

WCSJ  
Helsinki  
2013

8TH WORLD CONFERENCE OF SCIENCE JOURNALISTS HELSINKI, FINLAND June 24-28, 2013



Panel discussion and workshop 26.06.2013,  
14:00 – 15:30

## Data explored – a reader for science journalists

Co-produced by Dino Trescher, Lynne Smit, Anina Mumm, John Burn Murdoch, Duncan Campbell

Contact:

Dino Trescher, Constart Correspondent Network [trescher@constart.com](mailto:trescher@constart.com)

# Data explored – a reader for science journalists

Co-produced by Dino Trescher, Lynne Smit, Anina Mumm, John Burn- Murdoch

## Content

Preface .....	3
A recent data driven story.....	3
The digital turn and the relevance of data for journalism .....	3
The trend towards data journalism.....	5
Introductory reads and vodcasts .....	6
Data visualization.....	6
Good practice and selected case studies .....	6
Inspiring cases of data driven journalism.....	7
Contribution by Duncan Campbell, British freelance investigative journalist .....	9
Contribution by Anina Mumm, South African scientist and journalist, for Oxpeckers .....	10
Contribution by John Burne-Murdoch from the Guardian newspaper in the UK.....	11
An additional annotated list of selected tools and sources .....	12
Networks, organizations, links, notes, annotations .....	12

## Preface

According to World Wide Web inventor, Tim Berners-Lee, data driven journalism will play an increasingly [important role in journalism in the future](#). So what is data driven journalism and why is it considered the driving force in changing journalism? For Jonathan Stray, journalist and a computer scientist it means, “obtaining, reporting on, curating and publishing data in the public interest.” Paul Bradshaw, data journalism trainer and writer, stresses that it is “the convergence of a number of fields [... - from investigative research and statistics to design and programming.”

Data journalism has a fundamental role to play, because unique stories are often hidden in mountains of data, documents and statistics and are distributed all over the internet. So the question that many journalists need to answer is: **How to gain and explore data and use it in a skilful way for reporting?** Journalists in many countries are pioneering, utilising new digital tools, exploring data for reporting - thereby adding unique value to their work, and discovering facts and interrelations based on their data. In spite of this, the role of data-driven journalism in science journalism is minimal.

This reader is a collaborative document, based on numerous sources<sup>1</sup> prepared before and supplemented with the main aspects of the WCSJ 2013 panel discussion and interactive workshop: [Data explored – the code that underpins the future of journalism](#). It provides food for thought and case studies, stories, skills and tools for data driven science journalists. The goal of the workshop and the reader is to introduce and inspire science journalists to explore the sphere of data journalism. So it seems reasonable to consider data a key element and a kind of “need to know set of skills for journalists”. One example which shows what kind of stories can evolve from a data driven approach made the headlines in New Zealand.

### A recent data driven story...

Keith Ng, a freelance journalist from Wellington, New Zealand, extracted data from a self-service counter in a social security office and thereby accessed an enormous amount of government data. The files from New Zealand’s Ministry of Social Development included names and addresses of children living in protected care, investigators and clients in fraud investigations, invoices, and medical prescriptions — the beginning of what became the [government’s largest ever security breach](#). But Keith did not sell the story to a newspaper or magazine, he [blogged about it](#) on a website called [Public Address](#), where he maintains a blog voluntarily. So instead of being paid for a few lines, he put a link at the bottom of the story that led to a [donation page](#) on [GiveALittle.co.nz](#). “This story took most of the week to do,” he wrote, “so if you like it, some money would be greatly appreciated.” A decision which made Keith Ng, according to a national newspaper “the best paid journalist in New Zealand.”

### The digital turn and the relevance of data for journalism

The situation is well known: the internet has put journalism and journalists in many ways under pressure. Journalism has lost its traditional central role as gatekeeper and information broker. The selection, accentuation and distribution control, journalism has lost in part. Furthermore, since the internet offers all players numerous opportunities to exchange and network without journalistic plat-

forms, journalism has lost major aspects of its self-legitimation, having privileged access to events, primary sources or persons. This has led journalists to reflect on their ability to process information, to evaluate and contextualise. Where traditional journalism fails in the hopeless race for the very latest news and exclusive news, resources become available that can be allow for a partial repositioning as a "guide in the knowledge society". In this way, journalists become professionals who are capable of penetrating knowledge and experts, delving deep into the truth of matters, finding out deep layers of knowledge, while at the same time preparing it for a large audience - essential services for a democracy that the internet itself cannot provide. In the USA, the UK and other countries this journalistic pilot function in particular is discussed and implemented under the heading of "data-driven journalism". It is obvious that the 'pilot function' can be professionally completed only when the journalist is able to develop relevant and large knowledge resources and weight them to identify suitable contacts in the world of experts and, finally, to present the results of this research so that it is easily accessible to a broad population.

Specifically for science journalists this realignment of journalism poses huge opportunities for shaping its profile, because the core of good science journalism is excellent data-driven journalism. Science journalists are in a privileged position due to their intimate knowledge of how science operates. They understand numbers and statistics and their working knowledge of the appropriate interpretation and criticism of data enables them to play an important in role this essential development in journalism.

It seems reasonable that this trend in journalism is of great importance in repositioning of the profession. We are seeing it evolving in many countries simultaneously and being considered cautiously in others. But the field is evolving and as in every professional field, there are [conferences](#) bringing together journalists, developers, statisticians, and designers interested in data journalism. Examples are [data harvest](#) in Brussels, [hackathons](#) in Montreal and the [NICAR conference](#) and various other occasions, where scholars in data driven journalism exchange and collaborate. Last but not least there journalists with quite a reputation and expertise in the field such as Stephen Grey, James Randerson, Andrew Revkin, Nils Mulvad, Paul Bradshaw and many others. A continuous flow of information is available via: [mailing list of EJC/OKF open data and data driven journalism](#), where a [free data journalism course is available](#).

But journalism as a profession is probably not changing and adapting as quickly as the previous paragraphs imply. One reason for the seeming reluctance in implementing data journalism is probably that the editors might be lacking the skills to screen databases, to process large amounts of data and to make it appealing and journalisticly relevant. Data driven journalism is not equivalent to the placement of research techniques and computer assisted reporting skills. Rather, the examples in the Anglo-American world show, that the success strongly depends on the data driven narration. This requires the adequate use of new media, for example the visualisation by more or less interactive graphic charts, innovative narratives and interactive websites<sup>2</sup>.

The interplay between knowledge-based journalistic research and new ways of the craft of science journalism can provide a previously unsuspected key role for science journalists. This new expertise can be an important key qualification within the development of journalism. So some argue science

journalism must claim this role and be part of the development if it wants to continue to be relevant in the knowledge society and to fulfil its essential functions.

The datablog editor of the UK Guardian, Simon Rogers, describes this paradigm shift in journalism itself: "Comment is free, but facts are sacred. And that's what *The Guardian's* about. I really did think that, you know, this says a lot about the web. Interestingly, I think that's how the web is changing, in the sense did a few years ago it was just about comment. People wanted to say what they thought. Now I think it's, increasingly, people want to find out what the facts are".<sup>3</sup> In the treatment of the facts and their knowledge contextualisation, Rogers' statement adds up to the central function of the quality of journalism in the future.

### **The trend towards data journalism**

Many journalists still think that statistics are dry, but in fact they can be exciting material and provide facts for stories and interactive graphics. In general, data have always played a great role in media: from stock prices and soccer results in newspapers to traffic and weather reports on the average morning radio show, statistics have been key. What is new is that more and more statistics are now freely available, through governments, large organisations like the World Bank or the investigative portal Wikileaks. The computer and smart software make it easier to rummage through mountains of data and display the results graphically. The aggregation of data, the links, and in particular the visual way in which messages are collected, presented and consumed, increasingly change. Leading international media such as the New York Times and The Guardian are pioneers in data journalism (or "data driven journalism", the term "computer-assisted reporting" (CAR) is increasingly falling out of favour).. An example of how data can be used was evident in July 2010 when Wikileaks published about 70 000 documents on the Afghanistan war. With the publication of the Afghanistan documents, by TIME, The New York Times, Guardian and German weekly "Der Spiegel" the future of journalism was revealed. The political consequences are far reaching. Networks between major publications are the future - whether in the form of cooperation between newspapers across borders, whether in the form of multinational search pools or working with foundations and universities."<sup>4</sup>

The Guardian revealed an interactive map on its website<sup>5</sup> that impressively visualised the attacks on the ISAF troops.: "It's not just about revelations, but it makes overall scenarios transparent," says Lorenz Matzat, from the German portal datenjournalist.de. Another example: In August 2010, the Washington Post in an open data file named "Top Secret America"<sup>6</sup>, studied the American military industrial complex. It shows how this has ballooned in the past decade as part of the anti-terrorist struggle to such dimensions, that efficiency of control can no longer be guaranteed.

### **The role of the reader in data journalism**

The principle of open data is not automatically compatible with the classic desire for news exclusivity. Journalists have to lose their traditional assumption that they have the right to know more than "ordinary" citizens. This can, however, provide a competitive advantage in an increasingly competitive market. Martin Belam, "information architect" at the Guardian, has a different journalistic perspective: In the media industry, one must begin to understand start that not only the own content is essential, but also external links. "Journalists are increasingly becoming curators of data and infor-

mation". This change in values is evident in The Guardian's data blog. The newspaper regularly publishes the data sources behind their stories and asks the reader to join in. For more case studies see the Guardian's datablog [www.guardian.co.uk/news/datablog](http://www.guardian.co.uk/news/datablog).

### **Introductory reads and vodcasts**

A good read and introduction is the [Datajournalism handbook](#) edited by Jonathan Gray, Liliana Bounegru and Lucy Chambers, an [Iberoamericano version](#) is currently being written collaboratively by 64 people. A short essay "[A fundamental way newspaper sites need to change](#)", by Adrian Holovaty, founder of EveryBlock, gives a brief view of how the digital turn is changing the way in that journalists should publish structured, machine-readable data, alongside the traditional 'big blob of text': For example, say a newspaper has written a story about a local fire. Being able to read that story on a cell phone is fine and dandy. But what I really want to be able to do is explore the raw facts of that story, one by one, with layers of attribution, and an infrastructure for comparing the details of the fire—date, time, place, victims, fire station number, distance from fire department, names and years experience of firemen on the scene, time it took for firemen to arrive—with the details of previous fires. And subsequent fires, whenever they happen<sup>7</sup>. And last but not least another good start for journalists are the many sessions speakers the [School of Data Journalism 2013](#) by the European Center for Journalism, in Perugia, Italy, especially about finding stories in the data, finding the smoking gun, exploring the patterns.

### **Data visualization**

Besides digging for data, journalists are utilising the rising information flood by borrowing data visualisation techniques from computer scientists, researchers and artists. Some newsrooms are already beginning to retool their staff and systems to prepare for a future in which data becomes a medium. But how do we communicate with data, how can traditional narratives be fused with sophisticated, interactive information displays? For [data visualisation](#) there is [an introduction by Geoff McGhee](#).

### **Good practice and selected case studies**

To show the range of possibilities of data journalism, here are some references to pioneering and outstanding projects that have arisen as a kind of state of the art.

In the widely publicised Data Blog The Guardian regularly publishes under the tag line "Facts are sacred" the data sources behind their stories and asks the reader to join in (see [www.guardian.co.uk/news/data](http://www.guardian.co.uk/news/data) blog). An excellent example is The Guardian's evaluation of the "spending of parliamentarians" - a classic crowdsourcing project where 458 832 documents can be viewed by users and marked as relevant or irrelevant. The Guardian journalists helped to evaluate the mountain of data and report on interesting details. Other examples are semantic web-based offerings by the BBC: Powered by RDF, XML, and dynamic aggregations of up to 800 per visit to the webpage "World Cup 2010" ([http://news.bbc.co.uk/sport2/hi/football/world\\_cup\\_2010/default.stm](http://news.bbc.co.uk/sport2/hi/football/world_cup_2010/default.stm)) and "BBC Earth ([www.bbc.co.uk/wildlifefinder/](http://www.bbc.co.uk/wildlifefinder/)). These are prime

examples of efficiency and the wise use of databases. When The Guardian - along with the New York Times and Der Spiegel - introduced the "**Afghanistan WarLogs**" of **WikiLeaks** as a data collection online, Belam said that the Guardian already had an advantage, "because we did it for years," And when the British government had to disclose the expenses of your MPs, the Guardian had the data (see <http://mps-expenses.guardian.co.uk/>). s

That British and US media seem to be leading the way through the data jungle, is no coincidence: in the USA data is readily available through freedom of information legislation. An additional thrust is provided by the so-called open-data policy in the US and the UK - always evoked by the clearly audible reputation of online and civil rights activists. The Open Data freedom of information means that it must be in machine-readable form so that programmers can process the data and combine with other sources. . This enables journalists such as Aron Pilhofer of the New York Times (NYT) to scour the statistics to harvest for specific stories. And programmers with the material can feed information portals, such as the website "Where does my money go?". Tis website shows interactively how the British government manages taxes, see [www.wheredoesmymoneygo.org](http://www.wheredoesmymoneygo.org). Sometimes the stories evolving from databases also carry the potential for tangible controversy. I n 2007, the Los Angeles Times began its "Homicide Report" which used statistics which led to killings in the city and enriched these cases with a mash-up of name, address, and mapped the locations and photo of each victim as well <http://projects.latimes.com/homicide/map/> . In another data-coup the newspaper has now logged the performance of almost the entire teaching staff of Los Angeles: It published the official rating of 6,000 primary school teachers in 470 public elementary schools in the city. The basis of assessment is the performance of students in standardised final exams. Readers of the LA Times can now see any teacher's record with information on career and "effectiveness" see <http://projects.latimes.com/value-added/> .

### **Inspiring cases of data driven journalism**

An informative selection of [30 projects in data journalism](#) is a good read to get a first impression on what has been already achieved by data journalists, hackers and their supporters. Here are additional cases which show what can be done ... inspiring food for thought.

#### **Farmsubsidy**

Farmsubsidy.org helps people find out who gets what, and why. Farmsubsidy.org was founded by EU Transparency, a non-profit organization in the UK and the Danish International Centre for Analytical Reporting. Using citizens' rights to access government information they try to obtain detailed information about payments and recipients of farm subsidies in every EU member state. Interesting movie: [Fields of Gold: Lifting the Veil on Europe's Farm Subsidies](#).

### **Las Vegas Sun: Do No Harm – Hospital Care in Las Vegas**

An amazing data driven story in terms of a multimedia research on maladministration in the hospital care of Las Vegas, which was awarded the Goldsmith Prize 2011 and is considered a contender for the Pulitzer Prize: <http://www.lasvegassun.com/hospital-care/> Background: Las Vegas Sun Shines Light on Nevada Health Care  
[http://www.cjr.org/the\\_observatory/las\\_vegas\\_sun\\_shines\\_light\\_on.php?page=all](http://www.cjr.org/the_observatory/las_vegas_sun_shines_light_on.php?page=all)

### **Los Angeles Times: High salaries stir outrage in Bell**

Exposing widespread city hall corruption in tiny Bell, California: An LA Times report on the huge salaries of top administrators and elected officials in the small, working-class city of Bell has ignited community anger, calls for resignations at City Hall and condemnation from politicians and civic groups. Most council members are earning nearly \$100,000 for part-time work, far more than the \$400-a-month stipend recommended for a city the size of Bell. The LA County district attorney and the state attorney general have opened inquiries. <http://www.latimes.com/news/local/bell/>

### **National Public Radio: Behind The Bail Bond System**

Two-thirds of the inmates in US jails are petty, nonviolent offenders who are there for only one reason: They can't afford their bail. Sometimes, it's as little as \$50. Some will wait behind bars for a year before their cases make it to court. And it will cost taxpayers \$9 billion this year to house them. <http://www.npr.org/series/122954677/behind-the-bail-bond-system>

### **ProPublica, NPR's Planet Money, Chicago Public Radio: The Wall Street Money Machine**

Enticed by profits and bonuses, Wall Street took advantage of complicated mortgage-based instruments to reap billions, only to exacerbate the eventual crash. <http://www.propublica.org/series/the-wall-street-money-machine>

### **Pnadodaily: [Blogger turns to tips, briefly becomes "best-paid" journalist in his country](#)**

Keith Ng, an freelance journalist from Australia, gathered data which later became the [government's largest ever security breach](#). But rather than sell the story to a newspaper or magazine he [blogged about it](#) on a website called [Public Address](#). The files from New Zealand's Ministry of Social Development included names and addresses of children living in protected care, investigators and clients in fraud investigations, invoices, and medical prescriptions — and that was just the start of it...  
<http://pandodaily.com/2012/10/15/blogger-lands-huge-scoop-takes-tips-becomes-best-paid-journalist-in-new-zealand/>



## **Contribution by Duncan Campbell, British freelance investigative journalist**

Duncan had some very good advice for delegates at the workshop. His knowledge of data journalism and his work in the renowned Offshore Leaks (<http://www.icij.org/offshore>) made his practical knowledge invaluable. According to Duncan, his alternative career as a forensic expert on computers gave him some very useful skills. He mentioned some of the investigations he had been involved in, in which data journalism played a key role. Data journalism uncovered UK government support for the US invasion of Iraq in 2003. Under Tony Blair, a disinformation campaign produced a dossier on Saddam Hussein's weapons of mass destruction.

The dossier was supposed to be the work of the British Intelligence Services. It was produced as a Word document, which meant that there was hidden metadata. An investigation identified ten members of the staff of Blair's policy office who had worked on the document – they were clearly not members of British intelligence. A text retrieval search showed that 15% of the document was copied from a post graduate research study. As much as 60% of the 'dodgy dossier' was plagiarised, and not intelligence at all. Some of the tips Duncan provided were:

### **Approaches and shortcuts**

- Every time you get a chance to get documents, grab them and index them!
- Learn how to use Excel, learn how to manipulate and visualise data in Excel.

### **Mistakes**

- Be careful with Excel. It is essential for small amounts of data, but big data needs more. Excel can strip out deep relationships and mean that a relational database has to be rebuilt.
- Big data needs big statistical packages or SQL.
- When data is not freely accessible, be very specific about what data you want and why you are asking for it.

### **Audience questions**

#### ***Is it feasible for scientific data to be opened to everyone else?***

DC: Yes! But be careful. You may access data from a few studies but there is no gold standard: researchers testing the same hypothesis may use different tools. We need a standard of openness. I believe scientists should open their data. It is a 21st Century version of peer review?

#### ***What about covert reporting? When is it acceptable to hack data?***

DC: Telling lies and putting out incorrect information may also be part of journalism, but this needs to be very strictly monitored. You must have a good, documented, prior suspicion that something wrong is being done. You must have a real, high, public interest in mind. You need to ask yourself "If we use a covert method, will the balance of public opinion be with us?" That's the litmus test. You have to be sure that normal methods won't work.

## **Contribution by Anina Mumm, South African scientist and journalist, for Oxpeckers**

Anina provided insights into Oxpeckers’\* recent investigation into rhino-horn smuggling syndicates in the United States, a topic previously unexplored in world media. The three-month investigation used a combination of data-mining tools and traditional investigative techniques to expose the US as a major hub in the illicit trade. Articles were published in [100Reporters](#) in the US and in South Africa’s [Daily Maverick](#) in late May 2013. The investigation was part of part of ongoing research by Oxpeckers into illegal wildlife trafficking across international borders, and its impact on near-extinct species in Southern Africa. Anina will touch on this and other Oxpeckers projects, including investigations into illegal logging, pollution hotspots and extractive industries. Oxpeckers is collaborating with [InfoAmazonia](#) to develop a comprehensive database of eco-offences in Africa. The portal uses best-of-breed mapping technologies and geo-data analysis to track and expose the crime syndicates, corrupt officials and greedy corporations that are looting Africa’s natural resources.

*(\*An Oxpecker is a type of bird found in the savanna region of Sub-Saharan Africa. They perch on large mammals like rhinos, eating ticks, botfly larvae and other parasites that pester the animals.)*

### **Questions and answers**

#### ***What code, skills and trends underpin the future of journalism? What projects are trendsetting?***

Most environmental stories play out over large geographic areas. Organized crime uses sophisticated circumvention tools and technology to avoid detection. Syndicates also use shell corporates or straw men to hide their involvement. The Oxpeckers investigative environmental journalism project is the first of its kind in Africa.

It combines traditional investigative reporting with data analysis and geo-mapping tools to expose eco-offences and track organised crime syndicates in sub-Saharan Africa. It is one of 20 digital journalism projects awarded technical support from the [2012 African News Innovation Challenge](#). Oxpeckers uses geo-anchored stories, such as the poaching of rhinos and elephants, or the plunder of rainforests by production-line syndicates, using forensic data analysis, evidence-based narrative and dynamic visualizations, all with re-usable digital tools.

#### ***How do you utilize software tools and social media applications?***

The Oxpeckers centre uses sophisticated mapping software, geo-data analysis and forensic data tools to compile and visualize environmental data. Complex stories are told visually using dynamic infographics, “single story” animated maps and data visualization to augment more traditional story packages.

#### ***What are modes of co-creation and collaboration between journalists & hackers across countries?***

Embedded Oxpecker fellows, on secondment from other media in Africa and Asia (including Mozambique and China), spend three-month fellowships with the Center using its resources and tools to investigate major environmental stories for Oxpeckers and for their home audiences. Oxpeckers has

also secured a software development collaboration with [InfoAmazonia](#), based in Brazil, to create its online visualization tools. Furthermore, most of the Center's investigations are transnational, with the underlying evidence and other source materials available to other media and environmental guardians for use in their own work. Each major exposé includes tip sheets and tutorials to help other journalists replicate our evidence-based techniques.

### ***What have you learned from successful projects and mistakes from case studies?***

Accessing current, official information on syndicates driving illicit cross-border trade in natural resources is extremely difficult, unless you have the time and money to attend every court case and follow law enforcement investigators on a daily basis. No media has the resources to do this. As a result, surprisingly little is known about the big picture behind the illicit trade afflicting sub-Saharan Africa. Who is driving it, how is it happening and where is the money going?

The illegal trade in wildlife is the fourth largest global crime stream, after narcotics, counterfeiting of products and currency, and human trafficking. Crimes against nature have serious implications for the security and prosperity of Africans, with the proceeds reportedly used to fund organised crime and terrorism on the continent. Oxeckers' journalists are investigating the links between on-the-ground poachers, the middlemen and the syndicates placing the orders for illegal wildlife commodities. They are tracking smuggling routes from sub-Saharan Africa through various overseas countries to Asian markets.

In order to ascertain the role of US smuggling syndicates in the global rhino horn trade, Oxeckers used the Freedom of Information Act to access court records and other official documentation. The lesson learnt is that these kind of investigations take time and effort. Available skills for processing data and turning it into compelling narrative have also been a challenge, but the future looks positive as the fields of data journalism and science journalism are now rapidly growing in the region.

### **Contribution by John Burne-Murdoch from the Guardian newspaper in the UK**

John was unable to attend the workshop, but he very kindly agreed to share some information for this reader. For **scraping and mining**, he suggests using Scraperwiki (<http://bit.ly/scrpwk>), OutwitHub (<http://bit.ly/outwthb>), XML, JSON(<http://bit.ly/jsonxml>), SQL, DBMS and Google Forms (<http://bit.ly/ggforms>). For **cleansing**, he suggests using Google Refine (<http://bit.ly/ggrefine>), Stanford Data Wrangler (<http://bit.ly/wranglr>) and Excel Functions (<http://bit.ly/mjwescel>). For **mapping**, he is using Google Fusion (<http://bit.ly/ggftabs>), Carto DB (<http://bit.ly/cartodb>). For **dashboards**, he prefers Tableau (<http://bit.ly/tabpublic>) and Qlikview (<http://bit.ly/qlikview>). Illustrative charts are best drawn with DataWrapper (<http://bit.ly/datawrpr>), Google Charts (<http://bit.ly/ggcharts>) and Infogr.am (<http://bit.ly/infgm>). **Network graphs** can be created with Gephi and d3 offers rich interactive.

John points out, that journalists cannot escape the need for statistics, and John suggests taking a look at The Spirit Level (<http://bit.ly/sprtvl>), Statwing (<http://bit.ly/sttwing>) and R (<http://bit.ly/rstat>). **Data is infiltrating the mainstream of journalism, he says, so take a look at <http://bit.ly/nsafiles>, <http://bit.ly/gdnbigdata> and <http://bit.ly/cannesbd>.**

## **An additional annotated list of selected tools and sources**

[DataWrapper](#) is an open source tool helping everyone to create simple, correct and embeddable charts in minutes.

[Tableau Data Visualization](#) is an easy to use, drag-and-drop software that can help anyone see and understand their data. In the free 'public version' really useful, besides that you cannot save results, but screenshots do the job as well.

[Pajek Network Analysis](#) is a package for the analysis of large networks. It can handle very large networks, with thousands of nodes and tens of thousands of links. Both single node and bipartite networks can be analysed. A wide range of published network analysis algorithms have been implemented in Pajek, including such staples as network diameter, cluster coefficient and density, and most run very rapidly.

### [Jigsaw](#)

Software for huge amount of PDF document analytics for exploring and understanding document collections

## **Networks, organizations, links, notes, annotations**

### [Global Investigative Journalism Network \(GIJN\)](#)

The GIJN is composed of nonprofit investigative journalism organizations that produce stories, conduct training, provide resources, and encourage the creation of similar nonprofit groups. It was created in 2003 when more than 300 journalists from around the world gathered for the second Global Investigative Journalism Conference in Copenhagen. Since then it has grown to more than 80 member organizations in 38 countries.

### [IRE](#)

Investigative Reporters and Editors, Inc. is a grassroots nonprofit organization dedicated to improving the quality of investigative reporting. IRE was formed in 1975 to create a forum in which journalists throughout the world could help each other by sharing story ideas, newsgathering techniques and news sources.

### [datadrivenjournalism.net](#)

A free online data journalism course with five leading experts

### [Journalismfund](#)

Journalismfund.eu (formerly known as Fonds Pascal Decroos voor Bijzondere Journalistiek vzw) is an independent non-profit organization established with the purpose of promoting quality cross-border and in-depth journalism in Europe. <http://www.journalismfund.eu/dataharvest13>, <http://www.wobbing.eu/news/data-harvest-conference-2012>

### [Openknowledgefoundation](#)

A community of civic hackers, data wranglers and ordinary citizens intrigued and excited by the possibilities of combining technology and information for good – making government more accountable, culture more accessible and science more efficient. A good read is [openspending Digest](#) and discussion list

### [European Journalism Center](#)

The European Journalism Centre (EJC) is an independent, international, non-profit institute dedicated to the highest standards in journalism, primarily through the further training of journalists and media professionals. Building on its extensive international network, the Centre operates as a facilitator and partner in a wide variety of training projects.

**Sources:**

---

<sup>1</sup> Contributing institutions and websites: Investigative Reporters and Editors (IRE), European Journalism Center (EJC), Global Investigative Journalism Network (GIJN), Initiative Wissenschaftsjournalismus, Germany; Dossier: Hintergrundmaterial Computer-Assisted Reporting (CAR) , March 2011,

<sup>2</sup> A little selection on the websites: [www.farmssubsidy.org](http://www.farmssubsidy.org), <http://fishsubsidy.org/>, <http://datajournalism.stanford.edu>, <http://www.gapminder.org>; [www.wheredoesmymoneygo.org](http://www.wheredoesmymoneygo.org)

<sup>3</sup> Quoted as in <http://www.niemanlab.org/2010/08/how-the-guardian-is-pioneering-data-journalism-with-free-tools/>

<sup>4</sup> [www.zeit.de/2010/31/01-Afghanistan-WikiLeaks](http://www.zeit.de/2010/31/01-Afghanistan-WikiLeaks)

<sup>5</sup> [www.guardian.co.uk/world/datablog/interactive/2010/jul/26/ied-afghanistan-war-logs](http://www.guardian.co.uk/world/datablog/interactive/2010/jul/26/ied-afghanistan-war-logs)

<sup>6</sup> <http://projects.washingtonpost.com/top-secret-america/>

<sup>7</sup> [http://datajournalismhandbook.org/1.0/en/introduction\\_4.h](http://datajournalismhandbook.org/1.0/en/introduction_4.h)