



REUTERS
INSTITUTE for the
STUDY of
JOURNALISM

WORKING PAPER

Does Open Data Need Journalism?

Jonathan Stoneman

DATE: September 2015

Executive Summary

The Open Data movement really came into being when President Obama issued his first policy paper, on his first day in office in January 2009. The US government opened up thousands of datasets to scrutiny by the public, by journalists, by policy-makers. Coders and developers were also invited to make the data useful to people and businesses in all manner of ways. Other governments across the globe followed suit, opening up data to their populations.

Opening data in this way has not resulted in genuine openness, save in a few isolated cases. In the USA and a few European countries, developers have created apps and websites which draw on Open Data, but these are not reaching a mass audience.

At the same time, journalists are not seen by government as the end users of these data. Data releases, even in the best cases, are uneven, and slow, and do not meet the needs of journalists. Although thousands of journalists have been learning and adopting the new skills of datajournalism they have tended to work with data obtained through Freedom of Information (FOI) legislation.

Stories which have resulted from datajournalists' efforts have rarely been front page news; in many cases data-driven stories have ended up as lesser stories on inside pages, or as infographics, which relatively few people look at.

In this context, therefore, Open Data remains outside the mainstream of journalism, and out of the consciousness of the electorate, begging the question, "what are Open Data for?", or as one developer put it – "if Open Data is the answer, what was the question?"

Openness is seen as a badge of honour – scores of national governments have signed pledges to make data open, often repeating the same kind of idealistic official language as the previous announcement of a conversion to openness. But these acts are "top down", and soon run out of momentum, becoming simply openness for its own sake.

Looking at specific examples, the United States is the nearest to a success story: there is a rich ecosystem – made up of government departments, interest groups and NGOs, the media, civil society – which allows data driven projects the space to grow and the airtime to make an impact. (It probably helped that the media in the US were facing an existential challenge urgent enough to force them to embrace new, inexpensive, ways of carrying out investigative reporting).

Elsewhere data are making less impact on journalism. In the UK the new openness is being exploited by a small minority. Where data are made published on the data.gov.uk website they are frequently out of date, incomplete, or of limited new value, so where data do drive stories, these tend to be data released under FOI legislation, and the resulting stories take the form of statistics and/or infographics.

In developing countries where Open Data Portals have been launched with a fanfare – such as Kenya, and more recently Burkina Faso – there has been little uptake by coders, journalists, or citizens, and the number of fresh datasets being published drops to a trickle, and are soon well out of date. Small, apparently randomly selected datasets are soon outdated and inertia sets in.

The British Conservative Party, pledging greater openness in its 2010 manifesto, foresaw armies of "Armchair Auditors" who would comb through the data and present the government with ideas for greater efficiency in the use of public funds. Almost needless to say, these armies have never materialised, and thousands of datasets go unscrutinised by anybody.

In countries like Britain large amounts of data are being published but going (probably) unread and unscrutinised by anybody. At the same time, the journalists who want to make use of data are getting what they need through FOI, or even by gathering data themselves. Open Data is thus being bypassed, and could become an irrelevance.

Yet, the media could be vital agents in the quest for the release of meaningful, relevant, timely data. Governments seem in no hurry to expand the “comfort zone” from which they release the data which shows their policies at their most effective, and keeping to themselves data which paints a gloomier picture. Journalists seem likely to remain in their comfort zone, where they make use of FOI and traditional sources of information.

For their part, journalists should push for better data and use it more, working in collaboration with open data activists. They need to change the habits of a lifetime and discuss their sources: revealing the source and quality of data used in a story would in itself be as much a part of the advocacy as of the actual reporting.

If Open Data are to be part of a new system of democratic accountability, they need to be more than a gesture of openness. Nor should Open Data remain largely the preserve of companies using them for commercial purposes. Governments should improve the quality and relevance of published data, making them genuinely useful for journalists and citizens alike.

Overview

This study has been compiled through conversations with datajournalists, and by examining published stories, as well as statements made by various governments over the last decade or so. Finding out precisely what data have been used in the research of a particular story is difficult, if not impossible. Sometimes it is not possible even to discern that data have been used at all. Several stories were specifically identified by the authors to fellow datajournalists who are members of the NICAR-L email list.

The study is shaped as follows:

A look at the history of data and Open Data in journalism, including the growth of datajournalism training and bodies such as NICAR. (Pages 3-6)

A look at three country case studies: how Open Data has, or has not, had impact in Kenya (p.6-8), the UK (8-9) and the US (9-10), as well as more recent cases where Open Data, though in place, has not taken off: Burkina Faso, Croatia to see how Open Data loses momentum after a big fanfare-led launch. (11-13).

Then we look at openness in context – how a country’s culture of openness or relative secrecy affects expectations of Open Data. We also look at obstacles to Open Data being used by journalists, and the relationship between supplier (government) and demand (end-user), as well as how data are used by journalists. (13-20)

This leads to a look at the impact of data on audiences, including how far people say they get their news from graphs and graphics. (20-21)

We look at whether there is a need for an alliance between journalists and advocates of Open Data in order to increase impact, and discuss whether the question “does open data need journalism?” has a corollary in “does journalism need open data?” – both questions

invoke the issue of disintermediation: the process by which the audience receives news without needing journalists to process it on their behalf. (21-23)

This leads us to the concluding suggestions for action (24-25).

A list of sources begins on p.26.

Argument

Open Data has been touted as the “next big thing” for almost a decade: many governments, promising greater transparency, now publish huge amounts of data online, covering all kinds of government and civic activity; spending, expenses and salaries, planning data, data on emissions and water purity.

But, what is the effect on the great majority of tax payers of this openness? Do people feel better informed, or is Open Data the preserve of a nerdish minority? Do published data find their way into news stories, or are they doomed to remain unviewed by all but a handful of citizens on the host website?

When it comes to the possible role of journalists in the analysis of data, at one end of the spectrum there is the idea that “if journalists don’t examine published data, nobody will”. At the other end is the disintermediated world in which complete openness means there is no role for journalists (the same argument which was made as Social Media and citizen journalism arose to – potentially – supplant the mainstream or legacy media; “people will find out what they need to know without any intervention from the media”).

In this study, we will look at what has been happening in three countries – the USA, the United Kingdom, and Kenya. We will also look briefly at what is happening in countries with the most recent government commitment to openness, such as Burkina Faso. Each case shows the impact of Open Data at a different stage of development, with its own set of clues as to what the next steps might be.

The UK and US are at the top of one of the key data performance league tables – the Open Data Barometer¹, and both are highly placed in a similar assessment by the Open Knowledge Foundation². Kenya is the highest placed developing country.

Open Data Initiatives

In 2009, President Obama, in his first full day in office, signed a memorandum intended to “create an unprecedented level of openness in government”. The move to greater openness as part of participatory government was not a surprise. In the transition period before the inauguration, (Lathrop 2010, 116) “visitors to change.gov were invited to “Join the discussion” on topics such as healthcare reform, the economy, and community service....Finally there was the “Citizens’ Briefing Book” an attempt at making sure that at least some iconoclastic ideas from the public made their way, unfiltered, directly into the president’s hands.” After the inauguration the US Executive expanded the quality and

¹ <http://barometer.opendataresearch.org>

² <http://index.okfn.org/place/>

quantity of data available on USASpending.gov, which was intended to lay the groundwork for making the economic stimulus and recovery expenditures public, and creating a high-level process – itself conducted in a highly inclusive way – to develop open government policies under the Office of Science and Technology Policy.

In Chapter 2 of the Open Government, “Government as Platform” Tim O’Reilly said: (Lathrop 2010, 12) “There is a new compact on the horizon: information produced by, and on behalf of citizens is the lifeblood of the economy and the nation; government has a responsibility to treat that information as a national asset. ...Government is a convener and an enabler rather than the first mover of civic action. In the same chapter O’Reilly says – “the goal is not just to provide greater access to government data, but to establish a simple framework that makes it possible for the nation – the citizens, not just the government – to create and share useful data”.

This is a very optimistic view of the potential of openness, of Open Data and Open Government. Other countries followed suit. Open Data initiatives mushroomed in dozens of countries. But, crucially, in none of the numerous manifestos and policy papers, is “journalism” mentioned. The assumption is that the people will make use of their newly found access to Open Data in any way they please. President Obama made the link between openness and economic recovery at a town hall meeting in Elkhart, Indiana, in February 2009: (Lathrop 2010, 117) “I’ll enlist all of you. You can be my eyes and ears.” Taken at face value, this powerful rhetoric cuts out journalism, the primary purpose of which, according to *The Elements of Journalism* (Kovach 2001, 12) is “to provide citizens with the information they need to be free and self-governing”.

Some journalists had seen this coming. In 1973, Philip Meyer had published “Precision Journalism” – a book recommending that journalists adopt social science techniques of statistical objectivity and academic rigour. Then, as Meyer notes in the preface to the 4th edition: (Meyer 2002, vii) “In the next phase, many newspapers and broadcasters became interested in taking control of public opinion measurement away from politicians and commercial interests.” Technology was the key ingredient: “the aspect that first impressed its intended audience of journalists was the use of computers. ...since computers were expensive and mostly hard to use, their adoption by a mere news person was a definite novelty.”

Four decades ago the internet was a distant dream, and Open Data was not on the radar. Use of computers, and later the internet as basic tools in the newsroom grew rapidly. At the same time the market for data-journalism skills (also known as Computer Aided Reporting, or CAR) has also grown. The US-based organisation *Investigative Reporters and Editors (IRE)* grew the *National Institute for Computer Aided Reporting, NICAR*, to meet demand for training and discussion of journalism’s new skillset, which includes spreadsheets, scraping, relational databases, and visualisation tools.

The annual NICAR conference has grown in attendance to a recent peak of more than 1000 attendees in 2014. NICAR is only the biggest of many offers. The datajournalism training

industry is present in some form on every continent, and in the last few years online courses have also become more available, many of them free.

Computers, whether networked in a physical newsroom, or a virtual office in a reporter's kit bag, are more powerful than they have ever been, with the ability to run sophisticated software which is often open source and made freely available. Data is being provided by governments without anyone having to ask for it. It would seem that the ingredients are in place for Philip Meyer's "Precision Journalism" to start making a real impact. But what is the impact on journalism, and the news as consumed by citizens?

Put crudely, the results are mixed. It is surprisingly difficult to measure the true impact of Open Data on journalism. How does one identify the presence of data in a story? It is not always obvious, but data will be deployed in several ways:

- as figures in a text or a chart: these might appear as statistics, or simply as a number "coalition's £56M a day bill to private companies", "Devon and Cornwall police fail to investigate more than 30000 crimes"³,
- as part of a graphic, sometimes one with which the reader/user can interact and vary the inputs in order to produce different results eg. "Mapping poverty in America"⁴
- as part of a question to a politician or official
- as a fact in an otherwise text-based story – eg the names of politicians with the highest expense claims, companies which have received the most in government subsidies

Datasets are often just one of the elements in the research and presentation of a story. Just as they are not obliged to reveal all the sources which have driven or influenced their work, journalists don't necessarily draw attention to the fact that they have used data in their research.

Journalists in many countries have been used to obtaining data under Freedom of Information (FOI) legislation, and they have been understandably reluctant to signal to other journalists the precise dataset they have obtained.

(In July 2015 a group of US Federal agencies announced a pilot scheme which would mean a release of information in response to a request under the Freedom of Information Act (FOIA) would be made public⁵, ie a release to one would be a release to all. A blog post from the Florida-based Poynter Institute⁶ was headlined "the government is now giving away your scoops". The article was more nuanced, but resistance to the idea that everyone, public and journalist alike, might find out what a journalist had requested under FOIA may be little more than a general resistance to revealing one's sources. Arguably, knowing a dataset had been released at a competitor's request would not necessarily make it easier to work out what angle that competitor was planning to take. In the UK a similar suggestion to

³ Guardian, Exeter Express & Echo headlines 2011

⁴ <http://www.nytimes.com/newsgraphics/2014/01/05/poverty-map/>

⁵ http://www.huffingtonpost.com/entry/with-release-to-one-release-to-all-us-pilots-new-freedom-of-information-policy_559f1661e4b05b1d0290135b?

⁶ <http://www.poynter.org/news/mediawire/356961/the-government-is-now-giving-away-your-foia-scoops/>

make a release to one a release to all has been mooted as part of a revision of FOI legislation.

At this point we have only considered how to spot the use of any data in a story. The purpose of this study is to consider the impact of Open Data on journalism. “Open” is defined as follows: “Knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness”⁷.

How that level openness has an impact on journalism is the main question of this paper. And we will start by looking at countries which have Open Data initiatives in place. Kenya first, to see how momentum was quickly lost – if it ever truly existed. We will then look at the UK, where there is considerably more momentum, but the “ecosystem” of journalism, establishment, Open Data and popular understanding of it is incomplete; then on to the USA – where the Open Data movement was born, and where the “ecosystem” is in better shape. There too, more can be done. Finally, a short look at two cases where offering Open Data from the top down is not enough to create enough awareness of the possibilities it offers: Burkina Faso and Croatia are two recent examples of openness failing to make much impact, if any.

KENYA – early promise, stalled

Kenya, the highest placed developing country on the Open Data Barometer has an open data portal,⁸ about which the government’s introduction says “this site makes public government data accessible to the people of Kenya. High quality national census data, government expenditure, parliamentary proceedings and public service locations are just a taste of what's to come. Our information is a national asset, and it's time it was shared: this data is key to improving transparency; unlocking social and economic value; and building Government 2.0 in Kenya”.

In common with many national government Open Data initiatives, this declaration makes no mention of journalism or the media. The Kenyan Media Council’s annual awards website⁹ makes no mention of datajournalism or its use in winning entries. A look at the winners’ citations finds only one or two stories where Open Data might have been a needed in the research. In day to day journalism, a handful of journalists are making use of available data (census figures, spending on HIV) to match official claims against realities on the ground.

Eva Constantaras worked for *Internews* on a 5-month long project intended to incubate datajournalism in Kenya. This involved creating 3 teams, made up of journalists, graphics designers and developers, working on data-driven projects. None was based on Open Data – Freedom of Information requests were needed. The completed stories are listed in the references below. Constantaras¹⁰ says that “there is much less pressure on Kenyan media outlets to seek new forms of journalism – the press is not in the kind of crisis that it is in the USA, or, to a lesser extent, the UK. And you don’t need data to show corruption – since

⁷ <http://opendefinition.org/od/>

⁸ <https://opendata.go.ke/>

⁹ <http://www.mediacouncil.or.ke/en/mck/index.php/programs/media-excellence-awards>

¹⁰ Conversation with author

corruption is widespread and well-known. However, data has been used to name guilty officials, to quantify corruption, and to demonstrate misuse of international donors' funds": at this point datajournalism does have some impact. That said, Eva Constantaras argues that the Kenyan audience is not demanding more fact-based (ie potentially data-driven) reporting, that is, few bemoan the lack of reporting derived from Open Data because they perceive few "facts" in the stories they currently consume.

The stories which emerged from the Internews project covered mainly health, sexual health including use of contraception, abortion, unmarried mothers, spending on health. Open Government Data was the source of some of these stories, but, as in many statistics-based stories in Sub-Saharan Africa, NGOs such as the Africa Population Health Research Center (aphrc.org) are simply easier to use.

Four years after launch, Kenya's Open Data initiative is seen as stalling by some observers: in a blog entry on the Code4Kenya website in 2013, Nick Hargreaves said (Hargreaves 2013): "Today the portal has less than 3000 datasets with the latest one uploaded on December 12th 2012. Most of the data is from 2007-2011." *(In June 2015 the Kenyan data catalogue listed 257 datasets. The most recently updated were GDP figures and price index data up to 2009, although there was data such as Petrol Prices from 2015, and HIV infection data from as recently as 2014)*

Hargreaves assesses the reasons for this failure in this way:

"There's hardly any interaction on the portal anymore in terms of dataset suggestions and creation of visualizations. Only 12 apps are listed under 'community apps' and the mainstream media seems to have moved on, as it does when a story becomes stale. When the portal was launched, what was envisioned by most was better provision and access of services and accountability of government departments. Yet 2 years after the ribbon was cut and the champagne bottles popped we seem to be stuck at the beginning. What went wrong?"

He answers his own question, with this conclusion: "Clearly the whole program needs to be refreshed, with less hype and more work, and this time with policy guidelines in place that are strictly followed."

In Kenya, and in other countries where Open Data initiatives have not taken off, it could be that Open Data was not something the people demanded, but was delivered as a top-down exercise. In the developing world there is also an element of peer pressure – signing up to an Open Data initiative looks good (or, more accurately, refusing to sign up looks bad!)

Hargreaves says:

"Launching the portal should have been more than just Kenya wanting to be the new cool kid on the block, rather a push towards a transparent government that people can trust and for the government to give people better access to public services and amenities. So it has to be properly thought out and needs to be sustainable and active."

This is something which could be said of many Open Data initiatives. Even in a developed democracy such as the UK, some of the consequences of a top-down policy are evident.

UK – leading the way, but limited in relevance (and timeliness)

Britain is at the top of the Open Data Barometer, as defined by the Open Data Institute “for its leadership in Open Data by the Open Data Barometer (2013)”. For more than a decade, UK governments of different political hues have built on the Open Data initiatives of their predecessors to encourage Open Data to be released and reused across government: from the launch of the Power of Information Review by the Blair government in 2007, to the implementation of the data.gov.uk Open Data portal under the Brown government in 2009, to David Cameron’s letter to cabinet ministers in 2010, affirming previous progress in Open Data and setting out the coalition’s transparency agenda.

However it is hard to see access to Open Data breaking through into a many significant stories.

In 2009 when The Daily Telegraph won an award for its series on MPs’ Expenses, the documents which formed the basis of the story were not in the public domain (although they are now). The Guardian has, on the other hand, carried out several investigations into lobbying in Westminster using published declarations of interests of members of the both houses of parliament, and the interests of their assistants and secretaries. The Guardian’s Datablog¹¹ has published daily 2-3 data-based stories for several years. Many of them came from openly available data. Relatively few of them could be considered “front page news”.

This type of use of Open Data is rare. The British Journalism Awards¹² do not include a specific category for “data”; the awards categories include Business, Politics, Science, Sport, Campaign, New Journalist of the Year. In recent years no obviously data-based story has won a prize.

Looking at the kind of datasets available on the UK government website – data.gov.uk- it is not difficult to see why it is not generating many news stories: some datasets are too old to be of much use to the media – unless publication of the data is, in itself, a story, and that will be true only once! the subject matter is of dubious value too - the top ten UK datasets by number of downloads¹³ was as follows in May 2015:

Dataset Title
Road Safety Data
Bona Vacantia Unclaimed Estates and Adverts
English Indices of Deprivation 2010
Building Price and Cost Indices
Land Registry Price Paid Data
Statistics on Obesity, Physical Activity and Diet, England
Live traffic information from the Highways Agency
Land Registry Monthly Property Transaction Data

¹¹ <http://www.theguardian.com/data>

¹² <http://www.pressgazette.co.uk/awardshome>

¹³ <http://data.gov.uk/data/report/site-usage-publisher?metric=downloads>

Road safety would appear a fruitful topic for journalistic study – since thousands of incidents are well documented, including precise location, road type and condition, collision type, make and age of vehicle, and purpose of journey. But the potential interest is reduced when one finds that the most recent road safety data available dates from 2013. Not everything is so far behind, Land Registry Price Paid data¹⁴ are only a few months old, but even those few months' delay are arguably enough to reduce the dataset's journalistic value.

Relatively few data are updated in real time, and even fewer datasets cover the mainstream of story content – money, votes, and the data which government uses in its decision making. For everything else, journalists in the UK still tend to ask for the data they need, using the Freedom of Information Act (2000).

One of the founders of the UK's Open Data Institute, Sir Nigel Shadbolt, conceded that more needs to be done, in an interview in August 2015¹⁵ "It's the routine availability of data that provides a resource to make stuff happen. And we're quite a long way away from that". Despite the UK government agreeing the need for a National Information Infrastructure, and launching the project in 2013, Sir Nigel said:" The need for an Open Data infrastructure is essential. If we don't have the equivalent for data that we have with our roads, power grid, you won't build a 21st century digital state that is any more than little islands of privilege and monopolies."

USA – closer to a more perfect storm of data and journalism?

The United States has a longer history of openness, and its Freedom of Information Act of 1966 allows or encourages publication of much data which would remain under lock and key in many other countries. In recent years, Pulitzer prizes have gone to journalists (individuals as well as small teams) who openly acknowledge the importance of data in their work. The data involved have been acquired under the Freedom of Information Act rather than as freely available Open Data, and this is not surprising, given the Pulitzer's emphasis on Investigative Journalism.

The Pulitzer Prizes of 2015¹⁶ are obvious examples of this. The Wall Street Journal and the New York Times both had winners whose stories came from analysis of (very) big datasets. But both investigations – into political lobbying and influence for Eric Lipton at the New York Times¹⁷, and into Medicare by the WSJ team¹⁸ – relied on the dogged pursuit of Freedom of Information requests, rather than Open Data.

¹⁴ <http://data.gov.uk/dataset/land-registry-monthly-price-paid-data>

¹⁵ <http://www.computerworlduk.com/data/open-data-in-uk-job-half-done-3622607/>

¹⁶ <http://www.pulitzer.org/citation/2015-Investigative-Reporting>

¹⁷ <http://www.pulitzer.org/works/2015-Investigative-Reporting-Group1>

¹⁸ <http://www.pulitzer.org/works/2015-Investigative-Reporting-Group2>

Even when datajournalism is the specialism being rewarded, the pattern is the same – data are more often furnished through FOI than from open sources. The IRE/NICAR-run “Philip Meyer Awards¹⁹”, have most recently rewarded journalists for investigations into Medicare, health and safety for temporary employees, rising sea levels, wasteful prescription practices, inequality, and the methods used by political intelligence firms. When these stories were being researched, the data behind them could certainly not be considered “open”; it is common, of course, for either the journalist or the government body, or both, to publish the data once it has been released under the Freedom of Information Act, but this is necessarily “after the fact”.

That said, it is clear that many smaller US papers are able to produce useful stories on the basis of Open Data – covering such subjects as

- police shootings²⁰, deaths in custody²¹, stop and search²²
- pay and benefits of state and local officials²³
- city planning, local transport, road safety²⁴
- health care provision and standards of living – eg “the disappearing middle class”²⁵

The size of the journalism sector in the USA, and the relative coherence of the datajournalism community within it, has created a kind of ecosphere which helps to generate creative projects, and smart uses of available data. Coupled with a much more pervasive spirit of openness in the public sector, this has created a “perfect storm” which results in a much higher proportion of data being seen by the public at large, either through stories in the media, or through websites, or apps created by developers – followthemoney²⁶, and maplight²⁷ being good examples of sites which allow users to investigate the links between money and politics for themselves. It is the size and dynamic nature of this “ecosystem” which, I would argue, other countries not only lack, but need if Open Data is to become part of the relationship between government and governed.

Top down initiatives – still too early to see much impact?

Open data initiatives are launched frequently, sometimes at national level, often at city level. The Open Data Institute in London manages many of these, and summarises its activity on its website as follows²⁸:

¹⁹ <http://ire.org/awards/philip-meyer-awards/>

²⁰ <http://www.mypalmbeachpost.com/policeshootings/>

²¹

²² <http://www.theguardian.com/commentisfree/2015/aug/07/stop-and-search-uk-rural-black-people-police>

²³ <http://payday.cironline.org/>

²⁴ <http://www.delawareonline.com/article/20130714/NEWS1501/307140036/Delaware-s-dangerous-intersections>

²⁵ <http://www.oregister.com/articles/middle-523465-class-california.html>

²⁶ <http://www.followthemoney.org/>

²⁷ <http://maplight.org/>

²⁸ <http://theodi.org/nodes>

“ODI Nodes create Open Data impact globally. ODI Nodes contribute to the local, national and international development of Open Data. Operating at a city or regional level, they commit their local expertise and physical presence to:

- Develop and deliver training to build capabilities
- Connect people and businesses through membership and events
- Communicate stories and catalyse the adoption of standards, tools and processes

Nodes are organisations who wish to develop a sustainable business line that supports the business case for openness, and ensure Open Data is for everyone.

To underpin our approach, be clear how they operate and to protect the ODI’s reputation, each ODI Node adopts the ODI Charter. The charter is an open source codification of the principles and rules which the ODI uses to develop Open Data impact.

ODI HQ (based in London) charges ODI Nodes to be part of the network and shares revenue on income generated.”

Again – no mention of journalists as potential end users; the emphasis is on businesses and business models.

The list of nodes which have become part of this network is growing, but news of the activity of each node is not regularly updated, and it can be difficult to ascertain what has happened since a node’s accreditation. Searching using the name of the node or its portal (Open Data website), and words such as “hackathon” provides a rough and ready guide to activity levels.

One of the most recently accredited was Burkina Faso. The West African country, ranked by the UN as the third poorest in the world, launched the Burkina Open Data Initiative (BODI)²⁹ in June 2014. The (London-based) Open Data Institute said “The initiative aims to drive economic growth, boost innovation and demonstrate transparency and is supported by the UK’s Open Data Institute, its Paris Node: FivebyFive via the Partnership for Open Data, and the World Bank.” On the ODI website³⁰, Alfred N. Sawadogo, General Director, National Agency for ICT Promotion (ANPTIC), Ministry of Development of the Digital Economy and Post, Burkina Faso, is quoted as saying “The sharing and reuse of public data creates a dynamic of innovation in services to citizens and business. Ensuring that every Burkinabe has access to information in order to improve the quality of his or her life - this is one of the major objectives of the initiative.”

There have been a few hackathons in Burkina Faso in the last 12 months (data hackathons usually bring together data experts, coders, web developers, sometimes online startups, and occasionally journalists). Stories driven by data are as hard to spot there as anywhere else, because sources are not openly acknowledged. It is probably too early to expect many media stories to emerge from this newly Open Data, and it has not been possible to find a

²⁹ <http://data.gov.bf/>

³⁰ <https://theodi.org/news/burkina-faso-launches-open-data-initiative-with-mentoring-from-the-odi-and-funding-from-the-world-bank>

Burkinabe journalist willing to comment for this study, but it does appear that occasional stories such as a recent one³¹ on Government salaries, may be based on published data. That said, the government data portal contains more gaps than useful datasets: just 29 government bodies are indexed, and very few have so far uploaded data – none from the office of the President, and only 13 from the Finance Ministry, of which none is more recent than 2013.

Croatia, the most recent new member of the European Union, is another recently accredited ODI node. There too there have been hackathons and “Open Data Days”. Several individual Croatian cities have declared themselves “Open Data cities” – such as Osijek in Eastern Slavonia. But there is very little, if any, evidence of data reaching the media – despite, or more likely because of, the rapidly declining market for news in print there. At present the portal data.gov.hr contains fewer than 120 datasets.

Initiatives like this need publicity to take off. If Open Data is not publicised in the mass media, and journalists do not press for release of Open Data they can use in pursuit of stories, it is difficult to see Open Data breaking into public consciousness. Mainstream US papers including the New York Times and the Washington Post are increasingly covering stories about data as a subject in its own right – deficiencies in government data, releases of new datasets, and so on. In the UK there is minimal coverage of Open Data as a topic - “data” is more often covered in terms of breaches of the Data Protection Act, or civil servants losing memory sticks or computers. In 2014, plans to release health data en masse was seen principally as a story about privacy, and government making money from people’s data: hardly anything was written about the possibility of citizens or “armchair auditors” interrogating health data for themselves.

On the ODI’s News web page Ibrahim Elbadawi tracks recent progress of Open Data in the Arab world, drawing a conclusion in the final sentence which identifies the key problem for Open Data anywhere³²:

“So, several countries across the Arab region make noticeable progress towards embracing Open Data as a practice and culture. However, there are many issues need to be addressed to increase the value of these Open Data initiatives. These include filling up the legislative gaps, boosting the level of engagement of non-government parties, enhancing the quality of published data, and focusing on the actual value generated and of course building the capacity among the public sector Open Data teams to handle these initiatives.

Finally, it is worth examining how much engagement existing Open Data initiatives have had to understand whether improved rankings on eGovernment charts reflect a shift in citizen attitudes towards data and information exchange with government and how many Arab citizens have visited open government portals on Arab government websites.”

³¹ <http://www.fasopresse.net/societe/4288--soldela-realite-de-la-masse-salariale-du-personnel-de-letat>

³² <http://theodi.org/blog/a-sneak-peek-at-the-arab-regions-open-data>

Openness in Context

Openness cannot be achieved simply by publishing data. An Open Data policy itself fits into the context of how open government officials are to questions from the media, how quickly and easily they respond to FOI requests, and, as the Open Data Barometer notes:³³

“Effective Open Data policies require a degree of collaboration between the state, private sector and civil society. A balance is needed between governments with the capacity to create, manage, and publish data, and third parties with the technical skills, freedoms, and resources to use data as a tool for change. Governments that focus solely on increasing the supply of open data — without exploring ways to extend access to data literacy and skills, developing approaches to stimulate innovation, or putting in place foundations for data to be trusted — are likely to miss out on many of the benefits of Open Data”.

Cultures do not change overnight. The coalition government of Britain elected in 2010 offered citizens Open Data after centuries of government secrecy, or at best, government confidentiality. Reliance on “spin” and “spin doctors” long distorted the landscape, even as Britain’s government moved to make data available openly. For many years, British journalists were expected to rely on unattributable briefings as much as, if not more than, openly available sources of information. (Indeed, membership of the “lobby system” of off-the-record briefings required journalists not to acknowledge its existence in their articles!) Until the Blair governments of 1997-2010, the Prime Minister’s spokesman and most others would brief the media unattributably, so that journalists working in Westminster were obliged to refer to “sources close to xx”, as code which insiders would know could be translated as “xx speaking off the record”.

Some European countries, Sweden, for example, allow journalists to work in a culture where “everything is open by default”. Open Data there has added a new layer of information available to those able to make use of it. Datajournalism is practiced by most of the mainstream media, including the public broadcaster. Former Swedish TV journalist, Helena Bengtsson (now at The Guardian) says³⁴ training sessions there have already moved on from simply showing people how to deal with basic spreadsheet functions and are now demonstrating more sophisticated data tools such as SQL. Data visualisation and use of the various tools is also in demand.

Even so, the Swedish experience has its limitations – Bengtsson says Open Government Data is frequently used for relatively minor tasks such as geolocating children’s playgrounds, or public toilets. For “serious” stories Freedom of Information requests are still the normal way in.

The US government website, data.gov is a typical example of a mission statement which fails to mention the media as possible end-users³⁵: “Open government data is important because the more accessible, discoverable, and usable data is the more impact it can have. These

³³ <http://www.opendatabarometer.org/report/analysis/readiness.html>

³⁴ Conversation with author

³⁵ <http://www.data.gov/impact/>

impacts include, but are not limited to: cost savings, efficiency, fuel for business, improved civic services, informed policy, performance planning, research and scientific discoveries, transparency and accountability, and increased public participation in the democratic dialogue.”

However much governments say they want to see Open Data as a way of opening their business to the public at large, data remains predominantly the preserve of coders, developers and statisticians. Journalists are not seen as an end users of Open Data.

The British go a step further in identifying the end user – by inventing a new role. The British Conservative Party’s 2010 manifesto’s title was “Invitation to join the government of Britain” and said this about data³⁶: – “Drawing inspiration from administrations around the world which have shown that being transparent can transform the effectiveness of government, we will create a powerful new right to government data, enabling the public to request – and receive – government datasets in an open and standardised format”.

In May 2010, soon after taking over as Prime Minister, David Cameron said³⁷: “With a whole army of effective armchair auditors looking over the books, ministers in this government are not going to be able to get away with all the waste, the expensive vanity projects and pointless schemes that we’ve had in the past”.

If a government seeks “armchair auditors”, where will they be “recruited” from, and where will their work be published? A few individual citizens may take a brief look at published data out of curiosity or following up a particular interest, but in the main, the users of Open Data are much more likely to include pressure groups, businesses, and entrepreneurs seeking to make a profit by repackaging the data through an app or website. NGOs and other non-profit organisations are making use of Open Data: in the USA, transparency data in the fields of political finance, government expenditure, medical and transport data are all being made available for re-use through portals and online apps. Even so, only a small minority of citizens are aware of the tools available to them. Unless their publication has created a specialist app or website to crunch data, US journalists tend not to refer in stories to apps driven by Open Data.

There is another phenomenon which weakens the directness of access to Open Data by citizens or reporters. This is the number of government websites to place API’s (Application Program Interfaces) between the raw data and the end user. These have the effect of filtering the data, limiting access and analysis of it to the parameters intended by the publisher. Many API’s are open to any developer who wants to build a tool to interrogate the data, but this is a barrier to truly open use of the data. The knowledge required to create a workable API is greater, and arguably harder to acquire, than basic knowledge of a spreadsheet. As Robinson, Yu and Felten argue in Open Government (Lathrop 2010, 88), “APIs can be excellent, disappointing, or anywhere in between, but generally speaking,

³⁶[https://www.conservatives.com/~media/files/activist%20centre/press%20and%20policy/manifestos/manif
esto2010](https://www.conservatives.com/~media/files/activist%20centre/press%20and%20policy/manifestos/manifesto2010)

³⁷<https://www.gov.uk/government/news/pms-podcast-on-transparency>

providing an API does not produce the same transformative value as providing the underlying data in bulk.”

It is not just journalists who find aspects of open government data publication restrictive. In a blog post³⁸ of June 2015, Richard Speigal documents how UK Department for Education data makes life difficult for someone like him tracking school admissions data. He calls for improvements which include a publication schedule, a clear and prominent policy on API limits, consistent use of fields in the data, and a much better search engine within the data.gov.uk site.

Does Journalism need Open Data?

Leaving aside for a moment the identity of the intended end user, and the putative “armchair auditor”, we ought to consider briefly an inversion of the original question of this paper. Instead of asking “does Open Data need journalism?” we should also consider the extent to which journalists need Open Data.

When the Open Data movement began, and governments started opening their books, and publishing data, several journalists and editors, saw Open Data as a potential saviour of an industry in decline, with splintered audiences, rising staff costs, and increasing access to all of sources of information, there was a need for people with expertise to acquire, analyse and explain the data which governments were opening up. Simon Rogers, The Guardian’s first Data Editor said of data in general – not Open Data specifically - in a Guardian Short (e-book), Facts are Sacred (Rogers 2011): “There are still reporters out there who don’t know what all the fuss is about, who really don’t want to know about maths or spreadsheets. But for others, this new wave represents a way to save journalism. A new role for journalists as a bridge and guide between those in power who have the data (and are rubbish at explaining it) and the public who desperately want to understand the data and access it but need help. We can be that bridge.”

This has not happened, to the extent, or at the speed which Rogers and others may have hoped for. While journalists see part of their role as being watchdogs, holding power to account, even in relatively “clean” states those in power tend to see journalists as a nuisance.

In February 2012, on the “Talk about Local” blog³⁹, former UK civil servant William Perrin discussed the shortcomings of openness with government as “supply” and journalists as “demand”. He says: “the demand side needs much more work to get a public social economy of data working, by which I mean, helping regular folk find and use public data to achieve social outcomes – say a campaign on local development, for a new school or stopping arson. This is often well outside the Whitehall comfort zone.”

The “comfort zone”, whether Whitehall’s or any other country’s, is partly defined by the need to avoid criticism or blame when policies fail and/or public funds are wasted: the “gotcha” moments – when power is held to account and found wanting. In Open

³⁸ <http://www.webunknow.co.uk/play/posts/when-open-data-sucks>

³⁹ <http://talkaboutlocal.org.uk/open-data-forward-strategy/>

Government, Fung and Weil (Lathrop 2010, 107) look at the link between Open Data and the media's watchdog role "Open government efforts can [thus] plug into a media apparatus and public political culture that together make a very effective "gotcha" machine." They go on to argue that this is what is holding back the march towards true openness: "the larger responsibility of citizens is not just to judge when officials behave badly, but also to provide feedback on their performance in more nuanced ways, including registering approval when government performs well....Unfortunately, the current discourse of transparency – focused as it is on accountability and issues such as corruption – produces policies and platforms that are particularly sensitive to government's mistakes but are often blind to its accomplishments."

Governments cannot, of course, have it both ways – the same openness which sheds light on its accomplishments will inevitably expose mistakes and shortcomings. Opening data to examination means literally anyone can carry out the examination – and that means journalists as much as armchair auditors. Governments are going to have to adjust themselves to a new comfort zone, in which accomplishments and mistakes can be viewed by anybody, and to reassure themselves that opening up data to public view means giving everyone the tools to examine a government's record objectively.

How are data being used in journalism?

Turning the telescope round to look at openness from the media's point of view, it is clear that journalists still have much to do. For many journalists (and their audiences) the word "data" is practically synonymous with "statistics", and is therefore seen as something requiring special expertise.

At the same time, and perhaps partly as a way of avoiding the need to undertake a more thorough training in statistics, the advent of cheap/free visualisation tools has meant that data are frequently used to make charts, graphs, and interactive web pages. Organisations which have embraced data have put journalists in teams with developers and coders, and the result is often innovative and engaging.

At its best, datajournalism teams produce stories which could not have been completed without access to data. Examples include The Guardian's most recent interactive *The Counted*⁴⁰ – mapping victims of police shootings and killings across the USA; ProPublica⁴¹'s exposés of misspending and fraud within the Medicaid system, including the recent - *The Prescribers*⁴² and *Dollars for Docs*⁴³ in 2014; and The SunSentinel in Fort Lauderdale, Florida, which investigated off-duty policemen routinely flouting speed limits with impunity, and won a Pulitzer Prize in 2013 for it⁴⁴. The Guardian's site uses a combination of open data, items from news reports, and data requested under FOI legislation. ProPublica's data is a

⁴⁰ <http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database>

⁴¹ <https://www.propublica.org/>

⁴² <http://www.propublica.org/series/prescribers>

⁴³ <https://projects.propublica.org/docdollars/>

⁴⁴ <http://www.pulitzer.org/works/2013-Public-Service>

combination of data requested under FOI, and some published on the Medicare/Medicaid websites, while the SunSentinel's data was requested under FOI.

At the other end of the spectrum datasets are routinely turned into graphics and interactives with very little news value or journalistic content. Innovation is seen more in the presentation than the journalism: this is not to say that the journalism is poor, or not worthwhile, it is just that it tends to concentrate on minor stories, or on the minutiae of daily life – *ie the way we live now*. For example, The Guardian's most recent stories published on its datablog included, UK Premier League transfers, Australian politician's pay and benefits, the impact of parental leave at the company Netflix, why beer prices vary so widely around the EU. Even allowing for the fact that this was the summer "silly" season, they don't make a great case for datajournalism in general, or Open Data specifically.

When it comes to data which is turned into infographics, a look at The Economist's blog "Graphic Detail" for July and August 2015 included the following – Australia's carbon footprint; defence spending per capita, or "the price of paranoia"; gun ownership plotted against homicides in various countries; the impact of the Republican candidates' TV debates in the US (twice); death tolls of WW2 bombing campaigns, "the long shadows of Hiroshima and Nagasaki".

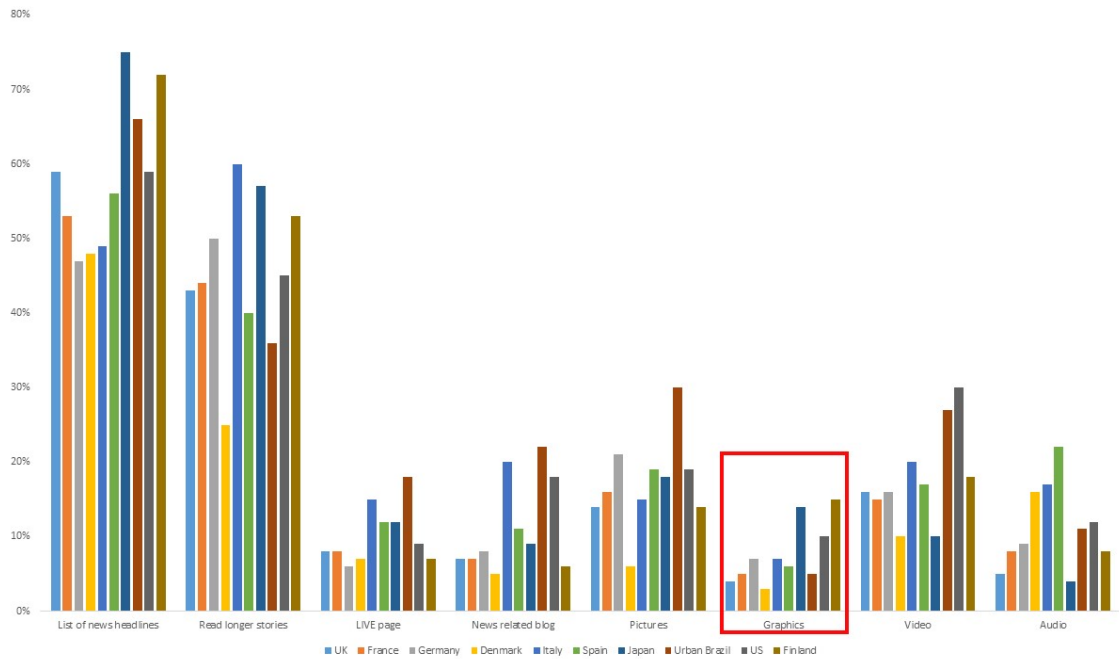
Further afield, a search for news stories invoking the word "infographic" during August 2015 throws up topics such as "why we need more women in tech", the rising number of parents in Yorkshire being fined for letting their children miss school, the most dangerous US cities for red-light running. The data underlying these reports is often but not always Open Data, but they exemplify the growing reliance on visualisation of data to show patterns in data.

Results are mixed. Some of the effort appears to have limited impact. In its Digital News Report, 2014, the Reuters Institute noted (Newman 2015):

"over the last year newspapers, broadcasters, and digital pure players have been producing more visual content, pictures, data-rich charts, animated gifs, and video itself. This is partly because of the multimedia capabilities of new devices, partly because distinctive visual content works well in Social Media – but also for commercial reasons. Video in particular is attracting much higher advertising premiums.

"Despite this, our research shows that, for the moment at least, most users remain wedded to words (traditional articles and lists), though we do find that pictures and videos are relatively more important in the US, Brazil, and southern European countries like Italy and Spain."

Breaking the responses into groups by country, and type of content, the disparity between news consumption through "words" and through "graphics" is all too clear:

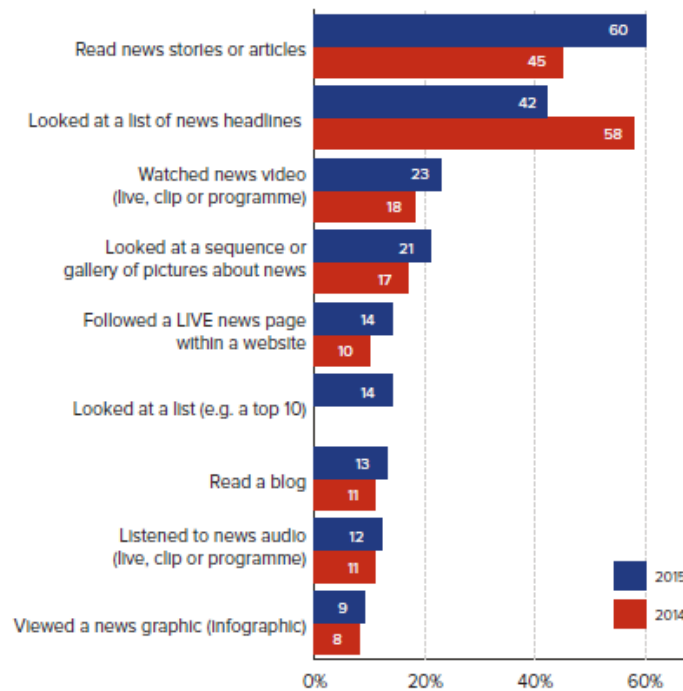


The countries surveyed were UK, France, Germany, Denmark, Italy, Spain, Japan, urban Brazil, the USA, and Finland.

The two national peaks within the (red-ringed) graphics group are Japan and Finland, but it's clear that no country's consumption of graphics comes close to the lowest consumption of headlines and stories (the two groups on the left).

In the 2015 report (Nic Newman 2015) the overall rating has risen by one percentage point. The percentage of respondents who said they had got their news by viewing an infographic ranged from 5% in Japan to 14% in the USA: slight increases in several cases from the 2014 responses, but Infographics, one of the mainstays of datajournalism output, are still not seen as a major source of news by consumers.

PERCENTAGE OF WAYS OF CONSUMING NEWS ACROSS ALL COUNTRIES



Q11. Thinking of the way you looked at news online (via any device) in the last week, which of the following ways of consuming news did you use? Please select ALL that apply. Base: All. Note: 'Looked at a list' added in 2015. 'Read news stories or articles' was previously 'Read longer news stories or articles'.

The implication for datajournalists is that turning data into graphics or tables is much less likely to make an impact on the audience than headlines and words.

These figures should not be taken to imply that datajournalism is a waste of time – online graphics derived from datasets tend to be “sticky”: web analytics show users spending longer on those pages, making use of the interactive elements, and for those who like such elements, the intention is to let users “play” with the data, to examine it within their own context, and/or look for patterns.

For those whose news consumption habits are more traditional, data needs to be embedded in the journalism in text-based stories just like any other element – press release, statement, interview, background briefings, etc.

Philip Meyer said as much in 2002, in the introduction to the 4th edition of *Precision Journalism* (Meyer 2002, 16): “generating piles of description is not enough. Just as storytellers need plots to hold their descriptions together, social scientists and precision journalists need theories to give their data its fullest meaning. There are times when we are so proud of the data spewing out of our computers that we’d like to just print it unaltered in the paper or online. But we can’t.”

Meyer's warning – that storytelling would still be needed – is echoed by several observers of journalism. In his blog, in May 2015⁴⁵ David Higgerson of the Daily Mirror's data team said: "Data journalism isn't easy journalism. It isn't about just taking data and making a map. It only works when journalists ask the question: 'How do I make this interesting and useful to readers?' Data journalism shouldn't be a passive process of interpreting data others choose to release – we need to be asking for data. Context is key. Data for data's sake makes no sense for readers."

"Data for data's sake" cuts both ways.

Simon Rogers said in *Facts are Sacred* (Rogers 2011): "Governments love data because publishing it can make them seem transparent." Appearing transparent by opening up data was in part a response to perceived image problems shared by several governments in developed countries in the mid-to-late 2000s. Once President Obama had published his administration's commitment to Open Data in 2009 other leaders were keen to demonstrate similar commitments. The result was the release of data as a top-down process, rather than a bottom-up response to popular demand – the public became passive recipients of Open Data.

In a blog post dated 4 June 2015, on the Democratic Audit website, Mark Frank⁴⁶ outlined the distinction between active transparency, and passive transparency:

"Success is typically measured by the amount of data that is published – governments and departments make much of the number of datasets – and the golden rule is all data should be published unless there is good reason not to. Many writers strongly resist the idea that government should select or interpret the data. It is not surprising that the result is mostly input transparency. Input data are on the whole easier to obtain and publish than output or outcome data. ... The national information infrastructure is a government initiative to select datasets based on users' needs that are key to the nation and demand special attention. But the passive transparency model remains dominant and while it does Open Data is unlikely to have a significant political impact."

Open Data - not having a significant impact on the public in many countries

While governments in all the countries we have discussed here (and many others) have publicly embraced Open Data, the selection of data, of formats, of licensing for re-use, is in their gift. For everything else there is (or may be) Freedom of Information legislation, which adds an extra step in the data-gathering process. Journalists who make use of data have so far either made passive use of what governments publish, or active use of datasets received in response to direct requests. This is not a recipe for progress towards real openness. As Aron Pilhofer, formerly of the New York Times, now Executive Editor of Digital at The Guardian, says⁴⁷:

⁴⁵ <https://davidhiggerson.wordpress.com/2015/05/25/infographic-how-to-create-great-journalism-online/>

⁴⁶ <http://www.democraticaudit.com/?p=13606>

⁴⁷ Email exchange with author

“ Ideally, the Open Data movement and journalists have a sort of symbiotic relationship, where new data is acquired (or liberated) and immediately that new data is eagerly lapped up by journalists who provide the analysis and storytelling the public needs. At least in the States, I'm not sure that has quite happened to the degree it should or could. And so if I were asked whether journalists need the Open Data movement, I would have to say, unfortunately, no. The inverse, however, is also just as true: the Open Data movement so far has not seen journalism as crucial. I hope that changes, because I do think a more formal alliance between journalists and open-data advocates would be a very powerful thing.”

A “more formal alliance”?

To date in most countries there is no discernible “alliance” between the media and Open Data advocates, that Aron Pilhofer half wishes for. Nor do we see the kind of collaboration between the media and government which would allow Open Data Initiatives to have real impact.

When proponents of openness make their case, the end-user is never named. Is the taxpayer, the “armchair auditor” whose analysis will “unlock value” and suggest ways of improving services? This is unlikely in more than a handful of cases.

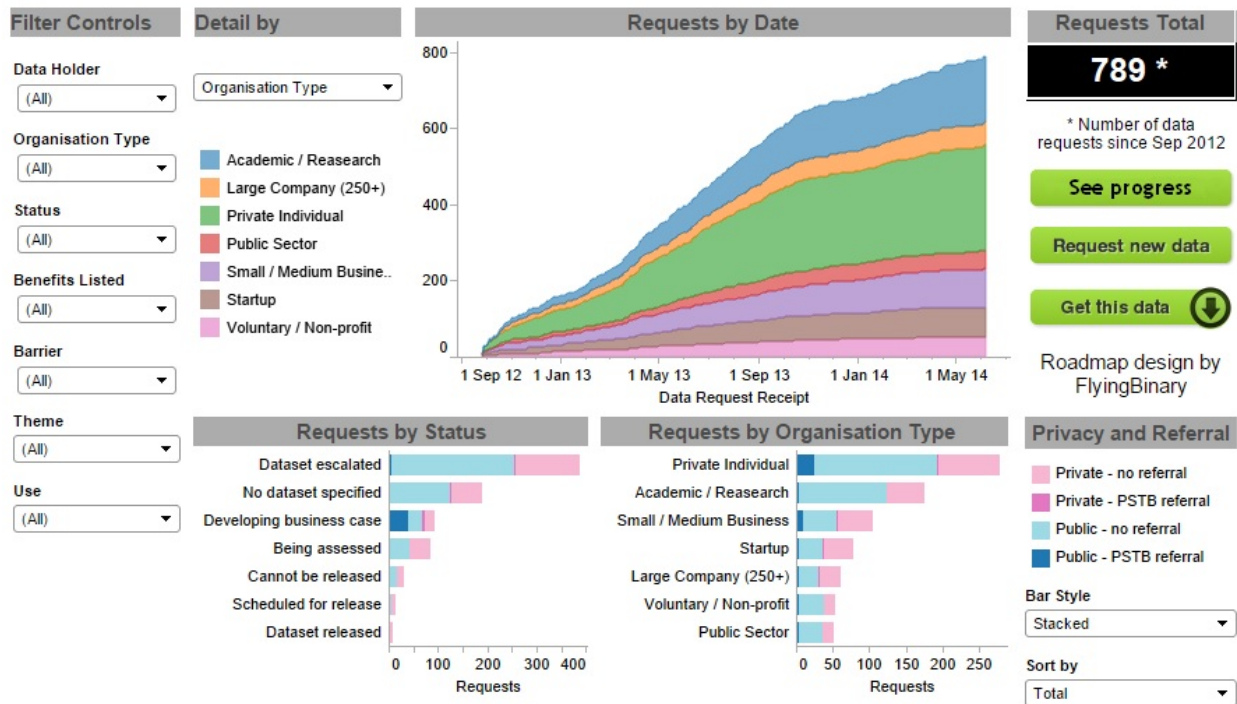
So far, the biggest community of end-users is made up of small and medium sized technology companies which build tools and applications to analyse or display data in a user-friendly way, and they provide this as a service, pro-bono, as a means of selling their application, or to sell the data in its repackaged form. This is making use of “passive” transparency, and they have little interest in pushing for more “active” transparency.

What do journalists need to do?

Journalists will always work with whatever is available, or can be made available by whatever means. Open Data is just one of those means. Freedom of Information requests are another. In a competitive market journalists have long preferred to keep their sources to themselves: openness is as new to them as it is to the publishers of data which until recently was kept away from public view.

There is currently a kind of inertia borne of equilibrium – Open Data initiatives are growing in number worldwide, but make little visible impact. From the journalists’ point of view datasets are not being published fast enough to be considered as timely sources for journalism, and the subjects covered are rarely of journalistic interest.

Journalists could press governments for more openness: where a dataset might form the basis of a news story, they are more likely to request it under their Freedom of Information Act than through the government data portal. In the UK, for example, the number of datasets requested is relatively small – 789 across the whole of government in 3 years. More importantly, neither “journalism” nor “media” is listed among the categories of requesting organisation types.



48

The Open Data movement needs publicity. It is rarely spoken about outside its own relatively specialised community. As we have seen journalists rarely describe their sources in enough detail to discern that data was involved in the writing of a story, and as we've seen, governments do not appear to consider journalists and the mass media important end-users. So there is a stand-off. Yet both sides need each other more than they admit.

As well as increasing the public profile of Open Data, and making better use of data in published stories, Open Data is only a relatively specialised component of any country's culture of openness.

The British case is a good example of a culture in need of change – the UK civil service has a natural tendency to keep things from the public, so it is not surprising that datasets are missing, or incomplete, or published so long after the fact as to be of little value to anyone. For example – Road Safety data, detailing 20 000+ incidents attended by the police in Great Britain every year is complete only up to 2013⁴⁹. Half way through 2015 the 2014 data has yet to be published.

For all the talk among Open Data-crats of a “fire hose” of data, there are still plenty of people trying to ensure that the water passing through the hose is of perfect quality, rather than let others do the purifying later. Since 100% accuracy is unlikely, in my view delaying publication in order to clean up data is at best a well-meaning act, and at worst a way of limiting the potential impact of published data.

⁴⁸ <http://data.gov.uk/odug-roadmap>

⁴⁹ <http://data.gov.uk/dataset/road-accidents-safety-data>

Much data is now generated in real time – instantaneously in the case of rapid transit passenger movements; less frequently, but nevertheless contemporaneously in the case of police records of accidents, arrests, “stop and search”; daily, weekly and monthly in the cases of UK health data (bed occupancy, ambulances waiting are daily; hospital visits, weekly; waiting times, drug prescribing are recorded monthly). Clearly - releasing all government data unchecked as it is created is often impossible or inadvisable, or both.

Where data is published only once a year, the responsible agency will check it for accuracy while waiting for the publication deadline. Other data such as UK prescribing data at doctors’ surgery level could, in theory, be published in real time. It is too large to be checked – made up of up to 4 million lines every month, the data is full of errors, but is still not published until it is 3-4 months old. Other data could affect markets or prices, and needs to be withheld and checked before release for commercial reasons. And then there are official statistics which are scrutinised and polished for public consumption; a process which takes time – and even so, is not immune from inaccuracy and/or criticism once published.

Journalists are just one group who rely on up-to-date information, and we have seen that they are not the intended end-users. Whoever the user, out of date data is of limited value, and may even temper the potential impact of research based on it. And, where the public are being encouraged to be “armchair auditors”, the more recent the data, the more likely they are to engage with it.

If the quantity, quality, and timeliness of Open Data is an indicator of government openness, then the transparency of sourcing of stories is its own indicator of the journalistic culture: even in a world in which much information is readily available to anybody 24 hours a day, the prevailing journalistic culture is to retain control over the way they quote their sources in such a way that “open source news” is still in its infancy, if not an alien concept. Some, notably The Guardian in the UK, made a practice of publishing links to the datasets on which its data-driven stories were based, but overall, the accepted tendency is not to list the sources of a story, or to “show one’s working” in arriving at an article’s conclusion.

Cutting out the middle man?

Ever since the advent of cameraphones, and the parallel development of Social Media, and of “Citizen Journalism”, there has been talk of “disintermediation” in news – the removal of journalists from the chain between events and consumers of news. So far this has not happened, except, arguably, in parts of the Social Media sphere. On the other hand, there have been precious few “armchair auditors” which governments were supposedly hoping to see poring over their accounts and suggesting efficiencies. “Holding power to account” still sounds like a role for the mass media. If journalists don’t make use of Open Data, who will?

As David Donald, formerly of Pro Publica and the Center for Public Integrity put it⁵⁰: “Open data has had a significant impact on journalism, but it doesn't stop there. It really has an impact on the functioning of democracy, from data journalists mining Open Data for

⁵⁰ Email exchange with author

government fraud, malfeasance or even incompetence to citizens hunting through the data themselves to keep their elected officials honest. I think gauging how open a government is with the people's data is one measure of how healthy that democracy is. For if it can allow investigative and data journalists to do their work, then all parts of that democracy can function more efficiently and successfully". Echoing several of the contributors to "Open Government", this is a somewhat rosy view of the US ecosystem, which, although advanced, and functioning well, has a way to go.

There is a lot still to do on the journalists' or analysts' side. Andy Cotgreave⁵¹, of the data visualisation software company, Tableau, tweeted four "data challenges" in June 2015:

1. Untrained analysts
2. Data reliability
3. Measuring the right thing
4. Unjustified data collection

Allowing for the brevity of a tweet's 140 character limit, this short list includes the key obstacles mentioned above. The 4th challenge is a question of ensuring a good public image for data collection in the post-Snowden landscape where many citizens remain wary of the extent to which their own data are being collected and how they are used. This is outside the scope of this paper, but it is worth noting that "big" and "open" data are referred to interchangeably, and in ordinary people's minds, may well be mixed with the kind of big data Snowden's material indicated governments were harvesting.

Conclusions

Datajournalism has grown in popularity rapidly but is likely to remain a minority specialism within journalism as long as the data available remains patchy, of dubious relevance to the general public and/or out of date. This leads to a kind of "push-me-pull-you" effect: journalists will not feel the need to make greater use of Open Data until they see it as a rich seam of material, while data policy-makers won't feel the need to improve the stream of data until they come under pressure to do so.

The publication of Open Data – with myriad declarations of openness from governments around the world – answers a question which few people had asked. There is no army of armchair auditors poring over published data: this will not change until calls for openness grow, and governments come under pressure to publish data which is useful to a sizeable element of the population.

Next Steps

Open Data is here to stay – in the sense that few governments would want to announce that they were reducing their openness. At the moment, the process is top down, with the public largely passive. Journalists need to embrace it much more as a tool, and understand how they can make use of it, making the transition to fuller openness much more bottom up.

⁵¹ <https://twitter.com/acotgreave> (tweet of 29/6/15)

Journalists should be making more noise about the data they can use, and what they cannot, and should be leading calls for greater openness. Audiences should be becoming more aware of their government's record of openness and exposing shortfalls in Open Data –its quality, timeliness, and relevance. Governments should accept that journalists are natural end-users of Open Data, and be working with them explicitly to make datasets more complete, more up-to-date, and more useful and relevant

Journalists who use data in their reporting should be publishing more links to that data, and showing their working.

Governments wanting to be seen as truly open should be working to help their populations understand what data can do for them, and what they can do with data, rather than leaving it to the coders, developers to remain commercially-minded intermediaries between data and end-user. This means involving the media in the release of data, in publicising its existence, its potential uses, its successes and its failures.

Bibliography

2015. *2013 Pulitzer Prize winners - Public Service*. June. <http://www.pulitzer.org/works/2013-Public-Service>.
2015. *2015 Investigative Reporting Winners - New York Times - Eric Lipton*. June. <http://www.pulitzer.org/works/2015-Investigative-Reporting-Group1>.
2015. *2015 Pulitzer Prize Winners, Investigative Reporting - Wall Street Journal Staff*. June. <http://www.pulitzer.org/works/2015-Investigative-Reporting-Group2>.
- Andrews, Kehinde. 2015. "Comment is Free - Stop and search is a disgrace across the UK." *The Guardian*. 7 August. Accessed September 2, 2015. <http://www.theguardian.com/commentisfree/2015/aug/07/stop-and-search-uk-rural-black-people-police>.
- BBC News. 2011. "BBC News." *Every death on every road in Great Britain 1999-2010*. 2 December. Accessed September 2, 2015. <http://www.bbc.co.uk/news/uk-15975720>.
2015. *British Journalism Awards*. May. <http://www.pressgazette.co.uk/awardshome>.
- Campbell, Ronald. 2013. "Orange County Register." *The economics of the fading middle class*. 20 September. Accessed September 2, 2015. <http://www.ocregister.com/articles/middle-523465-class-california.html>.
- Center for Investigative Reporting. 2014. "Pay Day California." *Pay Day California*. 4 December. Accessed September 2, 2015. <http://payday.cironline.org/>.
- Christopher Snowbeck, MaryJo Webster. 2014. "Nation & World." *Twin Cities.Com*. 1 March. Accessed September 2, 2015. http://www.twincities.com/nation/ci_25247140/where-insurance-premiums-are-highest-new-health-laws.
- Conservative Party. 2015. *Invitation to Join the Government of Britain (2010 Election Manifesto)*. June. <https://www.conservatives.com/~media/files/activist%20centre/press%20and%20policy/m-anifestos/manifesto2010>.
- . 2010. *PM's podcast on transparency*. June. <https://www.gov.uk/government/news/pms-podcast-on-transparency>.
- Cotgreave, Andy. 2015. *Tweet by Andy Cotgreave*. 29 June. <https://twitter.com/acotgreave>.
- Dan Beekman, Rocco Parascandola. 2014. "NYPD: Off-duty cops kills mugger in Brooklyn." *New York Daily News*. 8 December. Accessed September 2, 2015. <http://www.nydailynews.com/news/crime/off-duty-shoots-kills-crook-brooklyn-police-article-1.1013958>.
- Delaware Online. 2013. "Delaware's Dangerous Intersections." *Delaware Online*. 15 July. Accessed September 2, 2015. <http://www.delawareonline.com/article/20130714/NEWS1501/307140036/Delaware-s-dangerous-intersections>.

2015. *Dollars for Docs*. June. <https://projects.propublica.org/docdollars/>.
- Elbadawi, Ibrahim. 2015. *ODI news*. 4 January. <http://theodi.org/blog/a-sneak-peek-at-the-arab-regions-open-data>.
2015. *Follow the Money*. June. <http://www.followthemoney.org/>.
- Hargreaves, Nick. 2013. "The Missing Link in Kenya's Open Data Story." *Code4Kenya*. 16 August. <http://www.code4kenya.org/?author=2>.
- Higgerson, David. n.d. *INFOGRAPHIC: HOW TO CREATE GREAT JOURNALISM ONLINE*. <https://davidhiggerson.wordpress.com/2015/05/25/infographic-how-to-create-great-journalism-online/>.
- Howard, Alexander. 2015. *With 'Release To One, Release to All,' U.S. Pilots New Freedom Of Information Policy*. 13 July. http://www.huffingtonpost.com/entry/with-release-to-one-release-to-all-us-pilots-new-freedom-of-information-policy_559f1661e4b05b1d0290135b?
2015. *Impact [of open data]*. June. <http://www.data.gov/impact/>.
2015. *Investigative Reporting Prize 2015*. June. <http://www.pulitzer.org/citation/2015-Investigative-Reporting>.
- Jees, Charlotte. 2015. "Open data in the UK: a job half done." *Computer World UK*. 14 August. <http://www.computerworlduk.com/data/open-data-in-uk-job-half-done-3622607/>.
2015. "Kenya Open Data portal." June. <https://opendata.go.ke/>.
- Kovach, Bill & Rosenstiel, Tom. 2001. *The Elements of Journalism*. New York, NY: Three Rivers Press.
- Lathrop, Daniel and Ruma, Laura. 2010. *Open Government*. Sebastopol: O'Reilly Media inc.
2015. *MapLight*. May. <http://maplight.org/>.
- Meyer, Philip. 2002. *Precision Journalism*. Lanham: Rowman & Littlefield.
2015. *Monthly Price Paid Data*. July. <http://data.gov.uk/dataset/land-registry-monthly-price-paid-data>.
- Mullin, Benjamin. n.d. *Poynter Media Wire*. <http://www.poynter.org/news/mediawire/356961/the-government-is-now-giving-away-your-foia-scoops/>.
- Newman, Levy. 2015. *Digital News Report 2014*. June. <http://www.digitalnewsreport.org/survey/2014/>.
- Nic Newman, David Levy, Rasmus Kleis Nielsen. 2015. *Digital News Report 2015*. July. <http://www.digitalnewsreport.org/>.
2015. *ODI News*. July. <https://theodi.org/news/burkina-faso-launches-open-data-initiative-with-mentoring-from-the-odi-and-funding-from-the-world-bank>.
2015. *Open Data - Forward Strategy*. June. <http://talkaboutlocal.org.uk/open-data-forward-strategy/>.
2015. *Open Data Barometer*. August. <http://www.opendatabarometer.org/report/analysis/readiness.html>.

Open Data Barometer. 2015. *Open Data Barometer*. May. <http://barometer.opendataresearch.org>.

2015. *Open Data Burkina Faso - page d'accueil*. June. <http://data.gov.bf/>.

2015. *Open Data Request Roadmap*. June. <http://data.gov.uk/odug-roadmap>.

2015. *Open Definition*. August. <http://opendefinition.org/od/>.

Open Knowledge Foundation. 2015. *Global Open Data Index*. May. <http://index.okfn.org/place/>.

Palm Beach Post. 2015. "Police Shooting Probe - deputies shielded by badge." *My Palm Beach Post*. April. Accessed September 2, 2015. <http://www.mypalmbeachpost.com/policeshootings/>.

2015. *Pro Publica - Journalism in the Public Interest*. June. <https://www.propublica.org/>.

2015. *Programs page*. June. <http://www.mediacouncil.or.ke/en/mck/index.php/programs/media-excellence-awards>.

Rogers, Simon. 2011. *Facts are Sacred*. London: Guardian Books.

Sidwaya. 2015. "Societe." *Fasopresse*. May. <http://www.fasopresse.net/societe/4288--soldela-realite-de-la-masse-salariale-du-personnel-de-letat>.

2015. *Site Usage by publisher*. June. <http://data.gov.uk/data/report/site-usage-publisher?metric=downloads>.

2015. *The Counted - People killed by the police in the US*. June. <http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database>.

2015. *The Philip Meyer Award*. June. <http://ire.org/awards/philip-meyer-awards/>.

2015. *The Prescribers*. June. <http://www.propublica.org/series/prescribers>.

Transparency International. 2015. *Corruption Perceptions Index 2014*. June. <http://www.transparency.org/cpi2014>.

UK Department for Transport. 2015. *Road Safety Data*. June. Accessed July 13, 2015. <http://data.gov.uk/dataset/road-accidents-safety-data>.

—. 2015. *Road Safety Data*. July. <http://data.gov.uk/dataset/road-accidents-safety-data>.

2015. *Unless we change the way we think about transparency, open data is unlikely to have a significant political impact at the local level*. June. <http://www.democraticaudit.com/?p=13606>.

Bibliography

2015. *2013 Pulitzer Prize winners - Public Service*. June. <http://www.pulitzer.org/works/2013-Public-Service>.
2015. *2015 Investigative Reporting Winners - New York Times - Eric Lipton*. June. <http://www.pulitzer.org/works/2015-Investigative-Reporting-Group1>.
2015. *2015 Pulitzer Prize Winners, Investigative Reporting - Wall Street Journal Staff*. June. <http://www.pulitzer.org/works/2015-Investigative-Reporting-Group2>.
- Andrews, Kehinde. 2015. "Comment is Free - Stop and search is a disgrace across the UK." *The Guardian*. 7 August. Accessed September 2, 2015. <http://www.theguardian.com/commentisfree/2015/aug/07/stop-and-search-uk-rural-black-people-police>.
- BBC News. 2011. "BBC News." *Every death on every road in Great Britain 1999-2010*. 2 December. Accessed September 2, 2015. <http://www.bbc.co.uk/news/uk-15975720>.
2015. *British Journalism Awards*. May. <http://www.pressgazette.co.uk/awardshome>.
- Campbell, Ronald. 2013. "Orange County Register." *The economics of the fading middle class*. 20 September. Accessed September 2, 2015. <http://www.ocregister.com/articles/middle-523465-class-california.html>.
- Center for Investigative Reporting. 2014. "Pay Day California." *Pay Day California*. 4 December. Accessed September 2, 2015. <http://payday.cironline.org/>.
- Christopher Snowbeck, MaryJo Webster. 2014. "Nation & World." *Twin Cities.Com*. 1 March. Accessed September 2, 2015. http://www.twincities.com/nation/ci_25247140/where-insurance-premiums-are-highest-new-health-laws.
- Conservative Party. 2015. *Invitation to Join the Government of Britain (2010 Election Manifesto)*. June. <https://www.conservatives.com/~media/files/activist%20centre/press%20and%20policy/manifestos/manifesto2010>.
- . 2010. *PM's podcast on transparency*. June. <https://www.gov.uk/government/news/pms-podcast-on-transparency>.
- Cotgreave, Andy. 2015. *Tweet by Andy Cotgreave*. 29 June. <https://twitter.com/acotgreave>.
- Dan Beekman, Rocco Parascandola. 2014. "NYPD: Off-duty cops kills mugger in Brooklyn." *New York Daily News*. 8 December. Accessed September 2, 2015. <http://www.nydailynews.com/news/crime/off-duty-shoots-kills-crook-brooklyn-police-article-1.1013958>.
- Delaware Online. 2013. "Delaware's Dangerous Intersections." *Delaware Online*. 15 July. Accessed September 2, 2015. <http://www.delawareonline.com/article/20130714/NEWS1501/307140036/Delaware-s-dangerous-intersections>.
2015. *Dollars for Docs*. June. <https://projects.propublica.org/docdollars/>.

Elbadawi, Ibrahim. 2015. *ODI news*. 4 January. <http://theodi.org/blog/a-sneak-peek-at-the-arab-regions-open-data>.

2015. *Follow the Money*. June. <http://www.followthemoney.org/>.

Hargreaves, Nick. 2013. "The Missing Link in Kenya's Open Data Story." *Code4Kenya*. 16 August. <http://www.code4kenya.org/?author=2>.

Higgerson, David. n.d. *INFOGRAPHIC: HOW TO CREATE GREAT JOURNALISM ONLINE*. <https://davidhiggerson.wordpress.com/2015/05/25/infographic-how-to-create-great-journalism-online/>.

Howard, Alexander. 2015. *With 'Release To One, Release to All,' U.S. Pilots New Freedom Of Information Policy*. 13 July. http://www.huffingtonpost.com/entry/with-release-to-one-release-to-all-us-pilots-new-freedom-of-information-policy_559f1661e4b05b1d0290135b?

2015. *Impact [of open data]*. June. <http://www.data.gov/impact/>.

2015. *Investigative Reporting Prize 2015*. June. <http://www.pulitzer.org/citation/2015-Investigative-Reporting>.

Jees, Charlotte. 2015. "Open data in the UK: a job half done." *Computer World UK*. 14 August. <http://www.computerworlduk.com/data/open-data-in-uk-job-half-done-3622607/>.

2015. "Kenya Open Data portal." June. <https://opendata.go.ke/>.

Kovach, Bill & Rosenstiel, Tom. 2001. *The Elements of Journalism*. New York, NY: Three Rivers Press.

Lathrop, Daniel and Ruma, Laura. 2010. *Open Government*. Sebastopol: O'Reilly Media inc.

2015. *MapLight*. May. <http://maplight.org/>.

Meyer, Philip. 2002. *Precision Journalism*. Lanham: Rowman & Littlefield.

2015. *Monthly Price Paid Data*. July. <http://data.gov.uk/dataset/land-registry-monthly-price-paid-data>.

Mullin, Benjamin. n.d. *Poynter Media Wire*. <http://www.poynter.org/news/mediawire/356961/the-government-is-now-giving-away-your-foia-scoops/>.

Newman, Levy. 2015. *Digital News Report 2014*. June. <http://www.digitalnewsreport.org/survey/2014/>.

Nic Newman, David Levy, Rasmus Kleis Nielsen. 2015. *Digital News Report 2015*. July. <http://www.digitalnewsreport.org/>.

2015. *ODI News*. July. <https://theodi.org/news/burkina-faso-launches-open-data-initiative-with-mentoring-from-the-odi-and-funding-from-the-world-bank>.

2015. *Open Data - Forward Strategy*. June. <http://talkaboutlocal.org.uk/open-data-forward-strategy/>.

2015. *Open Data Barometer*. August. <http://www.opendatabarometer.org/report/analysis/readiness.html>.

Open Data Barometer. 2015. *Open Data Barometer*. May. <http://barometer.opendataresearch.org>.

2015. *Open Data Burkina Faso - page d'accueil*. June. <http://data.gov.bf/>.
2015. *Open Data Request Roadmap*. June. <http://data.gov.uk/odug-roadmap>.
2015. *Open Definition*. August. <http://opendefinition.org/od/>.
- Open Knowledge Foundation. 2015. *Global Open Data Index*. May. <http://index.okfn.org/place/>.
- Palm Beach Post. 2015. "Police Shooting Probe - deputies shielded by badge." *My Palm Beach Post*. April. Accessed September 2, 2015. <http://www.mypalmbeachpost.com/policeshootings/>.
2015. *Pro Publica - Journalism in the Public Interest*. June. <https://www.propublica.org/>.
2015. *Programs page*. June. <http://www.mediacouncil.or.ke/en/mck/index.php/programs/media-excellence-awards>.
- Rogers, Simon. 2011. *Facts are Sacred*. London: Guardian Books.
- Sidwaya. 2015. "Societe." *Fasopresse*. May. <http://www.fasopresse.net/societe/4288--soldela-realite-de-la-masse-salariale-du-personnel-de-letat>.
2015. *Site Usage by publisher*. June. <http://data.gov.uk/data/report/site-usage-publisher?metric=downloads>.
2015. *The Counted - People killed by the police in the US*. June. <http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database>.
2015. *The Philip Meyer Award*. June. <http://ire.org/awards/philip-meyer-awards/>.
2015. *The Prescribers*. June. <http://www.propublica.org/series/prescribers>.
- Transparency International. 2015. *Corruption Perceptions Index 2014*. June. <http://www.transparency.org/cpi2014>.
- UK Department for Transport. 2015. *Road Safety Data*. June. Accessed July 13, 2015. <http://data.gov.uk/dataset/road-accidents-safety-data>.
- . 2015. *Road Safety Data*. July. <http://data.gov.uk/dataset/road-accidents-safety-data>.
2015. *Unless we change the way we think about transparency, open data is unlikely to have a significant political impact at the local level*. June. <http://www.democraticaudit.com/?p=13606>.