by governments, companies and organisations, more and more data journalism will occur.

However, it is not only news organisations that will be providing information and services based on data, but new players will enter the field, such as independent publishers and transparency organisations, both of whom are expected to be increasingly using data. Most likely, they will take the space in providing data-analysing services to third parties rather than news organisations doing that.

News organisations, in turn, will focus on analysing data for journalistic purposes. In terms of workflow, journalists will adopt new, inexpensive data-analysis tools. In the long term, data journalism will become one of journalists’ basic tools for reporting. Better tools and increased skills will make data analysis cheaper and more efficient, and this will lead to journalists discovering story topics from larger data sets, cross analysed with other larger and larger data sets.

A symbiosis of the players in this emerging data ecosystem will develop. News organisations will provide data journalism based both on their own data sets, and link to other data sites and marketplaces. These sites will feature data-journalism stories, and link to analysis on news organisations' websites.

## 4 Conclusions

In the field of public data, there are vast differences among countries in both their governing laws and practices. Only 5.5 percent of countries do not publish reports related to their budgets, institutions and respective procedural documents while 64 percent of governments do not publish a list of civil servants and their salaries (Alonso, 2011).

When compared globally, the three largest categories of governments that are actively opening their data to the public emerge as the following: countries fighting corruption, post-authoritarian regimes that erect legislative barriers to promote open data, and countries that publish data due to tradition or conviction (Alonso, 2011). The majority of European countries have opened their data resources quite slowly, and this in spite of the directive given in 2003 by the European Commission.

Data journalism is finding its way to newsrooms. Reporters at US daily newspapers routinely turn to local, state and federal government websites to hunt for data that they can use in their stories, a recent [survey](#) by the Reynolds Journalism Institute found. The reporters ranked state data as the most accurate, and local as the least. The reporters said that their local governments were the most transparent and accountable, while the federal government ranked the lowest on those counts. According to the results of the survey, campaign-finance and lobbying data are the most available online; workplace-safety-inspection records were ranked as the least available.

An ongoing survey done by the non-profit European Journalism Centre of journalists showed that most of the respondents were already engaged in data projects. The most urgent training is needed in the developing of news stories from data. Also visualising and analysing data were mentioned as crucial skills in this field. Journalists also express an interest in learning better ways to find, verify and clean data.

Data journalism has become an integral part of several news organisations' online strategies. News organisations see data journalism as a part in the transition moving from being news-and-information sites toward becoming more interactive news-and-information platforms. In their new role, news organisations want to provide their readers with ways to interact with the content, and refine it further. Data journalism provides an opportunity for this.

According to the interviews conducted for this study, data journalism is perceived as a tool to produce better journalism with the help of statistical methods, visualisations and interactive ways to present information. Journalists see data journalism as a way to find hidden stories—stories that wouldn't otherwise get told. News organisations hope to serve public interest better with data journalism and, thus, have an impact for the better in the society.

Furthermore, data journalism is perceived as a way to increase transparency in journalistic processes. The data sets used in data stories are usually published online, and are accessible for anybody to do analysis on. This transparency potentially increases the public's general level of trust in journalism and media.

Data journalism adds new requirements to a journalist's profession. Now they need skills for data analysis, web development and community management, as readers are invited to participate in the stories. However, the most important factor for successful data journalism is the journalistic sense of what is relevant and interesting, and what questions need to be asked. Only after these are determined can data's usefulness kick in to address those questions.

News organisations are looking for sustainable ways to support data journalism, but viable business models have yet to be discovered. News organisations have visions of becoming number-one data stores and of adding the provision of data analysis to their oeuvre of paid service offerings. The most likely future scenario will see news organisations existing in a state of symbiosis with other data service providers, booming in the new ecosystem evolving around this area.

## Appendixes

### **Appendix 1. A selection of data journalism cases, especially of local journalism (Pauliina Lehtonen)**

In general, data journalism relates to the question how to visualise complex issues; how to tell a story with visualisation. A few examples have led the way by showing how data utilization can produce news pieces of global relevance. For instance, the famous example from *The New York Times*, *Der Spiegel* and *The Guardian*, publications who had access to US military's internal data regarding the conflict of Afghanistan. The data were obtained by Wikileaks. See Afghanistan War Logs:

<http://www.guardian.co.uk/world/the-war-logs>

<http://www.nytimes.com/interactive/world/war-logs.html>

Another famous example, again from *The Guardian* with the help of crowdsourcing, published data that revealed MPs' expenses:

<http://mps-expenses.guardian.co.uk/>

<http://www.guardian.co.uk/politics/mps-expenses>

<http://www.guardian.co.uk/news/datablog/2009/may/13/mps-expenses-houseofcommons>

The above examples have relevance on both a national and a global level. However, the open data cases that are introduced in this report focus more on the level of local journalism. The following open data cases are organised here in terms of their visualisation. Many of the examples utilize various visual tools, but here they are grouped based on their primary visualising tool: namely, if they utilize an interactive map, graphics or a table.

The examples have been selected based on their connection to a local level, but their scope and relevancy can be vaster. However, the examples depict how open data could be used in media houses to provide applications and services for audiences that would be of use at the local level. The examples illustrate, for instance, how to create visual presentations on issues such as education, health care, traffic and safety. The list is by no means exhaustive, but it presents a valuable sample of current data journalism examples.

In addition, a few generally interesting links related to the development of data journalism are listed after the following data-journalism examples.

### **Interactive maps**

- **Mapping the riots with poverty** (*The Guardian*) is an application that brings together data on deprivation and the riot incidents that took place in England in August 2011. The darker red colour represents poorer areas and the blue colours are the richer areas. By clicking on icons, the user can read a text box that provides information on the place and the incidents, such as location details, details about the local authority and an explanation of exactly what happened. There is also a link to related newspaper stories.

<http://www.guardian.co.uk/news/datablog/interactive/2011/aug/10/poverty-riots-mapped>

Here, the user can find key facts and figures presented with pie charts and other means:

<http://www.guardian.co.uk/news/datablog/2011/aug/09/uk-riots-data-figures>

Here's info on how to interpret the map:

<http://mattstiles.tumblr.com/post/8696184447/mapping-london-riots-deprivation>

This interactive map shows how the riots spread during the days in August:

<http://www.guardian.co.uk/news/datablog/interactive/2011/aug/09/uk-riots-incident-map#animation>

- **Homicide map of Chicago** (*The Chicago Tribune*) shows data on homicides, including date, block and community area, cause and victim name. It also provides links to newspaper stories of the incidents. The application utilises maps and tables.

<http://homicides.redeyechicago.com/date/2011/8/>

- **The Homicide Report** (*The Los Angeles Times*) is an interactive map and database that track homicides in Los Angeles County. It offers a forum for readers to discuss violence in

their communities. The maps provide data organised by various categories, such as ethnicity, age, neighbourhood, gender and method of death. Readers can link to the original homicide report to read archived comments and the original posts. It shows photos of the victims and their names, ages and details of the homicide.

<http://projects.latimes.com/homicide/map/>

- **Mapping L.A.** (*The Los Angeles Times*) is a service for local neighbourhoods in the Los Angeles County. It provides maps and information about demographics, crime and schools in neighbourhoods across the county. Users can choose a neighbourhood and click to see a detailed map with information about schools, income, demographics and news from the area.

<http://projects.latimes.com/mapping-la/neighborhoods/>

- **EveryBlock** (msnbc) combines different types of local news, including public records like crime reports, building permits, restaurant inspections, records of neighbours' discussions, people's photos from the neighbourhood and media mentions.

Every block shows what is happening within it. People can join in neighborhood conversations, and read news reports and event listings. The site also offers the ability to order an e-mail of the news of the neighbourhood.

<http://www.everyblock.com/>

- **The English school system** (*The Guardian*) is an interactive map that shows how England's schools break down by number of schools and students, race, free school meals, numbers of children whose first language is something other than English and appeals by parents for secondary places.

<http://www.guardian.co.uk/news/datablog/interactive/2011/jun/22/english-school-system-interactive-map>

- **Clout college application** (*The Chicago Tribune*) shows how many students from each Illinois high school applied, were admitted and attended each college campus, and how many were placed on the category I clout list each year. The app also utilises bar graphs.

<http://cloutcollege.apps.chicagotribune.com/cloutschools/#datagrid-for-loyola-academy>

- **2010–2011 school immunization data application** (*The Chicago Tribune*) provides a guide to vaccination levels in private and public schools across Illinois. The application calculates the percentage of students with required vaccinations against various diseases at the beginning of the 2010-2011 school year. It utilises bar graphs and tables to represent data about the number of protected students, students with medical exemptions, students with religious exemptions, etc.

<http://schools.chicagotribune.com/immunization/>

- **Fixing D.C.'s schools** (*The Washington Post*) is a service that provides information on the schools in Washington regarding such factors as crime, health and safety. The application

shows records for each category of data, and allows users to search for a particular school. The application was formed as a part of a *Washington Post* series that investigated public schools. School details show data on school populations, passing grades, qualified teachers, crime reports, school maintenance, school reports, etc. It utilises tables in presenting the data.

<http://www.washingtonpost.com/wp-srv/metro/interactives/dcschools/scorecard.html>

- **Chicago Mayoral election 2011** (*The Chicago Tribune*) provides information on voting rates in different parts of the city. In addition to maps, it utilises bar graphs. Precincts are coloured to show which candidate received the most votes.

<http://media.apps.chicagotribune.com/election-maps/mayor-2011/index.html>

- **Census 2010 Interactive Map: Voting Age Population By County** (*The Texas Tribune*)

The data provide a snapshot of the adult population (age 18 and older) of each county, showing how many people of voting age live in each area. This interactive map displays the totals in different shades, with darker colours signifying higher rates of adult population growth by race and Hispanic origin. Users can select a county to pull up a table of its figures.

<http://www.texastribune.org/library/data/census-2010-voting-age-population/>

- **Red light camera intersections** (*The Chicago Tribune*) is an application that provides data on traffic accidents at red light camera intersections. By clicking an icon on the map, users can access information in the .pdf file of weather and light conditions, type of collision, crash date, etc.

<http://www.chicagotribune.com/news/local/chi-091122-redlight-map,0,1513125.htmlpage>

There are also links to news stories that discuss the red light collisions:

<http://www.chicagotribune.com/news/local/chi-red-light-windfall-gallery,0,737477.storygallery>

- **Bike Accident Tracker** (*The Bay Citizen*) shows bike accidents on an interactive map along with statistics on interactive charts. The application lets users search for information about road conditions, lighting, accident type, violations and fault determinations. Links to related news stories are also provided.

<http://www.baycitizen.org/data/bike-accidents/>

- **Bicycle accidents** (*Seattle Times*) offers details of the city's bicycle collisions and their particular locations.

<http://seattletimes.nwsourc.com/flatpages/local/bicyclecollisionsinseattle.html>

- **Cardiff recycling rates** (*The Guardian*) shows recycling rates on a map where users can click on each ward to find out its recycling rate.

<http://www.guardian.co.uk/cardiff/2010/dec/06/map-cardiff-recycling-rates-by-ward>

Here's more information on how the map was done:

<http://hrwaldram.wordpress.com/2010/12/07/grappling-with-google-maps-creating-a-ward-map-of-cardiff/>

- **Landslides and logging** (*Seattle Times*) reveals information of landslides that were started after a logging company clear-cut slopes. It shows landslides in clear-cut, other slides and rainfall measure. The application also introduces photos and information boxes with such details as hazard zones and steepness before and after the logging.

<http://seattletimes.nwsourc.com/flatpages/local/landslidesintheupperchehalisriverbasin.html>

Here, find information on how the investigation was conducted:

[http://seattletimes.nwsourc.com/html/localnews/2008048858\\_logginghow13.html](http://seattletimes.nwsourc.com/html/localnews/2008048858_logginghow13.html)

- **Gypsy and traveller caravan sites** (*The Guardian*) are interactive maps that provide data on gypsy, traveller and travelling showpeople caravan sites by local authority. It illustrates which areas of England have the highest concentration of this kind of activity and shows, for instance, caravan capacity, number of pitches, site opening dates, etc.

<http://www.google.com/fusiontables/embedviz?viz=MAP&q=select+col0,+col1,+col2,+col3,+col4,+col5,+col6,+col7,+col8,+col9+from+1068644+&h=false&lat=51.767839887322125&lng=-2.08740234375&z=6&t=2&l=col2>

Data are also presented in a table format here:

<http://www.guardian.co.uk/news/datablog/2011/jun/29/gypsy-sites-england-local-authority>

- **Enjoy England** (*The Guardian*) is an application that provides information on great places in England. Users can upload photos, share tips and read news articles regarding the places.

<http://www.guardian.co.uk/enjoy-england>

- **Suburbified** (*The New York Times*) is an application that features a “Living In” column that profiles one suburban outpost or New York City neighbourhood. The colour of the dots on the map represents the date of the article. Red dots are the newest and orange dots are for older articles. The “show phrases” button at the bottom right of the map reveals a list of some common phrases that occur in more than one article. There are links to the original articles.

<http://suburbified.com/>

- **MovieAB** (*The New York Times*) info pages include excerpts from *NY Times* movie reviews and provides links to same. <http://www.movieab.com/>

Suburbified and MovieAb applications are based on *The New York Times*'s archives, which have been opened. *NYTimes.com* does not support or endorse these applications, but the apps work as examples of how readers can use the *Times*'s APIs. See the data clearinghouse of *The NY Times* here:

<http://developer.nytimes.com/page>

## **Graphics**

- **News Traces** (BBC, Guardian) is an application that allows users to search through BBC News headlines from the last 10 years. By typing a search term, they can instruct the application to show the pattern of reports as they unfolded over time.

Individual news items are displayed as small blocks laid out in a series of bars, so it is easy to see a keyword's history. Chronological bars composed of blocks of individual news items are revealing. When you roll over a news block, the title and summary outline the context of the story, and users can click links through to the full text of the BBC news page.

[http://www.data-art.net/guardian\\_newstraces/](http://www.data-art.net/guardian_newstraces/)

- **Do no harm** –project (*Las Vegas Sun*) focuses on hospital care in Vegas and provides information on healthcare flaws. The site offers source documents, reader stories, multimedia and interactive features.

<http://www.lasvegassun.com/hospital-care/>

For instance, various local hospitals' incidence of surgical injuries are shown on an interactive chart that uses bar graphs to present details.

<http://www.lasvegassun.com/hospital-care/surgical-injuries-interactive/>

The preventable harm application shows harm data acquired at a Las Vegas hospital, such as a foreign object retained, poor glycemic control or surgical site infection.

Users can search data by hospital name or event.

<http://www.lasvegassun.com/hospital-care/events-interactive/>

- **Teacher ratings** (*The Los Angeles Times*) is a database of value-added scores for Los Angeles Unified elementary schools and teachers. The application is used to estimate how schools contribute to students' learning. Ratings are based on the performance of students who took the California standards tests in math and English. Value-added measures the difference between the expected progress of the student body and its actual performance. The bars along the bottom represent the range of teacher ratings.

<http://projects.latimes.com/value-added/>

Information on the series of news stories is also provided:

<http://www.latimes.com/news/local/teachers-investigation/>

- **The Opportunity Gap** (Pro Publica) is an application that examines if states in the US provide high-poverty schools equal access to advanced courses and special programs. This application is the first to take a nationwide picture of which courses are being taken at which schools and districts across the country. Users can search for school names and then review and compare results presented in graphs and tables.

<http://projects.propublica.org/schools/>

- **Precios de canasta basica** (*La Nacion*). The visualisation presents data on how the prices of grocery products have developed over time. The interactive visualisation utilises charts and other graphics, and the visualisation is included as part of a news story.

<http://www.lanacion.com.ar/1381455-como-subieron-los-precios-de-la-canasta-basica-en-los-ultimos-anos>

**El sube y baja de las hortalizas** (*La Nacion*) is an interactive visualisation that uses chart graphics, to provide an account of how the prices of vegetables, such as potatoes, tomatoes and carrots have developed in relation to societal development. The visualisation is included as part of a news story.

<http://www.lanacion.com.ar/1382225-el-indec-se-aferra-al-precio-de-la-papa>

## Tables

- **Nursing home safety reports** (*The Chicago Tribune*) is a site that provides safety reports on nursing homes in Illinois. It includes data that are not presented on government sites, such as the number of residents who are convicted felons and sex offenders, crimes reported at Chicago nursing homes and fines levied because of deficiencies in care.

<http://nursinghomes.apps.chicagotribune.com/>

Provides links to news stories on the topic are here:

<http://www.chicagotribune.com/news/watchdog/nursinghomes/>

- **Dollars for doctors** (Pro Publica) produces data on the compensation medical companies have paid out to doctors. This database allows patients to search for their doctor, medical company or state (number of payments, amount). The application uses tables and pie charts.

<http://projects.propublica.org/docdollars/>

Here, they provide information on how the data were gathered:

<http://www.propublica.org/article/about-our-pharma-data>

- **Green buildings in Washington D.C** (*The Washington Post*) provides data on environmentally friendly buildings in the city:

<http://www.washingtonpost.com/wp-srv/metro/special/green.html>

Link to the news story:

[http://www.washingtonpost.com/local/dc-among-greenest-for-buildings/2011/04/13/AFFk8HZD\\_story.html](http://www.washingtonpost.com/local/dc-among-greenest-for-buildings/2011/04/13/AFFk8HZD_story.html)

- **Restaurant inspections** (*Seattle Times*) application shows the latest inspection results directly from the public health department. The application shows facility details, inspection type and scores in a pop-up window.

<http://seattletimes.nwsourc.com/flatpages/local/kingcountyrestaurantinspections.html>

**A few relevant links on data journalism**

(e.g. leading data journalists' blogs):

*Guardian* data blog & data store:

<http://www.guardian.co.uk/data>

*Guardian* open platform:

<http://www.guardian.co.uk/open-platform/dataviz>

The data desk of *LATimes*:

<http://projects.latimes.com/index/>

Searchable databases developed by *The Washington Post* and other local and national organisations and agencies:

<http://www.washingtonpost.com/wp-srv/metro/data/datapost.html>

Online journalism blog:

<http://onlinejournalismblog.com/category/databases/>

Tony Hirst's blog:

<http://blog.ouseful.info/>

Jonathan Stray's blog:

<http://jonathanstray.com/a-full-text-visualisation-of-the-iraq-war-logs>

The blog of David McCandless:

<http://www.informationisbeautiful.net/>

The blog of Anthony de Barros:

<http://www.anthonydebarros.com/>

The site of Adrian Holovaty:

<http://www.holovaty.com/>

Hacks/Hackers:

<http://hackshackers.com/>

Nieman Lab:

<http://www.niemanlab.org/tag/data-journalism/>

## Appendix 2. Subjects interviewed by Tanja Aitamurto

1. Simon Rogers, *The Guardian*, editor of the Guardian.co.uk/data, United Kingdom.
2. David Donald, data editor, iWatch News, Center for Public Integrity, United States.
3. Scott Klein, Pro Publica, editor of news application, United States.
4. Barbara Grady, *Oakland Local*, United States.
5. Shane Shafflet, *Bay Citizen*, software developer and reporter, United States.
6. Jeffrey Friedman, research director, Maplight, United States.
7. Angelica Peralta, *La Nacion*, editor of news applications, Argentina.
8. Joseph Germuska, *The Chicago Tribune*, senior developer, news applications, United States.
9. Peter Aldhous, *The New Scientist*

## References

Aitamurto, T. & Lewis, S. C. (2011). Open APIs and news organizations: A study of open innovation in online journalism. Paper presented at the International Symposium on Online Journalism, Austin, TX, April 1, 2011. Accessible at <http://online.journalism.utexas.edu/2011/papers/AitamuroLewis2011.pdf>

Alonso, P. (2011) Open data: A new international norm? in Owni.eu 02.08.2011. <http://owni.eu/2011/08/02/open-data-a-new-international-norm/>

Bounegru, L. (2011) What are the training needs for data-driven journalism. Preliminary survey results. Open Knowledge Conference. Berlin 30.06.2011. Slides on Slideshare: <http://www.slideshare.net/lilianabounegru/preliminary-results-of-ejc-survey-on-training-needs-for-data-journalism>

EJC (2010) Data Driven Journalism: What is there to learn? A paper on the data-driven journalism roundtable held in Amsterdam on August 24, 2010. European Journalism Centre.

Guardian Datablog (2010) How Canada can become an open data and data journalism powerhouse. Guardian.co.uk

Heinonen, A. (2011) Uusi rooli uutismedialle: datasolmu. Suomen Lehdistö 5/11 p. 19

Herzog, D. (2011) Highlights from reporter data use survey. rjionline.org. 29.08.2011 <http://www.rjionline.org/news/highlights-reporter-data-use-survey#>

Kayser-Bril, N. (2011) Presentation on data journalism. Prezi 13.04.2011 <http://prezi.com/e7tfgnu2zpuarepublica-xi-110413/>

Lehtonen, P. (2011) Open data in Finland – Public sector perspectives on open data. Next Media – programme, WP3, Hyperlocal, deliverable number 3.2.1.2

## TRENDS IN DATA JOURNALISM

Lorenz, M. (2010) Status and Outlook for data-driven journalism. In EJC (2010) Data Driven Journalism: What is there to learn? A paper on the data-driven journalism roundtable held in Amsterdam on August 24, 2010. European Journalism Centre.

[http://mediapusher.eu/datadrivenjournalism/pdf/ddj\\_paper\\_final.pdf](http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf)

Lorenz, M. & Kayser-Bril, N. & McGhee, G. (2011) Media companies must become trusted data hubs. Owni.eu 28.02.2011

<http://owni.eu/2011/02/28/media-companies-must-become-trusted-data-hubs-catering-to-the-trust-market/>

Nippard, C. (2011) Data visualizations emerge across newsrooms, online publishers.

Deutsche Welle. [www.dw-world.de](http://www.dw-world.de) 19.04.2011

Schellong, A. & Stepanets, E. (2011) Unchartered Waters. The State of Open Data in Europe. CSC. Public Sector Study Series 01/2011