Data-driven journalism: What is there to learn?

A paper on the data-driven journalism roundtable held in Amsterdam on 24 August 2010.

With additional material on data tools, DDJ innovators, and recommended websites and articles. The immediate goals are to improve access for interested journalists and to identify training needs for the future.

European Journalism Centre (EJC), 2010
www.ejc.net
# Table of Contents

Data-driven journalism roundtable, Amsterdam, 24 August 2010

**This Paper** .......................................................................................................................... 5
  Introduction: Why data matters for journalism ................................................................. 6
  Mirko Lorenz: Status and Outlook for data-driven journalism ...................................... 8

**Session 1: Data Production, Usage and Integration** ...................................................... 18
  Jonathan Gray: Open Data and Data Driven Journalism .............................................. 19
  Lorenz Matzat: Weatherstations - Citizen-Apps, eParticipation and Data journalism .. 23
  Richard Rogers: Tools for Data ....................................................................................... 26
  Simon Rogers: Free our Data! ......................................................................................... 29
  Tony Hirst: How to make the Data Flow ...................................................................... 32

**Session 2: Data Visualization** ....................................................................................... 34
  Stefan Fichtel: Data-driven visualization ........................................................................ 35
  Frank van Ham: How to use Data Visualization ............................................................ 37

**Session 3: Storytelling with Data** ................................................................................ 39
  Cynthia O’Murchu: Datastories ..................................................................................... 40
  Alan McLean: The data is dead, long live the data! ....................................................... 42
  Eric Ulken: Building a data desk ................................................................................... 44
  Nicolas Kayser-Bril: Data-journalism agency ................................................................. 47
  Gavin Sheridan: Data-driven transparency .................................................................. 51

**Session 4: New Formats for Presenting Information & Stories** .............................. 53
  Stijn Debrouwere: Baking a better cake ....................................................................... 54
  Burt Herman: Storify - making sense of the world ....................................................... 56
  Julian Burgess: The Role of Analytics .......................................................................... 58

**Future perspectives for DDJ** ......................................................................................... 60

**The #ddj lists 2010** ........................................................................................................ 67
  Data People (2010) ......................................................................................................... 70
  Data Examples ................................................................................................................ 71
  Data Articles .................................................................................................................. 74
  Data Books .................................................................................................................... 75
  Data Companies .......................................................................................................... 76
  How to connect and stay in touch .............................................................................. 77
  Acknowledgements ....................................................................................................... 78
    Contributors ............................................................................................................. 78
    Imprint ....................................................................................................................... 78

**Contributors**.................................................................................................................. 78

**Imprint** ............................................................................................................................. 78
“I think data-driven journalism is one of the big potential growth areas in the future of journalism. A lot of the forward-thinking discussion about the future of news focuses on the ‘glamorous’ possibilities, like video journalism and interactivity, but I often see data journalism being ignored.

In fact, I believe it is journalism in its truest essence: uncovering and mining through information the public do not have enough time to do themselves, interrogating it, and making sense of it before sharing it with the audience. If more journalists did this (rather than relying on ‘data’ from press releases) we would be a far more enlightened public.

My message to the next generation of journalists - or any journalist looking for a new niche or direction - would be to learn the skills and tools of data interrogation. It’s not glamorous, but it’s a skill not many journalists have, and one which will give one an edge in the market.”

Adam Westbrook, author of ‘Next Generation Journalist’, source: Interview by EJC, 2010
“One of our big goals in the storytelling process is to humanize the statistics. It’s hard for people to care about numbers, especially large numbers. How do you get your head around the death of 800,000 people in the Rwandan genocide? I think if you meet the individuals - see and hear the stories of the survivors - you can gain a better insight into the tragedy.”

Brian Storm, Mediastorm
Source: E-Mail interview by EJC, 2010
This Paper

This collection of resources presents an overview of data-driven journalism: what it is today and what it might become in the future.

The idea is to provide both experienced journalists and newcomers with a well-structured primer, while breaking down some of the barriers holding back coders and non-coders alike from starting to experiment. We need better solutions, good platforms and better reporting.

Data is not an entirely new field, nor is storytelling. But deep changes are afoot: while journalism’s old business models are crumbling, working with data provides new and attractive opportunities. Should you agree or disagree with anything in this paper, please drop us a line. It can only improve.

Summary goals
The aim is to provide food for thought and practical information, similar to the great papers of Mindy McAdams (Reporters Guide to Multimedia Proficiency) and Adam Westbrook (6x6 series, Next Generation Journalist). To provide a reference point for journalists adapting to new workflows, we present the most important information in one place: as a first overview on what data-driven journalism might mean and how it can provide a new perspective for journalists.

Additional material
An important addition here are chapters providing links to good articles on data-driven journalism, a long list of tools that allow you to play with data and dig deeper, even if you don’t know how to code.
Introduction: Why data matters for journalism

Ten, even five years ago, the use of data as a basis for reporting was difficult and costly, requiring IT skills far beyond what is common in the media. Databases were used mainly by investigative journalists. Editors and reporters usually relied on information provided by outside sources.

Today there is a notable change. Collections of data are becoming available online, often for free. There is a whole stack of tools for digging into ‘big data‘. Open source tools allow navigation and analysis of large amounts of data rather quickly. There are online applications that allow us to share and visualize data.

Developing the know-how to use the available data more effectively, to understand it, communicate and generate stories based on it, could be a huge opportunity to breathe new life into journalism. Reporters can find new roles as ‘sense-makers’ by digging deep into data, in turn making journalism more socially relevant. If done well, delivering credible information and advice could even generate revenues, opening up new business models beyond subscriptions and advertising.

In this context, the European Journalism Centre (EJC) in collaboration with the University of Amsterdam organized the first roundtable on data-driven journalism. The one day event gathered specialists in fields that intersect with data-driven journalism: data mining, data visualization and multimedia storytelling to discuss the possibilities of this emerging field, examine and understand key tools and workflows, while sharing their budding expertise in data-driven journalism. What can we learn from current projects? How can we integrate existing tools into journalistic workflows? What skills are needed to enter this field?

This paper tries to provide an overview, based on what was discussed in Amsterdam. Beyond this first summary, the idea is to publish a final paper that can be used by other journalists interested in working with data to get a deeper understanding, see examples of what can be done and learn about the tools that are available.
**Do journalists need to become programmers?**
This question has been debated extensively on many websites covering journalism. If the role of data grows as a basis for reporting and storytelling, then what are the skills needed to do this? What can be done with data? How will newsrooms change? What are the platforms and tools needed to do this?

There is a lot of confusion about this, with more questions than answers. To move towards the latter, the roundtable presented speakers from a variety of backgrounds including writing, information architecture, and visual design. The following chapters include a range of perspectives from various nations and disciplines, all with a shared interest in making better use of data.

The roundtable was planned and chaired by Mirko Lorenz, DDJ Project Leader for the EJC and a freelancing information architect from Deutsche Welle, who works on innovation projects. The event was organized by the EJC and partially funded by the Dutch Ministry of Education, Culture and Science.

**Websites**
http://www.ejc.net
http://community.ejc.net/group/datadrivenjournalism
Yet another buzzword is making waves: data-driven journalism or DDJ for short. It is based on the hope that journalism will find new structure and meaning by bringing data into the reporting workflow. It is debatable whether this is something really new or just a new flavour - but there is something in the air that hints that this new trend might be bigger and lasting for some time.

Interest in this field has grown only recently. The goal of the roundtable in Amsterdam was therefore to debate the status and outlook for journalists. At this stage, very basic questions have to be answered: are we all talking about the same
thing? Where do we agree and (probably just as important) where do we disagree?
Is the use of data an opportunity, and if that is the case, what are the barriers keeping us from actually doing it?

Another question is what we are missing. As of 2010 data is used increasingly to visualize very complex issues. The publication of the Afghanistan war logs by the New York Times, ‘Der Spiegel’ and the Guardian have raised awareness that better use of data might lead to very big stories. Similarly, the way the Guardian handled the expenses scandal of British MPs in 2009 has sparked interest in various elements might be involved in this. Think crowdsourcing. Think opening up large stores of public data and turning it into open data that everyone can share. Think uncovering scandals and being able to prove it with numbers. Think providing people with dependable services, helping them to decide when buying, insuring, participating or taking life choices.

To do that journalists will have to learn new tricks. They have to get used to working with tools that will help them to make data flow. Quite a few commentators on media and journalism think that this is a major obstacle because the average writer is not good with numbers. Is this true?

This first roundtable in Amsterdam provided an opportunity to discuss this based on practical work that has already been done. What were the experiences of the people that have actually taken a raw chunk of data or unsorted documents, when trying to make sense of it? Ultimately, we need to describe in clear terms what has to be learned before data-driven journalism can be integrated fully into newsrooms.

What?
Data in the house
Journalists have always worked with data and technology. Facts are data in some form or another, and facts are the basis for any story that can claim to be a journalistic work. In the past heavy number crunching was done elsewhere, in big rooms with specialists feeding machines. Journalists used the end result of this process, ranging from statistics, studies or stock quotes to new findings in science. Only very few journalists worked directly with the raw data at the start of this process.
Sure, there were specialists practising ‘CAR’ (Computer Assisted Reporting). They were trained and had technical skills. But CAR was and is primarily a technique, not a process affecting the whole workflow of journalism in a fundamental way. This is not about devaluing CAR. The use of computer searches in large databases remains an extremely important skill for investigative journalism. But data-driven journalism is rather different, making CAR one element in the chain of future events.

Today data-driven journalism can be defined as a workflow, where data is the basis for analysis, visualization and - most importantly - storytelling. With cloud computing, powerful PCs on the desktop and high bandwidth, we all have the potential to use technology to filter and process data that would have been a sensation just 15, 10, maybe even 5 years ago.

It is still unclear, however, what platforms are needed to achieve that. Today’s content management systems are page-orientated: they help you put together an article. But they greatly ignore the data in the text. Facts, which have been collected earlier, become a mess in this process. One other thing is business models: where and how could journalists make a living based on data? What formats would create

Source: Data-driven journalism - Status and Outlook
new income streams, either from subscriptions, advertising or plain selling of ‘information nuggets’?

**Unsolved questions of what DDJ might become**
Journalism is under siege. Newspaper business models that worked for more than 100 years are crumbling. Advertising is going after eyeballs, which means that more and more budgets have shifted to the web. Classifieds? Gone to Craigslist. Job offerings? Monster.com. And so on.

Brian Storm, a multimedia journalist who has influenced so many of us with his superb storytelling on ‘Mediastorm’ boils it down to this: “Don’t be the noise in the middle”. You can set out to create hilarious videos on YouTube with cats spinning around at one end of the market. Or go in the other direction towards artful, truth-orientated stories inquiring into the ‘state of the human condition’, as Storm describes his goals.

Good writing is simply not enough; with millions of bloggers out there you will find many voices from people who might not be journalists by trade or institution. But people understand and relate to good information.

The core technology of journalism has long been the printing press. Until recently, only media companies were able to print millions of copies overnight, distribute them into the home and create awareness for issues (and advertising). The Internet has ended this.

Understanding data might be an avenue to the future. But before starting this, we should ask questions where this might bring us. Skipping the questions would be like hopping into a Ferrari or a Porsche, thinking that driving this machine will be no different from a normal commuter ride, given that there is a steering wheel, a gas pedal and a brake. Beware: once you get up to speed the physics change dramatically - so the first road bump or curve might result in disaster.

Not asking any questions about what data-driven journalism might become would be a repetition of what people did in the ‘New Economy’, hopping into companies that by and large collapsed once the fever was over.
How can we regain relevancy with journalistic content?
A good example of what to look comes from the ‘New York Times’: go to Google in almost any country and punch in the search term ‘rent or buy’, chances are high that on the first page and even in the first listing will be a link to an interactive graphic from the New York Times. The little tool allows you to quickly fill in some personal data such as price of house, interest rate, etc. and it will tell you after how many years it is better to buy instead of rent.

This is an interesting tool. First, because it’s very usable. Second, because the New York Times enters a new field here, using state of the art browser technologies to help people make a decision on whether or not to buy. The reason why this is so interesting is partly because buying (and keeping) a home is the biggest financial deal most people will make in their lives. A journalistic publication by nature has different interests when providing such a tool than, say a real estate agent. Instead of trying to sell right away, the medium works as an intermediary, a consultant, an advisor. Could we build on such examples and develop new platforms? If you think real estate is essentially boring, look at Curbed. The blog is a big success, letting people peek into flats, condos and houses that are on the market. Connect this with a growing number of interactive tools that provide very clear answers based on a constant inflow of data, and you see one of the perspectives data-driven journalism might provide. Instead of commenting and reporting on the side, media brands could be the first destination to inform yourself before signing a cheque for anything.

So data-driven journalism can be viewed as a process of refinement, where raw data is transformed into something meaningful. As a result the value to the public grows, especially when complex facts are boiled down into a clear story that people can easily understand and remember.
To make this happen and let data-journalism grow into something bigger, we first need to ask ourselves some tricky questions. How can we use data? What systems do we need?

There is a proverb in Spain: ‘before you jump high, make sure that you stand on solid ground’. And this is why the questions asked might be puzzling to many, but should be discussed in depth.

**Are we solving ‘puzzles’ or ‘mysteries’?**

As a case in point, look at a story Malcolm Gladwell wrote in the ‘New Yorker’ in 2007. For some reason the text has two headlines: ‘Open secrets’ is how it is listed on Gladwell’s website; once you download the PDF the title reads ‘The Formula’. Both are fitting.

What is more telling is the sub line which reads: ‘Enron, intelligence and the perils of too much information’. Starting with the trials against defendant Jeffrey Skilling following the Enron collapse, he argues that there is a huge difference between ‘puzzles’ and ‘mysteries’.

“The distinction is not trivial,” says Gladwell. “If you consider September 11th to be mainly a puzzle, for instance, then the logical response is to increase the collection
of intelligence, recruit more spies, add to the volume of information we have about Al Qaeda. If you consider September 11th a mystery, though, you’d have to wonder whether adding to the volume of information will only make things worse. You’d want to improve the analysis within the intelligence community; you’d want more thoughtful and sceptical people with the skills to look more closely at what we already know.”

A little further, the difference between a puzzle and a mystery gets clearer: “If things go wrong with a puzzle, identifying the culprit is easy: it’s the person who withheld information. Mysteries, though, are a lot murkier: sometimes the information we’ve been given is inadequate, and sometimes we aren’t very smart about making sense of what we’ve been given, and sometimes the question itself cannot be answered. Puzzles come to satisfying conclusions. Mysteries often don’t.”

All this is important, as it defines how journalists should prepare for a data-driven future. As Gladwell points out: “The principal elements of a puzzle...require the application of energy and persistence...Mysteries demand experience and insight”.

This will be important when a current demand comes into reality: the opening of large data vaults, from statistical offices and governments. Once all this information is available (and presumably it will be) there will be a rapid shift from puzzles (missing information) to mysteries (making sense of too much information).

Most people today are confronted with usually too much information. Whatever product you look up, you can get it in many varieties. But is that after market, no-name replacement battery for a MacBook really as reliable as the one offered by the company? How can it be that the price is 50 per cent less? Often, people make decisions based on simple recipes, often the low price leads them to make a quick purchase. Only later do they find out that the bargain comes at a price.

**Just the facts might be not enough**

Equipped with data and sophisticated tools journalists will be able to tell apart what is true and what is not. But, as recent science findings show, even that is not enough. In an article entitled ‘How facts backfire’ (published in 2010 on Boston.com) scientific findings show that people who strongly believe in a certain position usually do not change their minds when confronted with information telling them that the opposite might be true. Such encounters lead to a ‘cognitive dissonance’ - the picture of the world, as it is seen is out of sync with reality.

The result is that - more often than not - people stick even more strongly to their (wrongly-held) beliefs. So journalists should expect data-driven journalism to take a long time to become an accepted element in the view of the public. Before developing a sustainable model, there will be a build-up process in order to gain trust.²

**Business models for data-driven journalism?**

Journalism is a craft. It’s searching, filtering and transforming a maze of aspects into information. Information then is defined as data that you can act upon. But as any journalists knows, writing, recording or filming can be tricky, even with years of practice. So, with old models how to make money fading away there is a deep anxiety that new ways might not work out or pose unexpected drawbacks.

But some relief should come from the fact that making money by selling information is neither new nor fading away. Just read a number of company stories to learn more. Take, for example, Thomson Reuters. Why are they named Thomson Reuters today? Arguably, because The Thomson Corporation, that started with a single newspaper in Canada and grew into one of the largest media companies in the world, divested early from newspaper holdings. Instead Thomson entered the field of specialized information, starting in the late 70s. Most of the brands the company owned were never of any interest to the public, but of high value to professionals in need of reliable information. This is not so far from what data-driven journalism promises.

---

By 2007 Thomson had merged with Reuters and is now well entrenched in the billion dollar financial information market. By the way, as of 2009, Reuters was not making any sizeable profits with news: the money comes from the numbers.

There are quite a few more examples of stable and successful companies, ranging from Bloomberg (which ‘packaged’ its offerings in the equivalent of a Porsche with its Bloomberg terminals) to eMarketer (successfully selling marketing insights via the web).

In Germany the consumer-oriented ‘Stiftung Warentest’ plans to stop receiving financing from the Government in the future and fund itself via accumulated profits, being fully independent. Their brand and test results are a strong selling point in the competitive German market. Getting a ‘Sehr gut’ (for very good) results in advertisers using this badge everywhere in their advertising.

**Future opportunities for data-driven journalism**

The ideas and arguments presented here might be debated and some assumptions might be wrong. But there is a sense of direction here. Summing up, some future directions media could follow by using data as an integral part are these:

- **Reduce time to search**
  With big databases and search technology journalists can add trust to a service, product or even political programmes. By providing readers/users with clear and understandable choices (not that much more than three in each case) there could be an interesting field for data-driven journalism insights opening up. A field that (for once) cannot as easily be invaded by content farms and scams.

- **Minimize time to reformat**
  Go into any modern office and you will see many people trying to filter information that is on the web or in a PDF into Word or PowerPoint. In many cases the tasks demanded by these programmes are keeping everyone from thinking; instead one spends another 30 minutes trying to reformat the mess of text into something readable. If we think of data-driven journalism as a process, why can’t we produce presentations as the end result of our work - nicely laid out and branded nuggets of information that can be easily understood, displayed on computer screens and even printed out? This is not to say that newspapers should look like business presentations. But the key information that the user actually searches for could be transformed in an automated process. One could even charge larger companies 25,000 Euros or US dollars per year, if they can get this information in their own company style, with logos and the option to add information. Where justifies the high price tag you ask? Well, if the time of just one employee per company is factored in the price could be even higher. Combine this with a strong, trustworthy
brand and it becomes understandable why ‘The Economist’ is presumably making good profit with its large information and consulting business.

- **Enable decisions**
  When we connect the dots from Gladwell’s findings to ‘how facts backfire’ the ability to help people make a clear and easy decision might be another opportunity. Right now, the process of searching and comparing anything (from bicycle tires to insurance) has become an increasingly time-consuming and complex task. Many service platforms offer to ‘compare insurances premiums’ and usually the first step is harvesting the address of the prospective client. Media and journalists could enter this market and simply be what everyone hopes they are: trustworthy.

**About**
Mirko Lorenz is a journalist/information architect based in Cologne, Germany. He holds a Master in History and Economics from the University of Cologne. After working for newspapers he founded an Internet strategy office in 1995, researching or editing for clients like Sony, Handelsblatt and others. Since 2007 he is a member of the innovation projects team at Deutsche Welle. Work assignments include development of future systems in areas such as P2P, semantics and cloud computing.
Session 1: Data Production, Usage and Integration

Before journalists can work with data to find meaning, they first need to learn how to access, structure and filter information. The first session discussed experiences so far and prospects for the near future.

Participants:
- Jonathan Gray, Community Coordinator, The Open Knowledge Foundation
- Lorenz Matzat, freelance journalist, Medienkombinat Berlin
- Richard Rogers, Chair in New Media & Digital Culture, University of Amsterdam
- Simon Rogers, Editor, Guardian Datablog and Datastore
- Tony Hirst, Lecturer in the department of Communication and Systems, The Open University
What is there to learn? (Abstract of the presentation)
Jonathan Gray focused on what will be the foundation of innovative use of data and what this might mean for journalists. His main point was that the eco system of open and linked data is in its early stages. For the future digital technologies have the potential to radically transform the way that knowledge is disseminated in our society.

But we still have a long way to go...
- the shadow of the print press (optimizing information for readers, not networks)
- datasets are to illustrate reports (not as a source of knowledge)
- publishing without reuse in mind (packaging vs. opening)
- culture of asking permission (leading to complexity when using it again)
- vast information silos (that are not interconnected or accessible)
- non-machine readable formats (making it difficult to extract meaning)
- broken links, vanishing content (calling for a better system in the future)

**Where are we going?**

An ecosystem of open data:
- small pieces, loosely joined
- easy to reuse, easy to recombine
- lots of contributors / maintainers
- distributed, decentralized
- divide and conquer
- innovation/unexpected reuse
- iterative, versioned, ‘wiki’-like - learning from open source

**From legal uncertainty to legal clarity**

Open data: free for anyone to reuse or redistribute for any purpose

**What does this mean for journalism?**

**Making the news:**
- finding new stories from datasets – creating a bigger picture by linking datasets - more pairs of eyes to spot patterns - harnessing more external expertise - analyzing the data behind the stories
- responding to interest from public
- putting stories into context
- publishing datasets with stories

**Spreading the news:**
- visually representing information
- demand-driven delivery
- datasets for others to reuse
- enabling users to comment/flag
- integration with other services
- connecting data to stories
What can journalists and media organizations do?
1. Publish data using an open licence
2. Work with existing communities
3. Use and support existing initiatives and technologies
4. Keep innovating!

Additional information:

Definitions
The Open Knowledge Definition (OKD) sets out principles to define ‘openness’ in knowledge – that’s any kind of content or data ‘from sonnets to statistics, genes to geodata’. The definition can be summed up in the statement: “A piece of knowledge is open if you are free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and share-alike”.

The Open Software Service Definition (OSSD) defines ‘openness’ in relation to online (software) services. It can be summed up in the statement: “A service is open if its source code is Free/Open Source Software and non-personal data is open as in the Open Knowledge Definition (OKD)”.

Source: http://www.opendefinition.org/

The Open Knowledge Foundation
Founded in 2004 the not-for-profit organization promotes open knowledge: that’s any kind of information – sonnets to statistics, genes to geodata – that can be freely used, reused, and redistributed.

They organize events like OKCon, run projects like Open Shakespeare, and develop tools like CKAN and KnowledgeForge to help people create, find and share open material. A full list of projects and events can be found on their homepage.

About Jonathan Gray
Jonathan Gray is Community Coordinator at the Open Knowledge Foundation. He studied Philosophy at Corpus Christi College, Cambridge University, Social Sciences
at the Open University and is currently doing research in the German department at Royal Holloway, University of London. He is particularly interested in open government data, data visualization and digital technologies in the humanities. More information can be found at jonathangray.org

Contact:
jonathan.gray@okfn.org
http://twitter.com/jwyg
Lorenz Matzat: Weatherstations - Citizen-Apps, eParticipation and Data journalism

Member of the Open Data Network/Medienkombinat Berlin

What is there to learn?
Lorenz Matzat provided an overview on why and how open data can be used, focusing on developments in Germany. While there are notable projects in the UK and US, many other countries (including Germany) are slower in opening data and making use of it. This is important to evaluate the options for data-driven in journalism in countries around the world.

Key points of the presentation:
What is data journalism?
- access to or production of structured data-sets
- using these data-sets for research of a story
- discuss data/data source in published piece
• provide raw-data for the users along with the story
  • as a spreadsheet
  • and/or interactive research environment (e.g. map-mashup with a timeslider and filters)

**State of Data journalism in Germany**
• no datadesk, datablog or similar in German media right now
• several projects at newspapers and other media (in combination with investigative units)
• Wikileaks War Logs as wake-up call

**Open Government Data in Germany**
• no OpenGov initiative or data.gov
• some government sources for statistical data, often not very detailed and in non-machine readable formats (PDF instead of raw data)
• weak ‘Freedom of Information Act’
• But: Changing consciousness in bureaucracy, some Gov 2.0 applications (though Gov 2.0 is not OpenGov)

**Open Data and Civil Society**
• Open Data in Germany is happening bottom-up (activists, volunteers)
• Several groups and private persons are working on gathering (scraping) data and developing applications
  • Open Data Network
  • Open Knowledge Foundation
  • Government 2.0 Netzwerk
  • Netzdemokraten

**Evolving Ecosystem of Citizen Apps**
• Open Budget (offener-haushalt.de)
• Open Parliament (openbundestag.de)
• Lobbyregister (lobbypedia.de)

**Gov 2.0/eParticipation**
• eConsultation
- Citizen Budgets (Bürgerhaushalte)

**Citizen Apps as Datajournalism sources**
- This ecosystem could be used like a net of weather stations or seismographs:
  - measuring political temperature, pressure, etc.
  - journalists can find stories in data and usage of these applications
  - newspapers/media should sponsor or develop applications like this for reader/user involvement and strengthening of watchdog function

**About Lorenz Matzat**
Lorenz Matzat is a freelance journalist & media educator, living in Berlin. His academic background is as a political scientist. He is a member of the Open Data Network. Additionally he runs a blog about the topic in German language, where he reports on new developments and demonstrates ways to make practical use of data. Blog: http://www.datenjournalist.de
What is there to learn?
Richard Rogers provided an academic perspective and demonstrated just one tool out of about 30 that his department at the University of Amsterdam has developed for various uses. The various apps and tools show clearly that it is possible to provide access and query options for many uses.

Key points of the talk:
Rogers demonstrated just one application, the so called Lippmanian Device, which offers new ways to gather and visualize data. This specific application enables users to query companies or other organizations and filter out word uses. The goal here is to better understand agendas and how original plans are actually followed through, reflected by number of word uses.
One notable point was that the tool developed in Amsterdam uses a ‘hop on/hop off’ approach on Google, using the infrastructure of the search engine to deliver results.

Tools for scraping data and running queries
More (tools can be found on the Homepage.

About Richard Rogers
Richard Rogers holds the Chair in New Media & Digital Culture at the University of Amsterdam. He is also Director of the Govcom.org Foundation (Amsterdam) and the Digital Methods Initiative. Previously, Rogers worked as Senior Advisor to Infodrome, the Dutch Governmental Information Society initiative. He also has worked as a Researcher and Tutor in Computer Related Design at the Royal College of Art (London), as Research Fellow in Design and Media at the Jan van Eyck Academy (Maastricht), and as a Researcher in Technology Assessment at the
Science Center Berlin (WZB) and in Strategic Computing in the Public Sector at Harvard University (JFK School). Rogers is author of Technological Landscapes (Royal College of Art, London, 1999), editor of Preferred Placement: Knowledge Politics on the Web (Jan van Eyck Press, 2000), and author of Information Politics on the Web (MIT Press, 2004/2005), the ‘2005 Best Information Science Book of the Year Award’ presented by the American Society for Information Science and Technology (ASIST).
Simon Rogers: Free our Data!

Editor, The Guardian Data Blog and Data Store

About this talk
The Guardian has taken an early lead on all things data. With the publication of the MP Expense scandal documents in 209 and the Afghanistan War Logs the editors at the Guardian Datablog and Guardian Data Store showed an ability to move quickly and come up with new uses for data in reporting. The Guardian is probably the best website providing users/readers with new insights and opportunities to find out about an issue for themselves, with assistance from the media.

What is there to learn?
Simon Rogers made clear that the use of data has always been a part of journalism. He cited examples from the very first Guardian ever published by Florence Nightingale, who used data gathered in the Crimean War to change sanitary conditions in hospitals. The ‘lady with the lamp’ could also be called ‘the lady with
the data’. She was a gifted mathematician and even invented a format to visualize the data gathered on the causes of death in a clear and convincing form.

Example of polar area diagram by Florence Nightingale (1820–1910).
This "Diagram of the causes of mortality in the army in the East" was published in Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army and sent to Queen Victoria in 1858. This graphic indicates the number of deaths that occurred from preventable diseases (in blue), those that were the results of wounds (in red), and those due to other causes (in black).


**Using data for reporting**
The Guardian has a very clear mission: answering how data can be found, understood, visualized and shared, and it clearly has a lead in this. The main approach seems to be very agile: coming up with quick solutions relevant to the report on a dataset. So instead of having ‘one system’, the Guardian team has come up with different solutions for different cases. An important point here is that the Guardian manages to create such solutions in relatively short time -- weeks instead of months. Additionally the data that used is open for readers to explore and even
download. With this the Guardian aims at making the whole website more useful for readers. Sounds simple, but it is still novel. Who ever is interested in how data-driven journalism can evolve should become a regular reader of the data blog.

**Examples:**

**Afghanistan War Logs**
Special report page, providing information from many angles, based on the leaked documents published by Wikileaks.
[http://www.guardian.co.uk/world/the-war-logs](http://www.guardian.co.uk/world/the-war-logs)

**Investigate your MPs expenses**
An innovative crowdsourcing application allowing users to check 458,832 documents, adding indications whether the documents should be investigated further or not. (Note that there is no barrier or form asking for your e-mail address.)
[http://mps-expenses.guardian.co.uk/](http://mps-expenses.guardian.co.uk/)

**MP Expenses: Who claimed what? The full list**
(Including an open spreadsheet for every MP)

**Links:**
Guardian Data Blog
[http://www.guardian.co.uk/news/datablog](http://www.guardian.co.uk/news/datablog)
Guardian Data Store
[http://www.guardian.co.uk/data-store](http://www.guardian.co.uk/data-store)

**About:**
Simon Rogers edits the Guardian Datablog and Datastore - and is a news editor for the Guardian.
What is there to learn:
Tony Hirst is a gifted teacher who at the same time has invested considerable time into exploring how data can be accessed and used via free and open tools available on the web.

In Amsterdam he showed how by using an array of free tools data can be moved from a spreadsheet inside the Wikipedia into a Google map mashup, e.g. displaying the location of MPs in the UK.

The main message here: no (real) coding involved. Journalists and anybody else interested in this simply have to know what they want to do. The tools to do that are
out there: Yahoo Pipes, Google Docs and Spreadsheets, Many Eyes, Fusion Tables. And they are all free to use.

It would do Tony no justice to display just a shortened version of his talk here, so please check out his presentation (and many others he gave on other occasions) on Slideshare. His OUseful.blog comes up with other hints and tips almost every week.

**Link to Amsterdam Presentation:**
http://www.slideshare.net/psychemedia/datajournroundtable-5046053

**Collection of Tony Hirst’s presentations:**
http://www.slideshare.net/psychemedia

**Blog**
http://blog.ouseful.info/

**About**
Tony Hirst is a lecturer at the Open University in the UK.
Session 2: Data Visualization
What is there to learn:
Stefan Fichtel demonstrated how data can be used to make complex interactions or situations easy to understand. To achieve this goal, said the Berlin-based graphic artist, a process is needed where the data is first understood and then visualized in a way that makes the connections visible at first sight.

He showed a number of examples, most of them highly developed and impressive works of art produced for clients in the media such as the German ‘Handelsblatt’ or corporate clients such as BMW.
Key lesson: data-journalism can profit greatly from applying the know-how of graphic designers, given that the subject is very complicated. Visual design opens yet another opportunity to filter down the information overload to clear messages.

But specialists working on the visualization must understand and work with the data. They have to root their work deeply in data, not only in a journalist’s instructions.

**Examples of work:**
http://www.kircher-burkhardt.com

**About:**
Stefan Fichtel is chief infographics designer at Kircher Burkhardt Consulting, Berlin.
Frank van Ham: How to use Data Visualization

Staff Member, IBM Research

What is there to learn:
Frank van Ham is one of the original members in the IBM team that developed ‘ManyEyes’, a free tool to visualize and share information on the web. The main goal of visualization is to ‘amplify cognition’, make people understand something more quickly. But - as Frank van Ham pointed out - this power can also be used to distort the truth.

Differentiating ‘Data’ and ‘Story’:
There is a clear distinction between ‘Data’ and ‘Story’. Data itself is a form of information. To extract information the user has to go through a process, usually first cleaning up messy formats, structuring and sorting the data into a readable format before it can be visualized. Story is an interactive form of communication, where information is brought into a context that people can understand, remember, discuss and tell others about.
Key lessons:

- Two main reasons for using visualization:
  - Explorative: Understand what is there (e.g. in research)
  - Communicative: Display and discuss visualizations with the public, tell stories based on the data and the facts

ManyEyes

- ManyEyes has evolved into a platform where a wide variety of visualizations can be produced, even with large data sets.
- Users can upload, visualize and publish their projects

Tools for Data Visualization:

Excel
Still the most commonly used desktop tool to work with data

Wordle
producing word clouds from text in a matter of seconds
http://www.wordle.net/

Tableau
Commercial visualization software, simple to use
http://www.tableausoftware.com/

ManyEyes
Fernanda Viégas and Martin Wattenberg, the two visualization creators of ManyEyes, have recently been hired by Google.
http://manyeyes.alphaworks.ibm.com/manyeyes/
Session 3: Storytelling with Data
Cynthia O'Murchu: Datastories

Investigative Reporter, The Financial Times

What is there to learn?
Cynthia O'Murchu talked about her experiences researching and publishing data-based stories at the Financial Times. One key point: the data published by governments (and others) is not at all user friendly. More often than not, journalists have to scrape data from PDFs, although the same information would be available from a database or Excel spreadsheet. “If they call that transparency they need to buy themselves glasses”, was her punchline on this issue.

Examples of data-driven work from the Financial Times:

Oil and gas chief executives: are they worth it? (November 29, 2009)
http://www.ft.com/cms/s/0/190f9e7c-bd8d-11de-9f6a-00144feab49a.html
Currencies in context (October 12/2009)
http://www.ft.com/cms/s/0/9a2fdf12-b725-11de-96f2-00144feab49a.html?ftcamp=rss

Interactive Graphic: The carry trade explained (October 15, 2009)

The pension crisis (May 27, 2009)
http://www.ft.com/cms/s/0/e82a672e-4ab4-11de-87c2-00144feabdc0.html?nclick_check=1

About
Cynthia O’Murchu has worked in the media for over a decade on a range of projects in print, documentary film, radio and interactive. Until recently she was Deputy Interactive Editor at the Financial Times where she researched and produced multimedia features and data visualisations.

In her current position as investigative reporter, also at the FT, she uses her data analysis and computer-assisted reporting skills to produce stories across a variety of beats, both financial and non-financial.
What is there to learn?
The key advice from Alan McLean can be summed up in one important sentence: ‘move developers and editors closer together’.

What seems so obvious and simple is far from the norm. Actually people from newsrooms and developers don’t mix, at least not on any chart. But this one move can make the difference, leading to a better understanding on how to get from data to story faster and better.

Based on examples from The New York Times, Alan showed how the integration of data as a fundamental part of better and deeper stories is “almost to easy” when
considering that many of the constraints that such big stories face in a pure print environment are actually going away: there is an infinite number of stories if the data makes the stories flowing from it fit for publication.

This entry is intentionally kept very short, as the presentation of Alan is available on Slideshare and tells the story better. We have added a link to another presentation that shares some of the main points from Amsterdam, but goes even deeper into how the technology has changed, providing new opportunities.

**Link:**
http://www.slideshare.net/amclean/data-driven-journalism-telling-stories-online?from=ss_embed

Also by Alan McLean is a longer presentation: ‘Hacking the News’. It shares a few slides with the one above but further towards the need for new technology in newsrooms.
http://www.slideshare.net/amclean/hacking-news
What is there to learn?
Google the term ‘data-driven journalism’ and you are likely to get a link to a story from Eric labelled ‘Building the data desk at the Los Angeles Times’. Although he has now left the newspaper, Eric still calls the outfit his ‘alma mater’.

In Amsterdam he shared 10 pieces of advice how journalists can work in their organizations to get data into the workflow. Because such lists are a good way to get started, we repeat them here:
10 pieces of advice for data-journalists in media organizations (and elsewhere):

• **Find the believers** (meaning: convince colleagues and superiors)
• **Get buy-in from above** (a long standing rule for everything that involves IT)
• **Set some priorities** (do not relaunch the CMS, start with a map or specific set of data)
• **Templarize** (develop templates that help you come up with projects quickly, over time you will have a library of usable starting points for new projects)
• **Do breaking news** (this will speed up the process and therefore minimize the scope of the data projects -- a strategy usually considered as best practice)
• **Develop new skills** (such as avoiding the biggest visualization mistakes, learning to work with tools, embeds, APIs, scraping of data, and so on. The steps are easy in themselves, but there are many of them, so reserve some time and be patient about the progress).
• **Cohabitate** (work closely with developers as a journalist and vice versa; this advice is similar to what the NYT says, see Alan McLean’s presentation).
• **Integrate:** Make the produced piece accessible and integrate it well into the existing site, providing links, several views and options for people to understand it.
• **Give back:** Report and tell people in the organization what the benefits of the project were, tell them about lessons learned and new options for the future. This might even include opening up the data sets, as is common at the Guardian.

See a Video of Eric's talk in Amsterdam on Vimeo: [http://vimeo.com/14715185](http://vimeo.com/14715185)

The longer version of this advice can be found at Online Journalism Review, [http://www.ojr.org/ojr/people/eulken/200811/1581/](http://www.ojr.org/ojr/people/eulken/200811/1581/)

**About Eric Ulken**
He describes himself as an ‘online journalist and an untrained but enthusiastic information designer/tinkerer’. Eric Ulken left his job as editor for interactive technology at the Los Angeles Times in November 2008 to travel and report on trends and best practices in online journalism. He is a 2005 graduate of the communication management M.A. program at USC's Annenberg School for
Communication, where he was an editor and producer for OJR and Japan Media Review. He has been a web monkey at newsrooms in six states, including his native Louisiana. He has spent much of this time since 2009 travelling and working in different places, speaking at conferences and teaching as a visiting professor at the University of British Columbia.

Find out more at his website http://ulken.com

Sources: http://www.ojr.org/ojr/people/eulken/
What is there to learn?
OWNI is a data-journalism agency. This Parisian start-up aims to deliver services to news outlets based on data. The team comes from a background of media and programming, enabling them to come up with app-style data presentation in a matter of days, not weeks. Plus, in line with other speakers of the day the need to connect data with story was stressed.
Nicolas Kayser-Bril is head of data of the start-up. He drew an interesting comparison between complex journalism stories and ‘World of Warcraft’: Both are only successful if they relate to human interests and the senses. And both share the need and a good understanding for plots and characters. In his talk in Amsterdam he described in part what OWNI does and what are the goals of the project.

This approach has been put to the test recently with the Afghanistan War logs. Having a number of big media companies ahead of them, the team at OWNI made a quick plan (‘look at the documents from a French angle’), gathered volunteers and came up with an application very quickly.

This approach is clever and will hopefully find growing demand: instead of months of development and complex IT-integration, the app-style development philosophy is very close to how the Guardian solves the problem of suddenly having to deal with a new data set, a new situation and the goal to find an elegant, quick and useful view for that particular information.

About Nicolas Kayser-Bril

Nicolas is 24 years old and Head of Data at OWNI. In an interview with the Online Journalism Blog, he described his goals for this new idea as follows: “We want to enhance information with the power of computers and the web. Through software, databases, visualizations, social apps, games, whatever, we want to experiment with news in ways traditional and online media haven’t done yet.”

Nicolas already has an interesting track record: during school he and others produced a story about violence in French schools, partly by counting the number of security cameras installed. He spoke about ‘The World according to newspapers’, the visualizations from which have been a favourite of many visualization blogs in 2008, when the study was published. Together with Giles Bruno he created an interactive map, in which the relative size of the world’s regions change according to the number of articles published about that region in a Western Country.

Owni is seen as an experiment and an innovative approach to journalism. It is financed through 22mars, an agency based in Paris.

http://onlinejournalismblog.com/2008/03/23/the-world-according-to-newspapers/

Interview with Nicolas Kayser-Bril on ‘Online Journalism Blog’,
http://onlinejournalismblog.com/2010/03/19/interview-nicolas-kayser-bril-head-of-datajournalism-at-owni-fr/
What is there to learn?
Gavin Sheridan advocates transparency. Many of us do that, but few invest as much time and effort as him. Together with a colleague he has created the site ‘TheStory.ie’ where he publishes data-sets after getting them from institutions. Often this data is only handed out after considerable investments, running to hundreds of euros for each case. To finance this, the two founders rely on donations from readers.

As for the data: the sets are usually published right away, only then will the examination start. What started out as an experiment remains something like a hobby. But the approach has clearly caused a stir. That can be told from quite a number of positive remarks from other journalists and academics. As Gavin describes on the site: “The Story is dedicated to sharing documents, combing and
combining data and promoting transparency in public life: an experiment in journalism and crowdsourcing hoping to shed light on the government. If you’re spending the Irish taxpayers’ money, you’re on the radar.”

This approach - getting things moving in reality, not planning for ages, is impressive. There are many, many data sets collected with tax payers money around the world, some are published, many are not. And even the data that you can get, is often presented in a way that makes it extremely complicated to get a clue what is means.

The Story.ie is a new approach: first crash the gates, then start to ask questions. The answers will be found over time, don’t you think?

Watch a video of this talk on Vimeo: http://vimeo.com/14716295

About Gavin Sheridan
Gavin Sheridan is a blogger and journalist from Cork. He established one of Ireland’s longest running blogs, Gavinsblog.com and helped set-up Kildarestreet.com. Kildarestreet is designed to allow citizens keep tabs on what their TDs and Senators are up to in the Oireachtas. You might see his name pop up in the Irish Examiner the odd time too. (gavinsblog AT gmail DOT com). He works at a new media start-up founded by former broadcaster Mark Little. (Source: The Story.ie)
Session 4: New Formats for Presenting Information & Stories

Speakers:
- Stijn Debrouwere, freelance journalist and programmer
- Burt Herman, Founder, Hacks/Hackers
- Andrew Lyons, Commercial Director, Ultra Knowledge
- Julian Burgess, Editorial Developer, The Times
What is there to learn?
Stijn is very critical of publishing systems in digital media, specifically how they process data in published articles. Instead of creating searchable information, today’s content management systems are usually ripping apart what has already been researched by reporters.

In order to create context - which should be the ultimate goal of electronic publication - much is vanishing and good data is lost. To change this the ‘parts’ of any information should be better structured and searchable via good accompanying metadata.
His key message to larger media organizations is to change the strategy: instead of looking for the ‘cherry on top’ (using digital publication as another way to create income through advertising, the focus should shift to ‘baking a better cake’.

This problem - Content Management Systems that are not designed to structure data - is usually beyond the control of a journalist. Still, without switching to better systems, publishers are more or less powerless and in need of help from others, such as Google and even Facebook to structure, filter and distribute stories in an efficient way.

To end with a quote from Stijn:
“**What we, as an industry, need to learn, is that digital strategy is not about sprinkles, but about baking a better cake. If we don’t see the opportunities, somebody else will.**”

Stijn was invited to the conference based on his excellent text titles ‘We are in the information business’ (see links below). He noted the main points of his talk in Amsterdam in another text published on his website.

**Sprinkles on top - live from DDS (August 24)**

**We are in the information business (April 2010)**

The article is part of a series covering ‘Information architecture for news websites’. The other installations can be found here:

**About**
Stijn Debrouwere is an information architect mainly working in media projects, based in Belgium.
Burt Herman: Storify - making sense of the world

Founder, Hacks/Hackers

What is there to learn:
Burt Herman, former bureau chief for Associated Press in various locations around the world is an example of a journalist moving deeper into programming and entrepreneurship.

One way of doing this was by founding Hacks/Hackers, a community that explores how to use technology and know-how to ‘make sense of the world’.
Another new project is Storify, where Tweets from Twitter can be sorted and structured into easy to understand patterns of information, creating an overview for others.

**Links:**

Hacks/Hackers


This site aims to bring together journalists and technologists. The main idea is to bring the journalism and technology communities together at casual face-to-face gatherings to trade ideas and find potential collaborators. The founders of the site and network are Aron Pilhofer of the New York Times and Rich Gordon from Northwestern University’s Medill School of journalism and Burt Herman, Knight journalism Fellow at Stanford University.

**Storify (in beta)**


The site aims to develop simple tools to curate the social web.

**About**

Burt Herman graduated from Stanford University and started his journalism career in 1996, as a reporter for the Associated Press. Other positions for AP included being the editor of the International Desk in New York, followed by postings to Berlin, Moscow, and Uzbekistan. He has reported on events such as the Afghanistan War and the Beslan school massacre, Russia, the U.S. invasion of Iraq and the Republic of Georgia’s ‘rose revolution’. From 2004 he was bureau chief in Korea. He is a John S. Knight Fellow.
What is there to learn:
As Julian Burgess pointed out in his talk, there are a number of new ways to monitor and react to website statistics and real-time analysis of issues, e.g. by accessing Twitter stats.

The way this is done is usually through APIs (Application Programming Interfaces), allowing the user to extract and look at data from big platforms such as Twitter.

In many newsrooms, however, the use of analytics is not a part of the reporting process, as blogs are usually analyzed elsewhere in the organization. But the trend in analytics is accelerating towards real time analytics (instead of being sent a chart once a month or once a week).
Closing the gap on this is another avenue to be explored in the future. Journalists should be aware when a story breaks, how readers/users react to stories published and should be enabled to act on such information.

The possible uses of such data range from being a tool in reporting to optimizing the pages of a news website, to make information easy to find.

**About**

Julian is a developer who became the first programmer to be hired specifically by The Times’ editorial team. He works with Jonathan Richards on data and visualisation projects.

Below is additional material on this particular issue that has been discussed widely in the journalism space:

Online Metrics Report (September 2010) - The Journalism School, Columbia University
Future perspectives for DDJ

What are next steps of data-driven journalism?
The opening of data, new interactive technologies and the pressure to find new roles for journalism will drive this area forwards. In the meantime, how can we structure a comprehensive training course: one that would enable journalists to use data for reporting?

What came out of the roundtable was that there will be a wide range of activities to be understood: from data scraping to programming new platforms to handle content better, from filtering and visualizing to storytelling with HD cameras.

Will every journalist need to learn all this? Clearly: no.

For the future we need journalists who are sympathetic to data and welcome every developer or self-taught data-digger. Future reporters should be specialists in one of the areas of application and have a good understanding of the problems and opportunities of other crafts. But in general it will be better to have ‘real’, trained programmers (who may also be journalists) and ‘real’, trained photographers, cutters, producers, etc. If data-driven journalism will grow beyond today’s beginnings, specialization will become more important.

What is there to learn? You may be tired of this much repeated question. But this is so important to the future success of data-driven journalism that we ask it yet again. For a start: we need really good examples of data-driven journalism, big stories like the MP expenses scandal or the Afghanistan War Logs. We need them more often, in evolving quality - because this will open up a market for data-driven journalists.

Programmes and training sessions should be structured into one-day idea/experience exchanges, followed by one-week intensive training focusing on (a) programming for news and data, (3) data filtering and data analysis, (4) data visualization and (5) storytelling.
The EJC plans to focus in this direction and will publicize courses on http://ejc.net/

We are in the trust business

Next steps for data-driven journalism
Mirko Lorenz, DDJ project lead/Deutsche Welle Innovation Projects

The data-driven journalism roundtable on August 24 in Amsterdam was a very positive experience. The speakers focused on their experiences, showed examples of their work and how they did it. There was a palpable interest in this evolving field and a lot of enthusiasm about its future prospects.

Can you contribute to the story?
Another clear message is that no journalist should feel excluded from this field. To clarify: not all journalists working in tomorrow’s newsrooms will be coders. There will be real coders, journalist/coders and journalists working with the output of such teams, being specialized in writing, photography and filmmaking. This is not so different from today’s newsrooms where you will find different talents too: some journalists are excellent in finding stories, some are masters on the phone, others maybe shy but great writers. And so on. The change in what knowledge skills are needed should not trick anyone into believing that only journalists with professional coding experience are needed for this. But we will need people with skills, talent and an interest in solving complex issues. The main question is: can you contribute to the story?

Questions we need to answer
On top of that there are quite a few questions that were not answered, some were not even asked or covered just for a few minutes. To give you an example: how can we make money with this? Another question: What are the areas of application? Investigative scoops or consumer service? Can data enhance reporting on business, economics and politics? How will all that work out?
One perspective: Tapping into the market for trust
This brief text will not manage to answer all these questions. But there is one important point: what market are we actually in when talking about data-driven journalism? My answer: we are in the trust business.

What is the scarce resource in the future of media?
In economics one basic principle is that only scarce resources have a price. Otherwise they are free goods, for everyone to use. Journalism as a business was built on the scarce resource of being able to print a hundred thousand or even a million copies of a newspaper and distribute it in the morning. There was content and an interest in news that drove demand by readers. And this provided the platform for advertising. That model, which was partly based on possessing a technology to produce and distribute using printing machines, cables or satellites is broken beyond repair. The reader’s perspective might look like this: sure I love to read a glossy magazine with nice photos, but why should I? All the content in there has been presented on my iPad yesterday.

The one resource that is scarce today is an intangible one: trust. The Internet is a dangerous place and so is the physical world. Just one example: how many billions in euros, dollars, etc. are wrenched from consumers around the world by banks not telling them about hidden fees and kick-backs for sales of mortgages, mutual funds? How much money is lost because we as humans are not very good at calculating compound interest (the interest that is adding up on a debt over time and that defines the final sum you actually pay)? Differences of half a per cent on a mortgage translate into thousands of euros or dollars, but most people don't know or care and as a result this ignorance are exploited day by day, year by year.

If we further develop applications like the ‘rent or buy calculator’, make them more versatile, clear and easy to use, my guess is that data-driven journalism can turn into a new revenue stream. As journalists have a strong motivation to be trustworthy, we go to lengths to make a story right and check the facts.
Let’s extend this thought of helping people to make decisions with data. Nobody wants to pay for online news. But will the picture change if we offer tools and databases to help people make clear and informed decisions? How to finance a house and keep it? How to find the best deal for your next car? Yes, it will be difficult to make people pay for such calculators. But how about offering to print out their personal finance plan towards their dream house as a full colour A3 poster, so that it can be posted on the wall? How about becoming the first place to visit?

There is still a lot of work to be done: we need to reduce the time needed to find an answer. We should create destination points where Google can (and will) send people to get unbiased, trustworthy advice. We could enhance the experience that people are usually not paying for information (this feels like paying for talking to someone), but that they do pay for ‘souvenirs’ as can be demonstrated on the app markets, in the games business, and in other markets.

Future formats represent another avenue for data-driven journalism. Did you see the little experiment on re-designing boarding passes? It is just one example of what can be done. To provide a clear model: research and information collections are today delivered in page formats, often as bulky PDFs. So go into any office and you will see droves of people spending their day ‘re-formatting’ information that is already there but does not look good: they copy sentences and pictures from PDFs and transfer them into text processors and presentations charts. Day by day, hour by hour.

What if future news organizations would better understand that the single items they are producing as articles could actually be transferred into meaningful analysis and would ideally be delivered in PowerPoint for large corporate clients? This might create a specialized information market worth thousands of subscription fees, year by year. There are examples of this, so this thought is not that new. Have a look at companies like eMarketer, The Real Story (reporting on CMS/IT systems), Statista and others. If you don't believe in anything proposed here, please read about the transformation of The Thomson Corporation from one Canadian newspaper to a media empire and then into a powerhouse for specialized information. It has been done before!
The last word on this, is a quote from Adam Westbrook, which puts a good perspective on what to expect and how to find new and better ideas to make this happen:

**Future directions: Towards good training for data journalists**

“It might not feel like it, but the future of journalism is being fixed right now. It is being figured out babystep-by-babystep, one small development at a time. Each new idea and business brings something new to the table. Each failed business model is a lesson learned,” Adam Westbrook, author of ‘Next Generation Journalist’ - [http://www.nextgenerationjournalist.com/](http://www.nextgenerationjournalist.com/)

We are looking for people to advance or challenge these ideas. Can you contribute to that story?
How to start working with data: A brief checklist for data-driven journalism

For anyone interested in entering this emerging field of journalism, here are five easy steps that can help you get to grips with data, visualizing and telling stories based on data.

1. **Watch these two videos: Hans Rosling (2006) and David McCandless (2010)**
   - [http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve Ever_ever_seen.html](http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve Ever_ever_seen.html)

2. **Look at some examples**
   Good places to start are ‘Where does my money go’, ‘Information is beautiful’, and The Guardian Data Blog. These show how data relates to storytelling and how journalists present their stories.

3. **Familiarize yourself with the concepts of Open Data**
   More and more data vaults are opening up to be used by the public and journalists. The Open Knowledge Foundation in the UK provides news and platforms such as CKAN. Additionally you should reserve some time to at least look at Data.gov (US), Data.gov.uk (UK) and the Data sections at the World Bank. By then you will have a good understanding of where we are heading.

4. **Find open data to work with**
   One place to start e.g. is Google Fusion tables where you find growing number of available datasets. You need to register with Google to get access.

5. **Experiment with visualization: Build a treemap and a bubble chart**
   Visualization tools on the web provide some new and innovative ways to visualize complex datasets nicely. Two popular forms are treemaps (e.g. to show how a budget is distributed) and moving bubble charts (to show how development of certain data relates over time). You have seen examples of this in the videos of Hans Rosling and David McCandless. Here is one place to start doing this: ManyEyes, where you will find examples and some tutorials to do this.
Flowing Data: An Easy way to make a treemap,  
http://flowingdata.com/2010/02/11/an-easy-way-to-make-a-treemap/

10 Lessons in Treemap Design:  
http://smartdatacollective.com/Home/16512

**ManyEyes: Bubble Charts and Tree Maps on ManyEyes**, Blog entry by Frank van Ham,  
http://manyeyes.alphaworks.ibm.com/blog/2007/05/25/more-flexible-visualizations/
The #ddj lists 2010

A collection of resources for data-driven journalism: Examples, people, tools, articles and books

A curated collection of links to articles, examples for data-driven journalism, tools, people and companies. If you think something is missing here, please send an e-mail to: mirko.lorenz@gmail.com

Websites

Data
Data.gov (US), http://www.data.gov/
Data.gov.uk (UK) http://data.gov.uk/
Worldbank, Data http://data.worldbank.org/
Scraperwiki, http://scraperwiki.com/
Open Knowledge Foundation, http://okfn.org/

Visualization
Flowing Data, http://flowingdata.com/
Well-formed Data, http://well-formed-data.net/
Information Aesthetics, http://infosthetics.com/
Good Magazine, http://www.good.is/
University of Amsterdam, http://www.digitalmethods.net/
Simple Complexity, http://simplecomplexity.net/
A Beautiful WWW, http://abeautifulwww.com/
Infografistas, http://infografistas.blogspot.com/
Datenjournalist. German blog on ddj. http://www.datenjournalist.de/

Data Tools
Google Docs and Spreadsheets (find on the web)
Google Fusion, - with access to many open data sets,  
http://tables.googlelabs.com/Home

Google Code Playground: Helps to explore Google data and tools.  
http://code.google.com/apis/ajax/playground/


API Playground - helping journalists to understand API Data.  
http://apiplayground.org/

Data Converter (Shan Carter): Simple Tool, transfers CSV to web-friendly formats, includes JSAON and XML.,  


Open Knowledge Foundation, CKAN, http://www.ckan.net/

Python/Django http://www.djangoproject.com/


Training & Tutorials
OUseful.info, http://blog.ouseful.info/

Multimedia/Storytelling


Multimediashooter: Collects videos, often points to interesting new productions.  
http://www.multimediashooter.com/

MediaStorm. The place to see some prime examples of storytelling. Starting tips: Sandwich Generation, The Marlboro Marine, Take Care.  
http://www.mediastorm.com/

Advancing the story, http://www.advancingthestory.com/

Finding the Frame: Interesting site where you can hand in multimedia pieces for review, http://www.findingtheframe.com/

Spill the Beans, German platform, presenting examples of multimedia storytelling.  
http://www.spillthebeans.de/

Data-driven journalism Articles: Online Metrics Report (September 2010) - The Journalism School, Columbia University

Data People (2010)
Our intention was to create a short, conceivable list. It is slightly unfair to single out just a few people, as most of them work in teams. This is not meant to exclude anybody, but driven by the intention to filter down to a reduced list of innovative people working on new ideas for data-driven journalism, better information filtering and storytelling. If you think somebody should clearly be on this list, again send us an e-mail.

Simon Rogers, Editor, The Guardian Data Blog and Datastore, http://www.guardian.co.uk/profile/simonrogers
Tony Hirst, http://blog.ouseful.info/
Michael Driscoll, Dataspora, http://dataspora.com/blog/
Data Examples
A list of resources on data, journalism and multimedia storytelling where you can find examples of new work, new insights and connect to other journalists. There are many, many more, but this should provide a quick entry into what data-driven journalism is all about.

Florence Nightingale: Polar chart. Perfect example how data, visualization can uncover the truth and make change happen.
http://understandinguncertainty.org/coxcombs

Video: David McCandless, The beauty of data visualization, TedGlobal, 2010,

Video: Hans Rosling, Gapminder, video on Ted.com,
http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html

The Guardian: MP Expenses. Elegant, smart and a prime example of crowdsourcing, Opening up formerly hidden information for everyone to see,
http://mps-expenses.guardian.co.uk/

The Guardian: Afghanistan War Logs. Providing a better overview through combination of different media,
http://www.guardian.co.uk/world/the-war-logs

Washington Post, Faces of the Fallen, an original project by Adrian Holovaty,
http://projects.washingtonpost.com/fallen/

New York Times: Run Well, Training Plan creator,
http://health.nytimes.com/run-well/start

New York Times, Rent or buy calculator,
The Billion Dollar Gram, Information is beautiful - billions of costs in a treemap, showing relations between issues and spending, 
http://www.informationisbeautiful.net/visualizations/the-billion-dollar-gram/

BBC: Super Power - Visualising the Internet - based on Top 100 websites, treemap with several layers, http://news.bbc.co.uk/2/hi/8562801.stm


New York Times: Rent or buy calculator. Interactive service, helping people to make a decision on whether it is better to rent or to buy. 

Where does my money go? Great visualization of the UK budget. They got calls from other government departments who did not have that kind of overview before and therefore wanted a poster. http://www.wheredoesmymoneygo.org/dashboard/

David McCandless: Information is Beautiful. Simply great, innovative work, 
http://www.informationisbeautiful.net/


Gapminder Website: Changing the view on complex issues with a very smart platform. Be sure to watch Hans Rosling’s presentations on Ted.com, if you haven't so far, http://www.gapminder.org/


BBC Superpower: Visualizing the Internet
http://news.bbc.co.uk/2/hi/technology/8562801.stm

EveryBlock: A newsfeed for your block, developed by Adrian Holovaty,
http://www.everyblock.com/
Data Articles
A short list of must-read articles on data-driven journalism. The goal here is to
guide newcomers to some insightful pieces, without trying to be complete. Feel free
to send us recommendations for additional input, best by posting your comments
on the data-driven journalism group at the EJC website.

Adrian Holovaty: A fundamental way newspaper sites need to change,

Stijn, Debrouwere, We’re in the information business, April 2010, stdout.be,

Eric Ulken: Building the data-desk: lessons from the L.A. Times, The Online
Journalism Review, Nov. 21, 2008,
http://www.ojr.org/ojr/people/eulken/200811/1581/

Journalism Needs Data in the 21st Century, ReadWriteWeb, Aug 5, 2009,

Rich Gordon, What Will Journalist-Programmers Do?, MediaShift Idealab, Nov. 18,

Rich Gordon: Data as journalism, journalism as data, Readership Institute, Nov. 14,
Data Books

Nick Davis, ‘Flat Earth News’, 2008. Criticism of ‘churnalism’ in newspapers, some very revealing back-stories how the truth is sometimes distorted by journalists. The most revealing story might be the crusade against heroin, which may have unintentionally helped to create a drug market. If the journalists had known the data, would they have argued differently?

Robert McKee, ‘Story’. An indispensable handbook for screen and storywriters. If multimedia journalism is to grow we need more knowledge on how to do it right.

Dan Roam, ‘The back of the Napkin’. Helps you to develop new skills in visualizing (by hand), but can be a good creative source for the future.

Ian Ayres, ‘Super Crunchers, Why thinking by numbers is the new way to be smart’, 2007 - sums up and describes techniques for data mining and introduces examples of how big data can help to make predictions for the future.

Malcolm Gladwell, ‘Outliers. The Story of Success’. Essentially most of these stories are data-driven stories, the findings are sometimes amazing and fun to read. E.g. the 10,000 hour rule that may help to understand the successful careers of programmers and musicians such as the Beatles.
Data Companies
This is not meant as promotion. Instead the idea is to guide the way to outfits and start-ups that work with data, will be able to help you if you are looking for someone to help with a project or for a job in this field.

Stamen (US), http://stamen.com/
Economy.com (US), belongs to Moody's, http://www.economy.com/default.asp
Tableau (US), http://www.tableausoftware.com/
Kircher Burkhardt (Germany), http://www.kircher-burkhardt.com/
Stiftung Warentest (Germany), http://www.test.de
Statista (Germany), http://de.statista.com/
DPA Infocom (Germany), http://www.dpa-info.com/
Golden Section Graphics (Germany), http://www.golden-section-graphics.com/
ThomsonReuters (US), http://thomsonreuters.com/
Bloomberg (US), http://www.bloomberg.com/
Economist Intelligence Unit (UK), http://www.eiu.com
How to connect and stay in touch

There is clearly something happening here, so make sure that you stay in the loop. There are several ways to do this.

**EJC data-driven journalism group**
To share your suggestions, insights or opposing views on anything contained in this paper, please join the data-driven journalism group on the EJC online community http://community.ejc.net/.

**Mailing List**
Additionally you can subscribe to the data-driven journalism mailing list, which is run by the Open Data Network (English): [http://wiki.opendata-network.org/DDJ-Mailinglist](http://wiki.opendata-network.org/DDJ-Mailinglist). The EJC will also use this list for future announcements.

**Twitter**  #ddj
Acknowledgements

The first roundtable on data-driven journalism was made possible by the European Journalism Centre (EJC), a non-profit organization made up of ‘journalists working for journalists’.

Anna Lena Schiller, a graphic artist based in Berlin, did what should be one goal of data-driven journalism in the future: boiling down each presentation into a single picture, and so filtering out the really relevant from an overload of information.

Her hand drawings from the event are part of each chapter on the following pages. The visuals are available for download via Flickr: http://wiki.opendata-network.org/Data_Driven_Journalism

Contributors
Mirko Lorenz, DDJ project lead for EJC/Information Architect Deutsche Welle
Texts, excerpts from talks and links

Edited by Howard Hudson, EJC

Liliana Bounegru
Organization of roundtable for EJC

Wilfried Rütten
Director EJC

Imprint
The material collected is published by the European Journalism Centre. You are free to distribute and share this material, but kindly asked to give credit when used elsewhere.