

# Section 5:

## Data journalism goes international

**Damian Radcliffe**

For journalists and academics, it's easy to find great examples of data journalism from the leading English speaking media markets of the USA and the UK. Yet, data journalism is a global phenomenon, and this journalistic reality can all too often be overlooked. In this section we begin to redress this, by hearing from a number of data journalists based outside of these primary markets. What's discernible from their contributions is the universality of several core challenges faced by data journalists. This includes: the data literacy of audiences, cultural change within large new organisations and the format of public data sources, irrespective of where they live and work.

These case studies also highlight how data journalists are driven by the same motivations as traditional journalists. Reporting using data is often underpinned by the rudimentary journalistic principles of seeking to hold authority to account, to tell stories that shine a different light on our world, and to support the creation of an informed citizenry.

A number of these characteristics can be seen in our first contribution, which explores immigration. It's a hot topic in many countries, and one where data journalists can play a pivotal role in discerning the differences between public perception and on-the-ground reality. The Italian journalist Matteo Moretti was part of a team who addressed this subject in the North Italian city of Bolzano. Their approach, which is easily replicable, offers a potential model for data journalists around the world who are interested in exploring this topic. The 'People's Republic of Bolzano' project used visualisations, social media and a

multi-disciplinary approach to discern if the area really was witnessing a Chinese ‘invasion’. Analysis of the data ‘painted a very clear picture: Chinese in Bozen-Bolzano were so few and their activities so fragmented across the whole city that an “invasion” was hard to imagine.’

Nonetheless, such perceptions continued to abound, and so the team undertook to ‘try to shift the perception of the Chinese among locals in the city’ by working with a cultural anthropologist ‘to tell the story of the Chinese community in the city via a series of personal narratives, something data alone could not do.’

Following this highly localised approach, Kathryn Hayes takes the temperature of data journalism on a country-wide level, examining how this field is evolving in Ireland. Featuring insights from qualitative interviews with journalism academics and journalist practitioners, she reveals ‘that while the landscape of data journalism is improving in Ireland, access to suitably formatted, open and linked data remains a struggle’.

Barriers to progress include ‘the unremitting reluctance of some government departments to release information’ and the need for appropriate training for data-hesitant journalists. There are, however, reasons for optimism, ranging from a newly strengthened Freedom of Information (FOI) Act through to a national track record of ‘a number of political and social controversies [which] have been unearthed through original and investigative journalism.’ These foundations often lay the cornerstone for data journalism to flourish, so it will be interesting to see if data journalism in the Emerald Isle is able to build on these fundamentals.

Eva Constantaras, who advises the NGO Internews on data journalism, offers a different perspective, building on her extensive experience of launching data journalism initiatives in developing countries such as Kenya, Afghanistan, China, Myanmar, El Salvador and Mexico.

Although clearly an advocate for this emerging practice, she notes that data journalism – globally – has yet to fulfil its potential. The reasons for this are multiple, and in developing markets they include apathy, social media buzz being mistaken for real change and impact, as well as ‘the assumption’ often by funders and Western media ‘of the role of the media itself as a public service watchdog’. Journalistic realities in these markets are often very different, she writes:

*‘Actually getting to the bottom of things and identifying a chain of responsibility can be much riskier than exposing corruption in general terms in countries where corruption, graft, mismanagement, incompetence are all expected and scratching too deep can be dangerous. Therefore, the motivation to exercise the kind of analytical thinking skills required for the kind of data journalism that helps solves problems is often missing.’*

Proponents of data journalism in developing countries therefore need to take account of these realities and consider this when funding and supporting new initiatives. To help them do this, Eva outlines how to overcome the ‘fatal design

flaws' that she identifies with the current boot camp model and the 'huge potential' afforded by global data investigations. 'It's no surprise that so many initiatives fail,' she says, 'but it is all the more important that we identify elements that work, build a new model for data journalism in developing countries and support rising stars.' After reading her contribution, you'll be inclined to agree with her.

One developing country with a growing myriad of data initiatives is India. Sanjit Oberai, data editor of Quintillion Media, walks us through a number of these enterprises, highlighting the importance of mobile in this market and aspirations of the county's wider 'Digital India' initiative. John Samuel Raja, of How India Lives, tells us that:

*'..data journalism that involves working with raw data and converting that data into interesting analysis hasn't picked up. That's because newsrooms don't have the resources or the required training skills to execute this kind of journalism.'*

However, he believes that there is 'immense scope' in this arena. 'We believe the way forward is to build communities and use them to gather data,' he says.

This optimism – although not necessarily the approach – is shared by many of his contemporaries in India's data journalism space. For this potential to be unlocked, barriers to use such as data cleaning and finding 'valid and up-to-date data' need to be overcome. The importance of this in India, as elsewhere, is increasingly important. As Factly's Rakesh Dubbudu explains:

*'Misinformation and rumour also frequently trend on social media, but little of it is substantiated. Wrong data/information or a morphed picture that might be sensational has a greater chance of being shared by many people on social media than genuine information. This was the motivation behind Factly, to make public data more meaningful to the public and encourage them to look for facts or genuine data.'*

It's an approach that data journalists the world over will empathise with.

# Reporting the ‘invasion’: perception and reality of Chinese migrants in Bolzano, Italy

**Matteo Moretti recounts how he and colleagues used data journalism to tackle Chinese migrant xenophobia in the Italian city of Bolzano**

## **Introduction**

The ‘People’s Republic of Bolzano’ is a visual journalism project that tried to debunk the cliché of a Chinese invasion in Bolzano, a south Tyrolean city in Italy, populated 25 per cent by German speakers and 75 per cent by Italian speakers. Due to historical and political reasons rooted as far back as 1918, when south Tyrol was annexed by Italy, the history of Bolzano and how the Italian and German communities live together, is a very long and not easy to tell story. In my personal experience as a lecturer and then a researcher at the Free University of Bozen-Bolzano since 2010, I found the local culture very sensitive to the concept of ‘invasion’ because of historical occupation by Italian troops. It is only in the last 20 years that the two communities learned to co-exist together, and migrants played a crucial role in this integration process. This change in dynamics raised new questions to answer.

This premise helps to explain why, despite the fact that a Chinatown doesn’t exist in Bozen-Bolzano, and Chinese make up just 0.6 per cent of the population - 633 of 105,713 citizens - (Astat, 2014) – some locals expressed fears of an ‘invasion’ of Chinese people in the city. A number of factors also contributed the feeling of an invasion. First of all, Chinese people work in public spaces – in restaurants, bars, shops and hairdressers; they are more visible in the city than other migrant communities that usually work away from public view in construction sites or in the apple fields that surround the city - South Tyrol is a well-known apple-producing region, producing 950,000 tonnes each season, equivalent to 50 per cent of the entire Italian apple harvest annually, and much of this work is done by migrants (Südtirol, ND). Second, the most read newspaper by Italian people in Bozen-Bolzano, the *Alto Adige*, stirred up migrant worries, running articles with headlines warning of the creation of a ‘Chinatown’

ghetto (Conti, 2011); of a 'Chinese advance' (*Alto Agide*, 2012), and 'Chinese megastores' (Valletti, 2012) depicting a context very different from the reality, most frequently in years of 'big' Chinese business openings between 2007 and 2011. In 2010 the city saw 19 Chinese businesses open. For a small city such as Bozen-Bolzano this is significant, as much from a social perception perspective as economic.

### 1996 - 2013

#### Opening of Chinese businesses in Bolzano

(Chamber of Commerce Bolzano 2014)

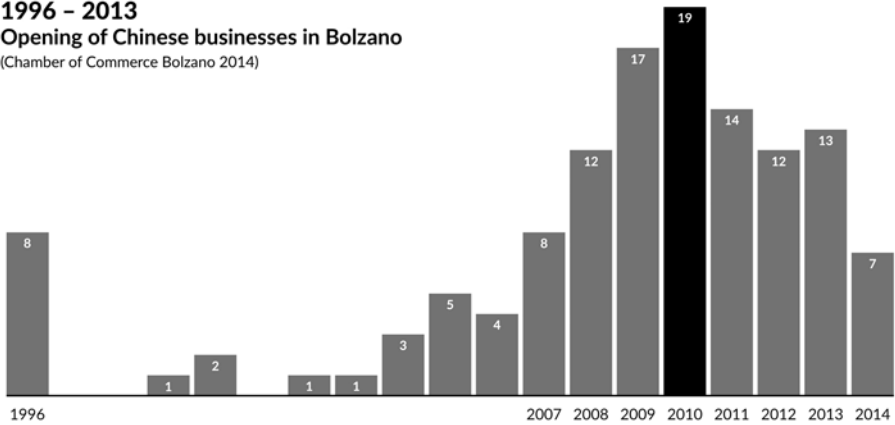


Figure 64: Chart demonstrates the opening of Chinese run businesses in the city between 1996 and 2014

### Reporting reality versus perception

All the factors above contributed the cliché of the 'invasion'. While there were issues in other Italian cities with migrants, the reality in Bozen-Bolzano would have been best described as integration instead (Chinese in Italy number 1.8 per cent of the population, while in Bozen-Bolzano the population was 0.6 per cent, as described above (Astat, 2014)).

With journalist Fabio Gobbato, we decided to investigate. We began experimenting with new ways to tell stories with data in order to open a window into the Chinese community of Bozen-Bolzano; to show who the Chinese living in the city were; what they thought of Italy; and moreover that the story of 'invasion' was far from reality. We started looking at the data: we asked Atstat – the Italian statistics agency – for details of the Chinese population of Bozen-Bolzano. The Chinese government doesn't allow dual citizenship, meaning Chinese migrants have to choose between their mother or destination country citizenship. Often, as is the case in Bozen-Bolzano, a majority keep their Chinese citizenship in order to spend their retirement in China.

The 633 Chinese we considered includes only citizens with a Chinese passport and not second generation Chinese-Italians, or Chinese that have taken Italian citizenship (though these are few in number). We didn't include them in our data analysis because from a legal point of view, they are Italian. As previously stated Chinese residents number just 633 out of 105,713 - 0.6 per cent of the city's

population. If we look at the population growth trend (see chart below) we notice an increase in 2007-2008 but the increase in the Chinese population in the city is very small; even newborns are few in number - just 16 in 2013. It was clear based on the statistics that fears of an 'invasion' were unfounded.

### **Economic 'invasion'**

We then analysed economic statistics provided by the local Chamber of Commerce to examine if an economic 'invasion' had happened in the city. We searched for businesses opened by Chinese citizens for the previous 20 years. We found that out of the 9,543 companies in Bozen-Bolzano, only 126 (1.3 per cent) were Chinese-run. The data from 1996 to 2013 showed us that the businesses were located all over the city and not in any particular spot; and that the type of Chinese-owned businesses had changed during the period (from bars and restaurants to a huge variety of activities such as shops and hairdressers). The data again refuted the idea of a Chinese invasion or a 'Chinatown ghetto'. We discovered some interesting statistics: out of 441 bars, 51 (11.5 per cent) were managed by people with Chinese nationality; of 261 restaurants, 32 (11.8 per cent) were Chinese, the 11.8 per cent of the total. Evidence certainly that Chinese nationals who have settled in the city are entrepreneurial in the service industry, but hardly evidence of much else.

We visualised the economic data on an interactive map, using a stop motion animation that displays the evolution of economic activity since 1996. We also split each timeline year unit into smaller coloured bars in order to visualise the proportion of different business activities opened in each year: in earlier years businesses concentrated around bars and restaurants, however since 2009 there has been a diverse range of businesses opened. Through the data map, we were also able to demonstrate that a Chinatown district didn't exist – as the map clearly showed businesses owned by Chinese nationals were spread throughout the city – despite reports by local newspapers.



Figure 65: Map demonstrating spread of Chinese-run businesses in the city

### **Shifts in perception**

The data painted a very clear picture: Chinese in Bozen-Bolzano were so few and their activities so fragmented across the whole city that an ‘invasion’ was hard to imagine. But for us it was very important not only to demonstrate this but also try to shift the perception of the Chinese among locals in the city. Too many clichés has been allowed to develop, and misinformation had contributed to the creation of a distorted perception.

To accomplish that, we decided to combine qualitative and quantitative investigative methods and took an integrated approach to the story. We met, and then included, Sara Trevisiol in our team, a cultural anthropologist with a strong knowledge of Chinese culture and deeply connected to the Chinese community of Bozen-Bolzano. With her, we shot eight qualitative interviews. This human-interest approach allowed us to tell the story of the Chinese community in the city via a series of personal narratives, something data alone could not do. The series also demonstrated how the Chinese community of Bolzano is composed of a huge variation of members, contrasting the perception that Chinese only own restaurants. Thanks to her work, we were able to represent how multi-faceted the Chinese community members of Bozen-Bolzano are. Through their voices, we were able to break through a number of misconceptions of Chinese people living in the city.

For example, one of our interviewees, YiYi Chen, discussed the concept of *Guanxi*, a traditional personal network of influence that allow Chinese people to save large amounts of cash in order to fund their businesses, rather than relying on banks for finance (Gold et al, 2002). This interview also allowed us to debunk the myth (very common in Italy) that Chinese businesses were supported by the Chinese mafia. Another interviewee, Yingjun, discussed the Italian education system, which she described as easier than the Chinese system. She arrived in Italy when she was 14 and managed to learn fluent Italian as well as graduate high school with honours alongside other Italian students her age within five years. These are just two of the eight stories that we collected in our efforts to tell the story of the Chinese people of Bozen-Bolzano. We believe these demonstrated that far from an invasion, rather we were witnessing Chinese migrants integrating. Thanks to the anthropologist’s work and approach, we accessed qualitative data that may have been difficult to collect otherwise.

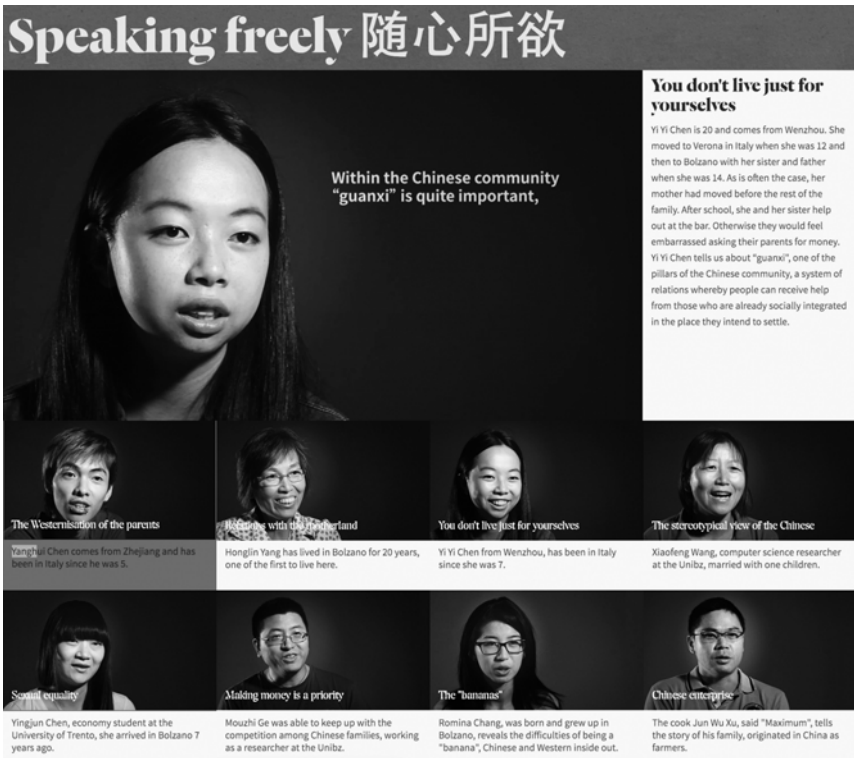
### **Telling the story from different sides**

The interdisciplinary team, composed of two designers, a journalist, a cultural anthropologist and a computer scientist, told the story from their respective points of view, revealing different sides of the same phenomenon. I found inspiration in the work of Otto Neurath, an Austrian philosopher, sociologist and political economist, who invented ISOTYPE - a universal language to educate people after the World War I. Neurath, was also foremost in the shift of thinking about data visualisation as purely a form of analytics toward thinking of visualisations as both educational and informative, through the work of interdisciplinary teams combining experts, graphic designers and a ‘transformer’ - a role that he defined as ‘responsible for organising the information in visual

*Reporting the 'invasion': perception and reality of Chinese migrants in Bolzano, Italy*

terms so that it could be understood easily' (Twyman, 1975: 7-18), a sort of visual and cultural mediator. This interdisciplinary approach was useful in our long-form investigation using data, where the journalist told the story; the designers visualised it in order to make it more appealing; the computer scientist created the interactions to engage the users in the navigation; and the qualitative interviews enriched the narrative.

The project took six months of work. We worked in a different way from the traditional (linear) approach where usually the experts provide the data, then the journalists write the story and finally the designer packages it. We constantly shared all information among the group in order to ensure everyone was up to date, and we worked simultaneously on the project, from each respective point of view. We could define this approach as 'organic', where a change in a single variable provoked the re-design of the whole story. We had a number of back and forth processes; we re-thought and re-designed part of the story according to the changes or as new data was found. It was a time consuming approach, most probably time that a newsroom couldn't afford, but it was worth it. Most of the six months spent on the project were spent thinking and re-thinking the story in order to develop the narrative, based on a complex and serious social issue.



Figures 66 and 67: Screenshots from qualitative video interviews with Chinese residents



The project, which can be seen at [www.peoplesrepublicofbolzano.com](http://www.peoplesrepublicofbolzano.com), was presented on 26 September 2014 at the 'Long Night of Research 2014' at the Free University of Bozen-Bolzano (2014), and was also published as a double page spread in the local edition of the national newspaper *Corriere Della Sera*. In order to change attitudes of the local community toward Chinese integration, we also opened our work up to public debate - especially among readers of the *Alto Adige* (the most read Italian newspaper in Bozen-Bolzano), in order to show them the difference between their perceptions of the issue and the reality.

After our work was published it gained a lot of traction on social media via our Facebook page, though little local coverage. However four months later the story was picked up by *Der Spiegel*. The Italian correspondent of the German news group spotted the story on Facebook and wrote an article for the magazine, which also appeared online. Due to the minority German-speaking population of the Bozen-Bolzano, and the fact that local media are very responsive to what happens in Germany and in the German media - in particular when the story concerns the city - our story began to get local traction. The week after publication in *Der Spiegel* we were interviewed by local television and a local newspaper, and even the *Alto Adige* newspaper published our work in print, on their website and - very important for measuring the debate - on their Facebook page.

**Project timeline 2014 – 2015**

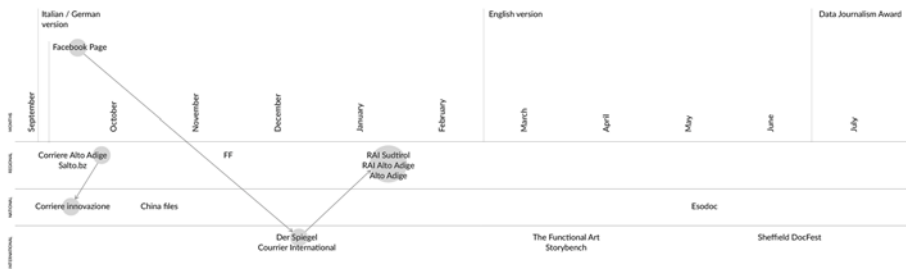


Figure 68: A timeline of the website interactions by the local, national and international media

We had observed that when the *Alto Adige* Facebook followers debated migrant issues in the past they usually they split in two, with a majority expressing anti-emigrant sentiment. In our case we compared two *Alto Adige* Facebook posts on the Chinese topic, the first post was about the article 'too much permission given to Chinese retailers, according to Italian traders' (*Alto Agide*, 2011) while the second was the one about our project, with the title 'There's no invasion: The Chinese are just 0.6 per cent' (*Alto Agide*, 2015).

We analysed sentiments expressed in the comments of both posts, in order to map and then compare the opinions. The results were interesting: in the first case, followers were split evenly in their views, with both negative and positive

sentiment expressed toward Chinese; while in the second case (the one relevant to our project) we noted that most comments were positive.

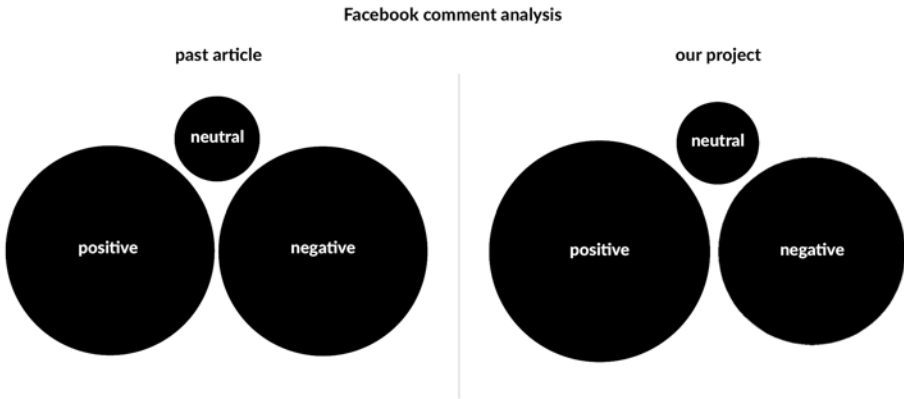


Figure 69: Comparison of the sentiment analysis of the post comments on Facebook

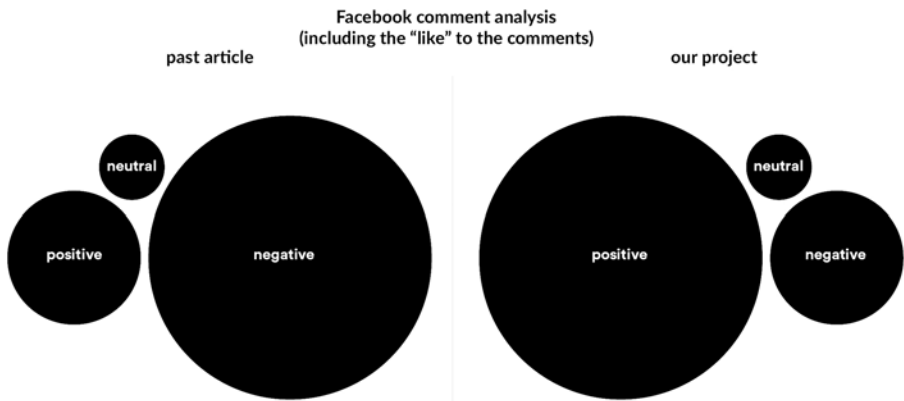


Figure 70: Comparison of the sentiment analysis of the post comments including 'like' on social media

Of course there are many caveats. Some time had passed between the two posts – the first in 2011, while the second was in 2014. Society had changed and grown, even if nationalist ideologies against migrants still persisted. Despite this, the positive comments expressed about Chinese integration was significant and cannot be ignored. We are happy to have contributed to this change in attitude toward Chinese migrants in the local community of Bolzano.

## References

- Alto Adige (2011) 'Protestano i negozianti di Bolzano: troppi permessi ai cinesi' via Facebook post, 14 June, available at <https://www.facebook.com/AltoAdigequotidiano/posts/209142142455113>, accessed 28 August 2015
- Alto Adige (2015) 'Macché invasione, in città i cinesi sono solo lo 0,6% della popolazione' via Facebook post, 8 January, available at <https://www.facebook.com/AltoAdigequotidiano/posts/10155024420955542>, accessed 28 August 2015
- Alto Adige* newspaper, available at <http://altoadige.gelocal.it/>, accessed 28 August 2015
- Conti, Alan (2011) 'Bar cinesi verso quota 150 La Chinatown bolzanina è in via Claudia Augusta', in *Alto Adige*, 13 March, available at <http://altoadige.gelocal.it/bolzano/cronaca/2011/03/13/news/bar-cinesi-verso-quota-150-la-chinatown-bolzanina-e-in-via-claudia-augusta-1.4214772>, accessed 28 August 2015
- Gold, Thomas, Douglas Guthrie, and David Wank (2002) *Social Connections in China: Institutions, Culture and the Changing Nature of Guanxi*. Cambridge: Cambridge University Press
- Istituto Provinciale di Statistica (Astat) (2014) available at <http://www.provincia.bz.it/astat/it>, accessed 28 August 2015
- Südtirol (ND) 'Eden's apple orchard' available at [www.southtyroleapple.com/en/our-territory/cultivation-area.html](http://www.southtyroleapple.com/en/our-territory/cultivation-area.html), accessed 28 August 2015
- Twyman, Michael (1975) 'The significance of Isotype in Graphic communication through ISOTYPE' Reading, UK: Department of typographic and graphic communication University of Reading
- University of Bozen Bolzano (2014) 'Long night for research' (translation) available at [http://langenachtderforschung.it/it/lunga-notte-della-ricerca-2?set\\_language=it](http://langenachtderforschung.it/it/lunga-notte-della-ricerca-2?set_language=it), accessed 28 August 2015
- Unnamed author (2012) 'Lega Nord: freniamo l'avanzata cinese' in *Alto Adige*, 28 February, available at <http://altoadige.gelocal.it/bolzano/cronaca/2012/02/28/news/lega-nord-freniamo-l-avanzata-cinese-1.4352991>, accessed 28 August 2015
- Valletti, Ricardo (2011) 'Il primo megastore cinese sfida i negozi di via Torino' in *Alto Adige*, 12 June, available at <http://altoadige.gelocal.it/bolzano/cronaca/2011/06/12/news/il-primo-megastore-cinese-sfida-i-negozi-di-via-torino-1.4250007>, accessed 28 August 2015

# Closed shop? Irish data pioneers battle to access information in ‘open’ regime

**Despite efforts by a few notable data pioneers, data journalism is still in its infancy in Ireland, where access to digital records remains difficult despite official commitments to open data, writes Kathryn Hayes**

## **Introduction**

While journalists in Ireland have been using data as part of their work for decades, Irish media organisations have been slower to engage with large data driven projects compared to their UK and American counterparts, where journalists have used open data to contextualise everything from the socioeconomic background of the London rioters to aiding the evacuation efforts of New York residents in advance of hurricane Sandy (Byrne, 2015). While the size of the country and economies of scale cannot be ignored, a lack of access to structured data is considered one of the main reasons for the delayed Irish data take off. An ‘innate conservatism’ in Irish media is also cited as partly to blame as is the deliberate obfuscation of many of the institutions, which hold the most valuable data sets in Ireland (Linehan, 2015). In recent years there has been a lot of rhetoric in Ireland around open government, however the culture of secrecy that has dominated the civil service since the foundation of the State has delayed any determined drive towards an ‘open data’ movement. The manner in which data held by some Irish government departments is still often released in PDF format or even sent by post has also hindered any large-scale data driven journalism.

Through a series of qualitative interviews with journalism academics, and journalists working in the national media in Ireland, this chapter explores the state of data journalism in Ireland; a country where despite a history of prolonged struggle with opacity in government, a significant number of political and social controversies have been unearthed through original and investigative journalism. Responses from those interviewed suggest that while the landscape of data journalism is improving in Ireland, access to

suitably formatted, open and linked data remains a struggle. The unremitting reluctance of some government departments to release information is also problematic. Appropriate training is also seen as necessary for some journalists who wish to overcome their fear of figures or increase their technical knowledge of how to extract and analyse information from large datasets.

### **Reviewing the Irish data landscape**

The turbulent conditions surrounding the birth of the Irish Free State in 1923 – following the Irish War of Independence against the English Crown and a consequent Civil War – meant openness and transparency in government were not high on the agenda of Ireland's new legislators, who were instead intent on defending against subversive elements in order to ensure survival. Democracy was born at a time of considerable violence and the new Irish civil service was heavily influenced by the Westminster model of government, where citizens had no right of access to information and their relationship with government was deliberately kept at a distance. Chubb (1992: 4) suggests that Irish political culture was heavily influenced by Britain as a result of geography, political, social and economic domination, and culture. From its foundation the fledgling Irish Free State was regarded as centralised and secretive and far from a 'model of openness and transparency' (Felle and Adhsead, 2009).

As an example of the State's efforts at control, the 1923 Censorship of Films Act was one of the first acts on the statute book of the new government. It gave power to an officially appointed censor to keep from the public films, which it believed to be 'indecent, obscene or blasphemous' (Censorship of Films Act 1923, Section 7(2)). Early Free State governments brought in the notion of Cabinet collective responsibility, and confidentiality in Cabinet decision-making. Cabinet handbooks stated that all memoranda for government were to be enclosed in special envelopes and marked for Minister's personal attention and sealed with wax. Documents not likely to be used again were personally burned by the minister's private secretary.

The Emergency Powers Act introduced in 1939 – on the eve of World War II – gave the government wide ranging powers to censor all broadcasts and newspapers in the Free State. In 1963 the Irish government amended the Official Secrets Act, making it a criminal offence for any civil servant to reveal anything, no matter how trivial, without express permission of the minister responsible. That early pattern of control and secrecy continued during the outbreak of the Troubles in Northern Ireland in the 1970s and the invocation of Section 31 of the 1960 Broadcasting Act, which prohibited Raidió Teilifís Éireann (RTE), the national public service broadcaster, from broadcasting anything that could be interpreted as supporting the aims or activities of organisations that 'engage in, promote, encourage or advocate the attaining of any political objective by violent means'. In essence the censorship was used to prevent RTE from using interviews with spokespersons for Sinn Féin and the IRA. The Section 31 ban as it was commonly known was not lifted until 1994 by

the then Arts Minister and current president of Ireland Michael D Higgins. Likewise, the Emergency Powers Act was also not lifted until the same year.

While Ireland did begin to open up, the pace was glacial. Modernisation of government as well as a number of other factors such as Ireland's membership of the EEC; the introduction of accountability laws around audits and Ombudsman legislation; and demands from civil society, all led to the opening up of Irish government and bureaucracy. Ireland introduced freedom of information legislation in 1997, and it was widely seen as a watershed moment in the relationship between citizens and the State in terms of openness, transparency and accountability. The former Labour Minister who introduced the legislation Eithne Fitzgerald said at the time she wanted to turn the culture of the Official Secrets Act on its head and in its place create a changed culture in the civil service (Fitzgerald, 2015). Finance documents were made available to journalists for the first time along with emails sent by civil servants. Citizens for the first time could access their medical and social welfare records. Ireland, many believed, was finally opening up.

But the FOI Act was, perhaps, too successful. Notwithstanding how these initiatives reflected a move towards challenging the 'cult of secrecy' in Irish administration (Doyle, 1996-7: 64-67), the culture of secrecy which dominated the State's first 50 years of existence had undoubtedly left an enduring legacy, which continues to haunt the open data movement. After a while civil servants stopped putting sensitive or controversial information in emails and started using the phone instead (Hennessy, 2008). A number of embarrassing exposés of incompetence by ministers and other rows and mishaps, as revealed by journalists following FOI requests, led in 2003 to the introduction by the then Fianna Fáil led government of the Freedom of Information Amendment Act. That Act introduced fees for making requests; extended the time when access to the records of government was available to 10 years rather than five; gave full protection to communications between ministers concerning matters before government; and introduced blanket bans on information that could be withheld. It was regarded as a 'severe setback' to openness in Ireland (Foley, 2015) and was described by the then Information Commissioner Emily O'Reilly as a 'step back from the commitment to openness, transparency and accountability' that had underpinned the spirit of the original Act (Office of the Information Commissioner: 2008, 13). A year after the amendments were enacted the use of the Act by journalist fell by 50 per cent (*The Irish Times*, May 17, 2004).

In 2014, a subsequent Fine Gael Labour coalition government reintroduced a strengthened FOI Act, reversing most of the amendments and abolishing fees. The legislation was also extended to cover the police and state financial bodies such as the Central Bank – important in the Irish context following the 2007 collapse of the Irish economy and near collapse of the banking system, in part blamed on weak banking regulation and poor oversight by financial watchdogs. That same government committed to an open data regime (see below).

While the history of Irish democracy and the story of FOI are not in and of themselves reasons for a lack of development of a strong data journalism community in Ireland, they do go some way toward explaining why media organisations have struggled to access data. The impact of the 2007 economic crash on media organisations and newspapers; shrinking newsrooms; lack of investment in innovation by media and lack of appropriate data training for journalists and in journalism schools has also contributed to the delayed Irish engagement with data driven projects.

### **Open government and open data**

Large national datasets are held, or at least in most cases controlled, by official bodies. While the Irish government has been historically slow to respond to calls for more openness, there have been moves toward data transparency and the government has committed to the principles of the Open Government Partnership – the worldwide movement toward open and transparent administration. Launching Ireland's Open Data Portal in 2014, the Irish cabinet secretary responsible, Brendan Howlin, said open data formed a core element of the country's first Open Government Partnership (OGP) national action plan.

The initial minimum target of Ireland's strategy was for open data to be published at a minimum 3-star format (non-proprietary machine-readable format). The ultimate aim was to strive towards the implementation of a 'five star open data deployment standard' where all future data would be released in an interactive HTML format or as linked data (Department of Public Enterprise and Reform, 2014). However, the lack of specified time frame to achieve this goal has already been highlighted as a cause for concern (Byrne, 2015) along with the fact that the foundation document for the development of Ireland's Public Service Open Data Strategy was itself released as a PDF – the most unusable format possible.

Notwithstanding the moves towards opening up government data, Ireland is still ranked 36th in the 2014 Global Data Index. The UK tops the list followed by Denmark, France and Finland. Germany is also ranked in the top ten after Australia, New Zealand and the US (Global Open Data Index, 2015).

### **Data successes**

Despite Ireland's continued legacy of secrecy and control, much of the data obtained for some of the biggest examples of public service journalism in Ireland during recent years was sourced through freedom of information requests as highlighted in a survey carried out by RTE's Ken Foxe and published in *the Guardian* (Greenslade, 2013). For example:

- In 2011, Independent TD (Member of Parliament) Michael Healy-Rae agreed to pay €2,500 to the taxpayer following revelations in the *Irish Daily Mail* about hundreds of premium rate phone calls to a reality TV show in which he appeared.

- The *Irish Mail on Sunday* revealed that the then education minister Ruairi Quinn was paid expenses for driving to and from his holiday home.
- Spending of around €105,000 a year on secretarial assistance and mobile phones for former prime ministers ended following a series of stories in *The Sunday Times*.
- The number of staff employed by the office of the Speaker of the Irish Parliament was reduced, thus saving €300,000 annually, following revelations by the *Sunday Tribune*.
- Ministerial travel in 2007 to coincide with St Patrick's Day festivities exceeded €500,000. In 2012, the figure was just €53,142. It followed a series of articles by the *Sunday Tribune* and other newspapers.

In the following sections, this chapter hears from the front lines where interviews with Irish data journalists explore the extent to which newsrooms are using data to tell stories; the problems they face accessing datasets; and what training if any is available to Irish journalists working with data.

### **The data pioneers**

As previously noted, apart from the work of some individual trailblazers, data journalism has been slow in establishing itself in Ireland. The founding of dedicated data initiatives such as *The Irish Times* Data Project follows years of journalistic work that has often borne the hallmarks of data journalism but not carried the label. The *Irish Examiner* has done a number of major data-based projects on health; the *Irish Independent* on policing; the *Irish Mail on Sunday* on clamping; and RTE's newly established Investigations Unit has done a number of programmes underpinned by the use of data journalism. In 2012 an RTE investigation into prostitution used data from escort websites to prove that prostitution was not about individual escorts working for themselves, but rather highly organised networks with links to people trafficking. However, the efforts of individuals were perhaps best described as rudimentary. 'Being very honest newsrooms just weren't adapted for it,' admits *The Irish Times* editor Kevin O'Sullivan, 'even though we would have done stuff over the years, I was very conscious that this is a very important digital front and that *The Irish Times* should be involved in it' (O'Sullivan, 2015). O'Sullivan's comments were echoed in other newsrooms.

#### *The Freelance*

Investigative journalist Gavin Sheridan is a doyen of Irish digital journalism and was considered a trailblazer when it comes to data driven reporting in Ireland. After becoming aware of [www.theyworkforyou.com](http://www.theyworkforyou.com), a web site dedicated to transparency in the British Parliament, Sheridan established [KildareStreet.com](http://KildareStreet.com), a service that tracks activity in the Irish Oireachtas (parliament) and which is designed to allow citizens keep tabs on their TDs (members of parliament) and



Senators. Sheridan describes the site as the 'go to' place for citizens who want to know who their local representative is, how often they speak in parliament, what their expense claims are and what donations they have received. Sheridan is also co-founder of TheStory.ie, which is a chronicle of his FOI requests to diverse branches of the Irish government and a growing public database of government records. According to its mission statement TheStory.ie is dedicated to 'sharing documents, combing and combining data and promoting transparency in public life' ([www.thestory.ie](http://www.thestory.ie)). One of the first posts on TheStory.ie involved the conversion of word docs and PDFs to spreadsheets, analysing all published donations to TDs (members of parliament) from 1997-2008.

### *The broadsheet newspaper*

*The Irish Times* Data Project was launched in February 2015 and is regarded as an important digital front by the newspaper, which is determined to give its readers more 'immersive journalism' (O'Sullivan, 2015). One of its opening data projects 'What's the most reliable car in Ireland?' (O'Brien and Scott, 2015) involved distilling complex sets of figures to work out what cars are most likely to pass the Irish MOT test. Another 'Is your name going out of fashion?' (Duncan, 2015) used tens of thousands of data points to work out which baby names are going in and out of fashion. The baby name story attracted more 170,000 readers in the first week and was hugely popular on social media according to O'Sullivan. In March 2015 an investigation by the newspaper into donations to the Sinn Féin political party in America saw the team digitise 20 years of filings made by Friends of Sinn Féin (FoSF), the party's US fundraising arm, allowing for analysis of cumulative donations made by donors (Duncan and Carswell, 2015). The filings are publicly available on the US Department of Justice website in PDF format, however as with most information held in PDF form, the data had to be uploaded into a format which would allow the data team to calculate totals for individual donors and the total amount raised in particular US states. The project also required that each and every row and column, 14,879 entries (or almost 60,000 individual cells), were manually checked. (Duncan 2015) According to the editor of the *Irish Times* Kevin O'Sullivan, the Sinn Féin expenses story required a huge amount of 'graft'; combined with old fashioned and modern journalism techniques, to ensure the figures were accessible and clearly understood.

*'It is demanding, and it does require special resources. But if it is fulfilling our needs in terms of being more digital and also generating journalism that is more immersive or journalism that has more of a news edge then we can justify doing it, and I will certainly think we are going to do more of it,'* (O'Sullivan, 2015).

### *The broadcast newsroom*

Ireland's national broadcaster RTE established a multimedia Investigations Unit in 2012 after one of its flagship programmes *Prime Time Investigates* was axed

following an investigation into the station's reportage of allegations against a Catholic priest. The station paid a six figure sum to the priest after unfounded allegations in the 'Mission to Prey' programme that wrongly alleged he had raped a minor and fathered a child by her while working in Kenya in the 1980s. After the Investigations Unit was established, part of its ambition was that it would focus not only on long-form TV projects, as it had done in the past, but expand its brief to include online and data driven projects (Foxye, 2015).

Since joining the Unit in 2013, reporter Ken Foxye has been involved in a number of investigative projects driven by the use of data including a story on political expenses (Foxye, 2015) which he admits was one of the most labour intensive projects he has been involved in to date. 'I probably would not have been able to do that project in any other environment other than RTÉ because of the time involved as the nature of the existing records is quite haphazard and an awful lot of it had to be done manually,' he explained.

Nonetheless, Foxye believes the idea that all data journalism projects are incredibly labour intensive is often more of a fear than a reality. One of the biggest problems in Irish journalism is inertia, he argues, and a fear of doing something because it might not work out, take too long, or end up causing embarrassment. Media companies are all fully aware that their future lies in the better use of technology – whether they are able to adapt to the change is more often the question, rather than their willingness to try.

*'I think perhaps we need to make a distinction between data-based journalism, which I think has been relatively common in Ireland – and what has become known as 'data journalism', which is probably the attempt to display this information in a fuller, more visual, or interactive way, rather than it just being an article or TV programme based on datasets or statistics. This modern interpretation of 'data journalism' obviously lends itself to an online experience for visual and interactivity reasons. And certainly Ireland has been behind the UK and the USA in online innovation right across the media, not just in this area,' (Foxye, 2015).*

#### *The mid-market tabloid*

The Independent News and Media (INM) group, which owns a number of national newspaper titles in Ireland including the *Irish Independent*, has put significant resources into its 'digital first' strategy and 'understands that the future of news for their readers is data led' (Caffrey, 2015). In February 2015 the *Independent* ran a six-part series 'Future Proof: Planning where we live' that examined the housing need in each county in Ireland, and explored the challenge of providing key services including water, transport, schools, childcare, and health services in communities of the future. The articles about the future of residential planning were based on data from Ordnance Survey Ireland, according to the paper's Environment Editor Paul Melia.

*'People are willing to engage in big worthy projects if they are presented right and if they are easy to understand and if the information is all in the one place. That's going to be*

*the trick for publishers. People want to see the trend but they want to be able to drill down to their own area,' (Melia, 2015).*

Another recent example of data-driven project undertaken by the *Irish Independent* highlighted the correlation between penalty points and accidents and examined where motorists were most likely to get penalty points in Ireland. Melia believes everyone is on a learning curve when it comes to embracing data journalism and lack of good data, time and skill shortages are also reasons for the delayed engagement (2015).

### **Getting to the data**

There have been some positive developments from individual public bodies. Located in north County Dublin, Fingal Council is the third largest local authority area in Ireland by population. Fingal Open Data launched in November 2010 was the first Open Data website in Ireland. Other public bodies in Ireland have since started publishing Open Data including Dublinked – the Dublin Region Innovation and Open Data Initiative; data.localgov.ie – the Irish Local Government Open Data Portal; and data.gov.ie, the Irish Government Open Data Portal. But the pace of publication has been painstakingly slow, according to RTE's Ken Foxe:

*'The government's open data programme is a positive step but obviously I would have concerns that it would turn into a bit of a cosmetic exercise to appear as forward thinking. You go to the data.gov.ie website currently and it just seems like a bit of a data dump at the moment. There is very little there to assist people in finding what they are looking for or to help them understand what exactly it is they are looking at,' (Foxe, 2015).*

While there is a certain amount of data available in Ireland from government departments and agencies such as the Central Statistics Office (CSO); the Health Service Executive (HSE); the Road Safety Authority (RSA); and the Environmental Protection Agency (EPA), there are significant amounts of data that is not readily available. Moreover there have been a number of high profile cases in Ireland where transparency only occurred after scandals were exposed. In 2005 an RTE *Prime Time Investigates* programme 'Home Truths' on the state of nursing home care at the Leas Cross facility in Dublin, prompted a national debate on how the elderly were being looked after in Irish nursing homes and mounted pressure on the HSE to publish inspection reports on private nursing home care. (Holland, 2006) Following the Leas Cross controversy, the Information Commissioner Emily O'Reilly said it was vital that such reports were made public (O'Reilly, 2006).

The widespread abuse of the penalty points system within the republic of Ireland's national police force, An Garda Síochána, was only exposed after it was highlighted in 2012 by whistleblowers Sergeant Maurice McCabe and former Garda John Wilson. The allegations led to the subsequent publication of the Guerin Report in 2014, which found that serious crimes were improperly

investigated by police. The report was also critical of the treatment of the whistleblowers (Guerin, 2014). Recent changes to FOI legislation in Ireland has seen the law extended to cover more public bodies including the police, however only administrative files can be requested. Data that is deemed sensitive or related to intelligence matters is not covered by the act (Freedom of Information Act, 2014). There is a lot of inconsistency in how data is released, and what data is available.

The release of high value linked datasets has the potential to reveal answers to important societal issues that could otherwise go unanswered; allowing journalists and academics opportunities to explore previously undiscovered trends while affording citizens opportunities to make more informed decisions about their lives. For example if someone wanted to know more about the area in which they were purchasing a house they should be able to access information about crime statistics for the locality and information on schools and local health services. 'If you want to buy a house and you want to get crime stats, the best way of finding out if your house is in a dodgy area is to ring up the local Garda (police) station and the local Garda will generally tell you. That's pretty much how it's being done,' said Paul Melia, Environment Editor, the *Irish Independent*.

While the Irish government's new data initiative has positive goals, including the mandatory incorporation of all public bodies under one uniform data release programme where reports will be released in the same format and at similar frequency, the portal where all this data is to be accessed, ([www.data.gov.ie](http://www.data.gov.ie)), has as yet been slow to reveal any of the promised linked data.

### **Culture and training**

But it's not just a lack of available data that has hindered the development of data journalism in Ireland. News organisations have been slow to embrace change, and many journalists have little training in data techniques. Fully embracing data requires a change in mind set from news organisations, which must start working towards embracing collaborative and team-based working models. Journalists don't do enough of that according to *The Irish Times* chief reporter and a member of their data team, Carl O'Brien:

*'We tend to be sole operators. When you look at other disciplines like maybe science it's very much a team based healthy model where you can bounce ideas off each other and cross pollenate ideas. It's a very effective way of working – that's a good learning curve for everybody.'* (O'Brien, 2015).

Most Irish journalists have also traditionally have had little or no numeracy or data training. With the exception of libel law training, historically journalism has never been associated with up-skilling and continuous professional development, which is often far more common in other disciplines. If Irish media organisations are to fully embrace data driven journalism, there is a need for appropriate training rather than simply relying on self-taught tech savvy reporters and developers or individuals willing to educate themselves. Those

interviewed identified a role for journalism schools to introduce data training as part of their journalism education programmes.

## Conclusion

The establishment of dedicated data initiatives is a positive step towards making data journalism an everyday practice in Irish newsrooms. The influence of the work that is being done abroad and by individual data pioneers in Ireland has also demonstrated the benefits of exploiting the full potential of this approach to storytelling. However, with the circulation of Irish newspapers in decline, it's difficult to determine where the resources will come from for large-scale data driven projects. According to figures from the National Newspapers of Ireland (comparing January-June 2009 and January-June 2014) Ireland's biggest selling daily the *Irish Independent*, had 26 per cent wiped off its circulation while *The Irish Times* lost almost 30 per cent. The population size and economies of scale cannot be ignored when comparing Irish media outlets with big hitters in the US and UK who have embraced data-driven journalism. With a population of 4.6 million Ireland has a limited audience of potential readers (for advertisers to target) thereby limiting the budgets available to newsrooms for investment in data projects.

However in a modern digital society, where information is increasingly stored in data it is imperative that journalism in its duty to inform fully explores the potential of data. Journalists must find ways to access, analyse and explain this information, especially if the 'truth lies in the data' (Linehan, 2015). Equally, the government's commitment to the principles of the Open Government Partnership must be honoured with the provision of good data.

The role of data journalism in the future of journalism is undoubtedly significant and while for some journalists it might remain a specialism, for others it will be used as an important tool in their armoury. The basic tenets of journalism – cultivating sources and finding stores – remain as relevant and valuable as ever.

## References

- Byrne, Robbie (2015) Howlin: 'New data strategy 'a major milestone'. So what's it about' in the *Irish Examiner* 19 June, available at [www.irishexaminer.com/viewpoints/analysis/howlin-new-data-strategy-a-major-milestone-so-whats-it-about-337973.html](http://www.irishexaminer.com/viewpoints/analysis/howlin-new-data-strategy-a-major-milestone-so-whats-it-about-337973.html), accessed 9 September 2015
- Chubb, Basil (1992) *The Government and Politics of Ireland*, (3rd edn) London: Longman
- Caffrey, Elaine (2015) 'Independent.ie – A digital first strategy' available at <http://digitalmarketingstrategy.ucd.ie/independent-ie-digital-first-strategy>, accessed 9 September 2015
- Department of Public Expenditure and Reform (2014) 'Ireland and the Open Government Partnership', available at [www.per.gov.ie/en/open-government-partnership-ogp](http://www.per.gov.ie/en/open-government-partnership-ogp), accessed 8 September, 2015

*Closed shop? Irish data pioneers battle to access information in 'open' regime*

Duncan, Pamela (2015) 'Is your name going out of fashion?' in *The Irish Times* Data Blog, 3 February, available at [www.irishtimes.com/news/ireland/irish-news/is-your-name-going-out-of-fashion-1.2089357](http://www.irishtimes.com/news/ireland/irish-news/is-your-name-going-out-of-fashion-1.2089357), accessed 8 September 2015

Duncan, Pamela and Carswell, Simon (2015) 'Sinn Fein's money' in *The Irish Times* Data Blog, 5 March, available at <http://www.irishtimes.com/news/ireland/sinn-feins-money>, accessed 8 September 2015

Felle, Tom and Adshead (2009) *Democracy and the Right to Know: 10 years of the Freedom of Information Act in Ireland*. Limerick: University of Limerick Papers in Politics and Public Administration, 2009/4

FitzGerald, Eithne (2015) 'Ending the culture of secrecy: What we wanted to achieve when we introduced the original FOI legislation' in Adshead, Maura and Felle, Tom (eds) *FOI@15: Ireland and the Freedom of Information Act*. Manchester: Manchester University Press

Foley, Michael, (2015) 'Keeping the State's secrets: Ireland's road from 'official' secrets to freedom of information' in Felle, Tom and Mair, John (eds) *FOI 10 years on: freedom fighting or lazy journalism?* Bury St Edmunds: Abramis

Department of Public Expenditure and Reform (ND) 'Foundation Document for the development of the Public Service Open Data Strategy', available at [www.per.gov.ie/en/open-data](http://www.per.gov.ie/en/open-data), accessed 10 September 2015

Foxe, Ken (2012) 'Ruairi Quinn's mystery mystery €1,400 mileage claim: Minister says he drove 5,100kms but his diary shows only 1,000kms' in *The Irish Mail on Sunday*, 19 February, available at [www.dailymail.co.uk/news/article-2103184/Ruairi-Quinns-mystery-mystery-1-400-mileage-claim-Minister-says-drove-5-100kms-diary-shows-1-000kms.html#ixzz3lG09dOVZ](http://www.dailymail.co.uk/news/article-2103184/Ruairi-Quinns-mystery-mystery-1-400-mileage-claim-Minister-says-drove-5-100kms-diary-shows-1-000kms.html#ixzz3lG09dOVZ), accessed 9 September 2015

Global Open Data Index, available at <http://index.okfn.org>, accessed 10 September 2015

Government of Ireland (1963) *Official Secrets Act*. Dublin: Government Publications Office

Government of Ireland (1997) *Freedom of Information Act 1997*. Dublin: Government Publications Office

Government of Ireland (2003) *Freedom of Information (Amendment) Act*. Dublin: Government Publications Office

Government of Ireland (2014) *Freedom of Information Act 2014*. Dublin: Government Publications Office

Greenslade, Roy (2013) 'Irish journalists alarmed at Freedom of Information pricing proposal' in *the Guardian* online blog, available at [www.theguardian.com/media/2013/nov/13/irish-journalists-alarmed-at-freedom-of-information-pricing-proposal](http://www.theguardian.com/media/2013/nov/13/irish-journalists-alarmed-at-freedom-of-information-pricing-proposal), accessed 8 September, 2015

Hennessy, Mark (2008) *The Irish Times* journalist in comments to expert focus group on freedom of information 'Democracy and the Right to Know' seminar at University of Limerick, Ireland, 29 February, 2008.

Holland, Kitty (2006) 'Harney to consult AG over Leas Cross report' in *The Irish Times*, 3 October, available at [www.irishtimes.com/news/harney-to-consult-ag-over-leas-cross-report-1.1010360](http://www.irishtimes.com/news/harney-to-consult-ag-over-leas-cross-report-1.1010360), accessed 8 September 2015

Guerin, Sean SC (2014) 'Review of the Action Taken by An Garda Síochána Pertaining to Certain Allegations Made by Sergeant Maurice McCabe' (Guerin Report), available at [www.merrionstreet.ie/en/wp-content/uploads/2014/05/Final-Redacted-Guerin-Report.pdf](http://www.merrionstreet.ie/en/wp-content/uploads/2014/05/Final-Redacted-Guerin-Report.pdf), accessed 8 September 2015

Information Commissioner (2008) *Annual Report of the Freedom of Information Commissioner 2007*. Dublin: Government Publications Office

Irish Government Data Portal [www.data.gov.ie](http://www.data.gov.ie)

Linehan, Hugh (2015) 'Data Journalism: Reinventing news for 21st century digital society.' in *The Irish Times* 27 February available at [www.irishtimes.com/opinion/data-journalism-reinventing-news-for-21st-century-digital-society-1.2119160](http://www.irishtimes.com/opinion/data-journalism-reinventing-news-for-21st-century-digital-society-1.2119160), accessed 9 September 2015

News Brands Ireland (2015) 'Circulation' available at <http://newsbrandsireland.ie/data-centre/circulation>, accessed 8 September, 2015

O'Brien, Carl and Scott, Paul (2015) 'What's the most reliable car in Ireland?' in *The Irish Times* Data Blog, 5 March, available at [www.irishtimes.com/news/environment/what-s-the-most-reliable-car-in-ireland-1.2073954](http://www.irishtimes.com/news/environment/what-s-the-most-reliable-car-in-ireland-1.2073954), accessed 8 September 2015

O'Reilly (2006) 'O'Reilly "disappointed" Garda not subject to FOI' in *The Irish Times*, 11 May, available at [www.irishtimes.com/news/o-reilly-disappointed-garda-not-subject-to-foi-1.782425](http://www.irishtimes.com/news/o-reilly-disappointed-garda-not-subject-to-foi-1.782425), accessed 9 September 2015

Saorstát Éireann (1923) *Censorship of Films Act*. Dublin: Government Publications Office

TheStory.ie (ND) 'Our Goals' available at [www.thestory.ie/our-goals](http://www.thestory.ie/our-goals), accessed 8 September, 2015

## Interviews

Duncan, Pamela, Data Journalist, *The Irish Times*

Foley, Michael, Lecturer in Journalism, Dublin Institute of Technology

Foxe, Ken, Investigative Journalist, RTE Investigations Unit

Melia, Paul, Environment Editor, *Irish Independent*

Kerrigan Mick, Principal Data Scientist, *Irish Independent*

O'Brien, Carl, Chief Reporter, *The Irish Times*

O'Sullivan, Kevin, Editor *The Irish Times*

# Global models for starting data journalism in developing countries

**There is a pressing need for data journalism training in developing countries but the appeal of the boot camp model is starting to fade, writes  
Eva Constantaras**

## **Introduction**

In a recent data journalism workshop hosted by the Media Initiatives Center in Yerevan, Armenia, a young journalist discovered that after years of leading the European Union in the rate of incarceration, with a steady uptick over 5 years, the rate had suddenly stalled. Not only that, it had dropped. How did she account for that? Had a bunch of prisoners been freed? Was there a drastic decrease in criminality in Armenia? Had the criminal justice system in Armenia changed policy or practice in a significant way? Instead of being excited about this possible scoop, the journalist looked at me and shrugged. She said that everyone knows that the government is corrupt and cyclically arrests and pardons enemies and allies and the media house wouldn't be interested in funding her to find out who had been released or why. Sure, she could learn to graph the trend, but, why bother? This is a question not only for the global data journalism community but also for the open data community in general. From the perspective of the open data community, Jonathan Gray recently questioned the certainty of the data revolution:

*'We might consider moderating some of the mythologies of spontaneously self-organising actors that will optimise society, if only we can create the right conditions for them to flourish. Who, specifically, do we anticipate will use public information? And how, specifically, do we anticipate that they might use this information to bring about the kinds of social and political objectives that we desire? ... If we do not scrutinise these questions we risk being left with, for example, data without users or analysis without action' (2015).*



Data journalists have a potentially tremendous role to play in acting out its role as a public-service watchdog by transforming the flood of publically available data into insight that facilitates citizen engagement in the democratic process. From Nigeria to Rwanda, Chile to El Salvador, Pakistan to Cambodia, Twitter is exploding with anecdotes of successful data journalism conferences, boot camps, hackathons and fellowships propelling forthwith propaganda-ridden, analog, ambulance-chasing yellow journalism into the world of news apps, investigations and citizen engagement.

But even in the United States, that's not an accurate description of the journalism revolution. But what appears on Twitter is not a reflection of the day-to-day reality of the practice of journalism either in Western or developing countries. Despite the potential of digital and data tools, in most newsrooms, they simply aren't there. According to *The Goat Must Be Fed* authors, 'Our biggest finding is that data journalism is out of whack with the hype – and we need to acknowledge that we've been part of the problem,' (Adair, Kamalakanthan, and Stencel, 2014).

### **Why digital tools are missing**

In donors' excitement to embrace the open government and open data movement, they have pumped lots of money into the quickest, cheapest and flashiest path to data journalism: boot camps, hackathons and conferences. Yet these approaches boil down the barrier to data journalism into one simple problem: technology. These boot camps are designed to provide technology solutions, with the tacit assumption that the rest will follow, but they have misdiagnosed the essential root problem. *Goats* and a vast array of experience in failed data journalism initiatives tell us, it's not the tools, at least not primarily.

The primary challenge is rooted in the assumption of the role of the media itself as a public service watchdog. It speaks to the apathy I saw demonstrated in Yerevan. Flipping through a few Kenyan newspapers, the number of headlines with direct quotes is overwhelming. The role of media in many places is not to report the truth, but rather, to quote powerful people espousing their version of the truth. Once the quotation marks are removed and replaced by data, the journalist assumes some responsibility for content verification and in a place where governments and their data are distrusted, this is not something many journalists want to stake their reputation on. Actually getting to the bottom of things and identifying a chain of responsibility can be much riskier than exposing corruption in general terms in countries where corruption, graft, mismanagement, incompetence are all expected and scratching too deep can be dangerous. Therefore, the motivation to exercise the kind of analytical thinking skills required for the kind of data journalism that helps solves problems is often missing. And, as my Armenian colleague pointed out, with high levels of public apathy, it might not be worth the effort. Those journalists who are interested in that kind of work are usually busy, really busy, and not likely to sign up for flashy data journalism events.

A critical barrier to data journalism in countries where it has the greatest potential for good (the most corrupt, unequal and impoverished) the biggest challenge is simple data literacy both among the media and among citizens. A recent study examines why journalism students at the University of Georgia don't study computer science, a key ingredient for data journalism, finds:

- They don't know they should
- They think they will fail
- They don't think they'll enjoy CS classes (Cook, 2015)

Data literacy for mid-career journalists in developing countries, which is also tech-intensive, can seem equally irrelevant, intimidating and unrewarding.

Thirdly, the absence of a media industry crisis, no matter how imminent, offers a paradoxical barrier to innovation and exploration of digital content, including data journalism. Publishers, editors and journalists see no need to engage in a difficult, expensive, risky endeavour when for now, their traditional business model is stable. People are still buying newspapers, listening to radio and tuning into the nightly news. Additionally, nobody has produced a compelling business model based on data that can win over publishers. Those plug-and-play tools that seem like an easy way to get newsrooms started in data are often not supported by the CMS. Web traffic is so low that editors don't want to bother adapting.

### **Models for growing data journalism**

In many cases, these challenges are glossed over and with no real metrics for impact of data journalism interventions, years later, the same activities are still being rolled out and failing despite the social media buzz they generate. To understand these challenges, I will evaluate three common models for seeding data journalism in developing countries— fellowship, boot camp and cross border—and recommend ways to improve them.

#### *The fellowship revolution*

Fellowships are a popular strategy for trying to blend journalistic and data science skills in the newsroom. Either a developer is embedded in a newsroom or aspiring data journalists are removed from the newsroom and participate in intensive training and project creation for the fellowship period. With the challenge to basic data literacy and widespread analytical capacity as high as it is, it is unrealistic to expect most journalists to take on the programming that is often used in developing interactive data products. The logical course is to try to lure developers and programmers into journalism (Susman-Peña, 2015).

Data literacy, journalist engagement and online platforms' limitations are often not factored into the developer fellow model, borrowed from Code for America. As this fellow is generally neither a journalist nor a teacher, often he or she faces a yawning data and digital literacy gap and unable to revolutionise the

newsroom alone, builds an isolated news app or dashboard that may or may not ever be migrated to the host news organization because of limitations to the CMS.

One of the few official studies of these embedded fellowship programs evaluated Code4Kenya, a project of the Africa Media Initiative Code4Africa Program and the World Bank in 2012. Code4Kenya aimed to develop applications that would create demand for government data and in the process, catalyse institutional change as well as how citizens engage with the government through fellowships with the mainstream media. According to the study, Code4Kenya sites received an average of only approximately 3,000 daily page views (with a 78 per cent bounce rate – i.e. visits to the home page without further interaction on the site).

Most of the websites are still in beta mode, incubated in the main Code4Kenya site as opposed to live sites maintained by host organisations. It seems that the media houses and civil society organisations did not have a sense of ownership of these applications and were therefore still relying on the Code4Kenya pilot program to maintain and develop further the applications (Mahihu and Mutuku, 2014).

Journalists of host outlets, some of whom had attended data journalism boot camps, did not write stories around the data. The Star Health Dodge Doctors App, developed over a year after the fellowship ended, has been touted as the future for data-driven business models, enabling citizens to pay a fee to research their doctor's credentials (Looney, 2015). Yet the link to the app on the Star Health site leads instead to a citizen reporting app and has only been downloaded from Google Play between 10-50 times and is only one of two apps available from Code4Kenya. There is not even a link on the Star homepage that leads to the Star Health page, which features the database.

More successful developer embed programs have taken place in media houses with an established data journalism team, such as the Knight International Journalism Fellowship program. This program has embedded a fellow for several years in La Nación of Argentina, South America's leading data journalism outlet and sponsored a fellow at InfoAmazonia, a data-driven environmental reporting platform in Brazil that is supported by several international donors.

Learning from the Code4Kenya experience, Internews in Kenya piloted a data journalism fellowship modelled after the Stanford Knight Fellowship. This paid fellowship removes media professionals from the newsrooms for an extended period of time to focus on skills building in and out of formal classroom settings, and develop ambitious journalism projects with technical and editorial assistance. The Internews in Kenya office, which had a 10-year history and relationships with all the major media owners, offered a competitive five-month fellowship to three journalists, a graphic designer and developer in 2013. These were young professionals from mainstream newsrooms who participated in an intensive training, mentoring and editorial support program with the five-person

Internews in Kenya data team. Editors and publishers committed to continued employment for the fellows as well as to publication of their completed stories.

The fellowship overcame data literacy barriers and produced high-impact investigations, though mostly through low-tech data visualisations distributed through print and TV. The 'Change for Health' series published by the *Standard* in Kenya prompted several county administrations to contact the journalist for further data and advice on developing a health budget for the following fiscal year (Wafula, 2013). 'When the sun sets in Turkana; Hunger stakes and stripes in the North' the lead news story on NTV (Juma, 2014) revealed that diverting funds from emergency food aid to food security programs could end famine and let to the implementation of a drought relief strategy. A front-page investigation led by the *Standard* fellow into corruption in the donor-funded system of cash payments for poor Kenyans (Wafula, 2014) led to an audit and overhaul of the system (Wafula, 2015) Policy changed, but only after five months of sweat and tears and continued follow-up.

### *Fine-tuning the fellowship*

The challenge of fellowships, deciding between mass media and elite media, again mirrors the division in the United States.

*'There's no question that some news organisations are doing wonderful, innovative data work that has created extraordinary journalism. They have mined large data sets, invented tools and used new reporting techniques to expose corruption and explain complex issues in ways that weren't possible before the digital age. They've been aided by millions of dollars from foundations that have funded innovation through grants, pilot projects, training and journalism education. And yet, hundreds of news organisations are still stuck in the analogue past, doing meat-and-potatoes reporting that doesn't take advantage of the new tools'* (Adair, Kamalakanthan, and Stencel, 2014).

Working with the latter groups is challenging but has a potentially greater public impact by democratising access to data-driven content. Bringing together journalists with other members of the technology community through group fellowships result in high-profile projects in mainstream media that engage citizens in a public debate and encourage a reaction from policy makers. These fellowships grow data literacy and public service journalism while also teaching the interactive multimedia content production skills that media houses seek. Civic hackers come to regard journalists as a valuable partner in disseminating data findings that shape public policy. However, these fellowships are only appropriate in strong media markets where media houses can afford to release a fulltime staff member for an extended fellowship. They are also resource intensive for the training institution, which requires a fulltime staff of locally-based data professionals. A less intensive approach may be a part-time year-long fellowship where media professionals commit to a certain number of days and stories per month.

Embedded fellows for niche data-driven media houses are often easier and more resource efficient. There are a growing number of niche digital data-driven media outlets across the world that could benefit from the embedded developer fellowship model. Many digital outlets have the data literacy to leverage embedded data experts to scale up production. The debate is whether the audiences for these media are too elite and whether there can be true public engagement through these platforms.

### **The data journalism boot camp buzz**

For many in the data journalism community, data journalism boot camps and hackathons invoke an uncomfortable déjà vu of years spent explaining to donors that transforming a media ecosystem – converting a propaganda-heavy media cycle to a bastion of independent media – requires more than a 40-hour workshop, no matter how well designed. Data journalism boot camps and hackathons, which began as a place to generate buzz around the open data movement, have now become a cheap substitute for actual sustainable investment in data journalism capacity. The media development literature is overloaded with lessons cautioning that short-term, parachute training engagements have limited to no impact, especially in low data and digital literacy and high development priority environments such as South Sudan where two-day workshops are popular yet ‘many journalists lack the most basic writing skills’ (Fojo, 2012). In many countries, journalists are poorly paid, a promise of a daily stipend is motivation enough to attend a workshop for a few days. Some form of long-term educational engagement designed specifically for adult learners will be necessary for data journalism (Susman-Peña, 2015).

Despite a flurry of Twitter traffic, guest appearances by famous Western stars of data journalism and flashy prototypes produced by invited developers, the glowing appeal of the boot camp model is starting to fade. Besides the immediate challenge of a crash course in data science, learning data journalism skills requires an uncomfortable attitude shift for many journalists to a more scientific and collaborative work style. As Paul Bradshaw, a veteran data journalism professor explains, ‘It is an open approach to reporting that borrows more from the culture of programming than journalism’s own culture of guarding information jealously’ (Bradshaw, 2015). Even if a run-through of a tool in a boot camp goes smoothly, once a journalist gets back to the newsroom, that tool is not going to work perfectly and the journalism may give up.

The popular boot camp model is crippled by a few fatal design flaws:

- Huge, 50-plus person events without a pedagogical strategy for teaching the necessary theoretical and technological skills for basic data literacy;
- Failure to recruit promising journalists with the analytical skills who are most likely to thrive in data journalism but would not prioritise attending a massive boot camp;

- Wasted resources to bring in high-profile speakers with little understanding of the local data or media environment;
- Focus on developing prototypes for news apps and recycling cutting edge-digital product ideas from previous boot camps that are far removed from the goal of enriching the quality of traditional media reporting with data and analysis;
- Lack of buy-in from editors and publishers to publish stories produced by budding data journalist and at worst, an after care system where a Western data journalist provides remote support to trainees and effectively produced a data story for him or her.

### **Rescuing the bootcamp**

Many media development organisations are experimenting with composition, intensity and post-event support needed to ensure data journalism workshops have the desired impact: getting data-driven stories out through mass media. There is a huge potential for intensive training to overcome key barriers to data journalism, namely:

- Sustained training in data literacy basics that lays the groundwork for more balanced, purposeful and objective reporting in countries with a weak history of independent journalism;
- Identifying and supporting journalists who want to pursue public service journalism and linking them with other members of the open data community who can work together to lower the barrier to data journalism production; and
- Actively persuading publishers and editors that data journalism can be a valuable asset in the face of digital convergence and discussion the transition for breaking news to in-depth reporting.

One approach piloted in Cambodia was a 14-day workshop funded by the German government and implemented by DW Akademie that trained 12 local participants including journalists, communication officers, and a journalism professor in data journalism skills. All trainees posted a final project including story and visualisation published on the training blog (DW Akademie, 2014). The project went a step further and conducted community meetings to ensure that the content made its way to an offline audience as well (James, 2014). IREX has implemented a similar approach in China where a combination of in-house workshops, grants, competitions and events has created a distinct space for data journalism in a country in the middle of a digital transition (2014).

In countries without the opportunity to implement a fellowship training program, such as in Afghanistan, Internews has designed a series of data journalism interventions around specific themes including elections, public health and economics, with mentoring and small production grants, to support

journalists to continue growing their skills between training periods. Stories produced by these trainees touch on governance issues such as arms trafficking, opium production, corruption perception and violence against journalists. Another model developed by Internews for pilot testing is the Data News Lab, where participants spend six weeks working intensively in a data-driven daily newsroom environment run by a team of trainers, combining capacity building and production for home outlets. More research into the goals of capacity-building activities, the impact of different training models and closer tracking of workshop outputs would help shift a model based on open data hype to one of journalistic content.

### **Cross-border global data investigations**

The burgeoning field of cross-border reporting has catapulted the media of several countries into the world of data thanks in part to efforts such as the Organized Crime and Corruption Reporting Project (OCCRP) and the International Consortium of Investigative Journalists (ICIJ). Both of these organisations harness national and international databases to uncover financial abuses in a globalised world. OCCRP provides a unique example of locally-driven investigations with a goal of changing regional governance with members spanning Eastern Europe through to Central Asia. Members come from media houses across the region to share data and produce collaborative stories such as 'The Russian Laundromat', which investigated the regional reach of Russian money laundering by organised crime networks from Russia to Europe (OCCRP, 2014).

Alexandre Léchenet's new research into cross-country data journalism collaborations found, many success stories come from media that prioritises working together for richer content. According to Léchenet, 'an open-source and collaborative culture lends itself well to projects that seek to untangle a story hidden in a large dataset. Finding information as well as writing complete stories afterwards can be complicated for one person alone' (2015).

The underlying logic of cross-border collaborations: discrete, non-competing audiences in different media markets and localised content for each member's audience and political context underpin a lot of the success of data-driven investigative projects. It applies even more to smaller media rather than international behemoths competing for the same online audience, which Léchenet cites as a possible future deterrent to collaborative reporting (*ibid*). Yet cross-border data journalism faces the same capacity issues that many data journalism initiatives do in developing countries. ICIJ, for example, isolated and distributed data relevant to journalists, country-by-country for 'Offshore Leaks'. OCCRP works primarily with already trained data journalists who are members of their network.

'Offshore Leaks' and 'Migrants Files' proved that strong project management, by people understanding the data well, is crucial in carrying out a big investigation of this kind. The data-journalist can assume this role by being the

main contact with other journalists who are not so good with the data (2015). A recent collaboration between ICIJ and African journalists illustrates the limitation. The African journalists involved in the 'Fatal Extraction' investigation of Australian mining companies in Africa were, at the most, junior partners, publishing independently in their own outlets and not as authors of any of the principle ICIJ products (2015). The Africa Network of Investigative Journalists played a supporting role. Contributions by African members provided local fact-checking of the situation, logistical, data and editorial support but little in-depth data analysis, further evidence that data skills do not simply 'rub off' on journalists through exposure to other data journalists but require sustained investment of resources. The muscle behind the database analysis was ICIJ's data team. Arguably, the local governments have the largest role in future regulation of extractive industries and thus exploration of local data had potentially greater policy impact. But, as with other cross-border projects, such as 'Migrant Files', the focus is primarily on Western media and a Western audience for a policy response.

In his criticism of another ICIJ project that examines displacement of populations by World Bank projects, Nicholas Benequista pointed out, to fix an international development issue, the most effective strategy is to go after the people who have a vested interest in their local reputation and better yet, go after them through local media (2015). This begs the question about the objective of cross-border data journalism. Is the goal to expose the wrongdoing in order to give the right public the information they need to solve the problem or is it about scandal? As Jonathan Stray suggests '...it is always about scandal - what has been called "*the journalism of outrage*". This has sometimes made investigative journalism powerless in the face of huge systemic issues without a clear locus of wrongdoing' (2015).

#### *Globalizing audience for cross-border data journalism*

With a dearth of quality international news coverage during a persistent media industry crisis, cross-border reporting may be the future for in-depth global coverage and journalists in developing countries should be taking a more and more active role. Projects such as 'Influence Mapping', a new initiative supported by Open Society Foundations seeks to document relationships between people, organisations, and political processes. Perhaps even more importantly, it seeks to facilitate collaboration on investigating these relationships. It will enable databases from different countries to talk to each other and hopefully spur their journalists to do the same as well. This should enable investigations that range from the global to the hyper-local through media most likely to prompt change.

#### **Conclusion: making the jump from tools to change**

Western media pundits are wringing their hands about the role of journalism, with terms like explanatory journalism, context journalism, and solution journalism being discussed as potentially both lucrative and competitive in a



digital media environment. This goal is a given for donors seeking to fund data journalism in developing countries with the express goal of fostering transparency and accountability. According to Stray, who is also one of the leaders of 'Influence Mappers', 'in a recent study comparing the same story with and without a proposed solution, readers who read the solution reported being more likely to share the article on social media, and read other articles on the same site or on the same issue.' Or, if the aim is to educate citizens to be more informed and active in the democratic and government process, that also requires a more sustained approach. 'If we're serious about the notion of an independent check on government, we need to get systematic about it' (2015).

In countries desensitised to corruption, efforts to grow data journalism need to get serious. Transforming the legacy media's messengers of breaking news into change agents for government accountability requires an array of skills, a shift in the media industry's attitude and the coalescence of an open data community around this common goal. It also requires overcoming public desensitisation to the daily scandal, front-page corruption stories and permanent dysfunction. With this in mind, overcoming apathy will require not just a couple of data driven stories, but a structured journalism approach to covering governance consistently over time. It's no surprise that so many initiatives fail, but it is all the more important that we identify elements that work, build a new model for data journalism in developing countries and support rising stars, such as Ani Hovhannisyán, another participant in the Armenia workshop, who after hitting a wall in getting hospital data, personally went to Yerevan's hospitals to document hygiene conditions and X-ray machines, both of which she suspected are widespread health hazards in the country (2014a, 2014b). She persuaded the Yerevan municipality to hand over budgets for renovations and inspection reports that document significant public health risks. Though conditions in the clinics have not yet changed as a result of her reporting, the recent public demonstrations against an energy price by the Russian-owned distributor gives her hope that change is possible and she is pursuing a Fulbright Fellowship to study data journalism in the United States in the hope of returning and continuing her campaign to reform the country through data journalism.

## References

- Adair, Bill, Kamalakanthan, Prashanth and Stencil, Mark (2014) *The Goats Must Be Fed: Why digital tools are missing in most newsrooms*. Atlanta, Georgia: Duke Reporters' Lab at the DeWitt Wallace Center for Media & Democracy in the Sanford School of Public Policy, available at <http://www.goatmustbefed.com>, accessed 26 August 2015
- Benequista, Nicholas (2015) 'Journalists Covering International Development Don't Get this One Thing'. Medium, 17 April, available at <https://medium.com/@benequista/journalists-covering-international-development-don-t-get-this-one-thing-7b94d697d761>, accessed 26 August 2015
- Bradshaw, Paul (2015) 'Data journalism isn't just a technical skill – it's a cultural one too', *Online Journalism Blog*, 24 June, available at

<http://onlinejournalismblog.com/2015/06/24/data-journalism-isnt-just-a-technical-skill-its-a-cultural-one-too>, accessed 26 August 2015

Cook, Lindsey (2015) 'Why Journalism Students Don't Learn CS', *Source Open News*, 7 May, available at <https://source.opennews.org/en-US/learning/journalism-students-and-cs>, accessed 26 August 2015

DW Akademie (2014) *DW Akademie Blog*, 31 January, available at <http://ddjtraining.org/blog>, accessed 26 August 2015

Fojo Media Institute (2012) 'Moving away from "parachute" journalism training in South Sudan' 4 September, available at <http://lnu.se/1.73416/moving-away-from-parachute-journalism-training-in-south-sudan>, accessed 25 August 2015

Gray, Jonathan (2015) 'Democratising The Data Revolution: A Discussion Paper'. Cambridge, United Kingdom: Open Knowledge Foundation, available at <https://assets.okfn.org/files/reports/DemocratisingDataRevolution.pdf>, accessed 26 August 2015

Hovhannisyan, Ani (2014) 'Expired X-ray Machines: Dangers of Radiation a Shut Topic in Armenia.' *Hetq Online*, available at <http://hetq.am/eng/news/59954/expired-x-ray-machines-dangers-of-radiation-a-shut-topic-in-armenia.html>, accessed 26 August 2015

Hovhannisyan, Ani (2014) 'Yerevan's Polyclinics: Take Advantage of the Free Health Care but Stay Away from the Toilets.' *Hetq Online*, available at: <http://hetq.am/eng/news/58378/yerevans-polyclinics-take-advantage-of-the-free-healthcare-but-stay-away-from-the-toilets.html>, accessed 26 August 2015

International Center for International Journalists (2015) 'Fatal Extraction', 10 July, available at <http://www.icij.org/project/fatal-extraction>, accessed 26 August 2015

IREX (2014) 'Chinese Journalists and Programmers Crunch Numbers' in IREX, 13 May, available at <https://www.irex.org/news/chinese-journalists-and-programmers-crunch-numbers>, accessed 26 August 2015

Juma, Mercy (2014) 'When the Sun Sets in Turkana: Hunger Stakes and Stripes in Turkana' in *NTV*, 21 January, available at <https://www.youtube.com/watch?v=Ga8CEYVALo4>, accessed 26 August 2015

James, Kyle (2014) 'Getting online data to offline communities', *OnMedia Deutsche Welle Akademie*, 19 August, available at <http://onmedia.dw-akademie.de/english/?p=20775>, accessed 26 August 2015

Léchenet, Alexandre (2015) *Global Database Investigations: The role of the computer-assisted reporter*. Oxford, United Kingdom: Reuters Institute Fellowship Paper, University of Oxford, available at <http://reutersinstitute.politics.ox.ac.uk/publication/global-database-investigations>, accessed 26 August 2015

Looney, Margaret (2014) 'Africa's newsrooms experiment with charging for data', *IJNet*, 12 March, available at <https://ijnet.org/en/blog/africa%E2%80%99s-newsrooms-experiment-charging-data>, accessed 26 August 2015

Mahihu, Christine M and Mutuku, Leonida (2014) *Open Data in Developing Countries: Understanding the Impacts of Kenya Open Data Applications and Services*. Nairobi, Kenya: iHub Research, available at <http://www.opendataresearch.org/sites/default/files/publications/ODDC%20Report%20iHub.pdf>, accessed 26 August 2015

Organized Crime and Corruption Reporting Project (2014) 'The Russian Laundromat', 14 August, available at <https://www.reportingproject.net/therussianlaundromat>, accessed 26 August 2015

Stray, Jonathan (2015) 'The Editorial Product' in *Tinius Trust*, 27 May, available at <https://tinius.com/blog/the-editorial-product>, accessed 26 August 2015

Susman-Peña, Tara (2015) *Understanding Data: Can News Media Rise to the Challenge?* Washington, DC: Center for International Media Assistance, National Endowment for Democracy, available at <http://www.cima.ned.org/publication/understanding-data-can-news-media-rise-to-the-challenge>, accessed 26 August 2015

Wafula, Paul (2013) 'Change for Health' in *The Standard*, 12 November, available at <http://www.standardmedia.co.ke/health/changeforhealth>, accessed 26 August 2015

Wafula, Paul (2014) 'Poverty Fund: Sh600m paid to "ghosts"' in *The Standard*, 17 January, available at [http://www.standardmedia.co.ke/?articleID=2000102507&story\\_title=poverty-fund-sh600m-paid-to-ghosts](http://www.standardmedia.co.ke/?articleID=2000102507&story_title=poverty-fund-sh600m-paid-to-ghosts), accessed 26 August 2015

Wafula, Paul (2015) 'Treasury, CRA on collision path over new formula to allocate Sh6b equalisation cash' in *The Standard*, 12 May, available at <http://www.standardmedia.co.ke/business/article/2000161913/treasury-cra-on-collision-path-over-new-formula-to-allocate-sh6b-equalisation-cash>, accessed 26 August 2015

# Baby steps: a slow start but data journalism in India is gathering pace

**Sanjit Oberai, data editor of Quintillion Media and one of the early pioneers of data journalism in India, talks to leading Indian data enthusiasts about their efforts to found various start-up data journalism initiatives. They have had mixed results but all are positive for the future**

## **Introduction**

Data journalism is at a nascent stage in India. There are currently very few websites that are operating in this area, and news organisations still have a long way to go to reach the global benchmark. The good news is that it has picked up and caught the attention of the Indian government, news organisations, analytics companies, and data visualisation companies. Data can be used to bring about accountability and transparency among people and that awareness is catching people's attention. It is also fast becoming an important part of the armoury of newsrooms in India and plays an important role in multimedia and digital reporting.

The concept of 'open data' is a relatively new phenomenon, and is gathering momentum as many governments voluntarily decided to release data to the public. While India may have been one of the first few countries to join the open data movement after the US and the UK, the idea of open data evolved over a period via the National Data Sharing and Accessibility Policy (NDSAP), since March 2012. This initiative gathered pace when the National Informatics Centre (NIC), in collaboration with the US government, created an Open Government Data Platform, [data.gov.in](http://data.gov.in), as an open source portal to provide single point access to all the datasets published by different government departments in open format. The NIC is part of the Indian Ministry of Communications and Information Technology. It assists in implementing information technology projects, in collaboration with central and state governments.

Further, the 'Digital India' initiative under Prime Minister Narendra Modi is expected to help transform India into a digitally empowered society and knowledge economy. Digital India is an initiative of the Government of India to integrate government departments and the people of India. It aims to ensure that government services are made available to citizens electronically by reducing paperwork, and to connect rural areas with high-speed internet networks.

Following these moves at the federal level, a number of start-ups and other media organisations have begun to explore data storytelling and visualising data, with varying successes. In the following chapter, I discuss with leading Indian digital pioneers their experiences to date in being a part of these initiatives.

### **How India Lives - *John Samuel Raja***

There are two types of data journalism: the first looks at data as a unit of information, processes the data and get interesting analysis out of raw data; and the second uses published data reports like national crime statistics or government survey reports to write stories around them. The second type of data journalism has picked up in India quite rapidly, with many media houses - both traditional and new ones - doing such stories on a regular basis. This is welcome because it will bring insights into reports and surveys that were not previously reported extensively. However, data journalism that involves working with raw data and converting that data into interesting analysis hasn't picked up. That's because newsrooms don't have the resources or the required training skills to execute this kind of journalism. One would need a person with programming skills who can help scrape data and arrange in a database, and then write code to visualise the data. A newsroom where a programmer works alongside journalists, and who understands the data, is needed to execute such work. With government moves to increasingly make data available, there is immense scope to do such type of data journalism. Such moves increase transparency, as we will be making it available to public so that anyone can make sense of it.

#### *How India Lives started*

How India Lives is a Delhi-based start-up. The internet-based application aims to organise a massive amount of public data on India, and make it available in a searchable, comparable and visual format. The idea is to offer something of value to everyone who uses public data, be it for decision-making - like the company executive, the government official, the researcher - or for information-seeking. It was founded by a team of journalists with more than 60 years collective experience and has a technology team that has executed large data projects for Mint, Shine.com, The Caravan magazine and others.

#### *What 'How India Lives' does*

How India Lives' efforts are two-fold. The first is to make data easy to search, compare and visualise. We feel there is a huge gap between people who know

how to access data, and those who don't. We would like to bridge this gap. We do this in a number of ways, namely:

- bringing data from different sources into a single database and making different datasets compatible so they can be cross-referenced
- enabling tag-based search for user friendliness
- visualising data for easy understanding

By making data available from different sources in an easily searchable manner, we believe more journalists will start using data. By making public data datasets available through data interactives, greater numbers of ordinary citizens can also access stories they are interested in, and be informed. For example, we used vehicle theft data for Delhi and visualised it like a dashboard (How India Lives, 2015). We believe the way forward is to build communities and use them to gather data.

#### *Challenges we face*

The biggest is data cleaning. Data is in difficult formats - often in scanned PDFs, so it takes the majority of our time just to clean the data. Second, a significant majority of people in India only have access to the internet via mobile devices, and most smartphones have limited storage capacity. This means there is no space to download more apps. So the importance of a mobile-friendly website is vital. And on top of that, presenting data interactives and visualisation in mobile-friendly formats is difficult.

#### **Health Analytics - Syed Nazakat**

Health Analytics India is a data journalism initiative dedicated to providing the most illuminating reporting on healthcare. We're not going to blindly accept the data, but we're not going to be blind to it either. We crunch the numbers, investigate the issues behind the numbers and turn them into facts and figures based stories that matter to people. Our aim is to make Health Analytics India a single-point source of healthcare data and information in India.

Our initial challenge was to set-up the infrastructure and to build the team. Designing and curating website for data visualisation and analytics was also quite a challenge. Health, as a subject, was chosen because it touches us all. And there is so much scope to improve the health reporting in India. Millions of people are dying in this country from the diseases, which are totally preventable. Lack of health facilities is such a shameful story. Yet the stories hardly make news. Our challenge remains to build engagement with audiences and to present data to them in a way that will help them to understand complex stories. We're conscious of data overload so our daily challenge is to handle and interpret large data.

One of the biggest challenges is to find valid and up-to-date data. We collect data on the healthcare system by searching for the different studies or reports of

the union health ministry, or from 29 state health ministries, or from other international organisations such as WHO and research institutes. Sometimes we're surprised by how much you are able to find. But then there are vast data gaps, which leave you with an impression that nobody really knows anything concrete about healthcare data in India.

Recently while doing a story about death of rabies, we were told that the union health ministry does not collect data on its own. The data comes from 29 states and seven union territories. But in many states there were no cohort studies; community-based studies; and until 2014 one of India's biggest states Punjab hadn't even a department to collect rabies data. Elsewhere, in many states, there are proper rules and guidelines about data collection – but there is no data.

Data journalism is one way in which the media can help the cause of information, transparency, watchdog reporting and quite legitimately hold the government to account. However the lack of data and loopholes in data collection are problematic.

### **Factly - *Rakesh Dubbudu***

Factly is a platform that brings various aspects of life that directly or indirectly affects the common man but with one major difference: each news story on Factly is backed by factual evidence/data that is either available in the public domain or that is collated/gathered/collected using tools such as the Right to Information (RTI) Act.

#### *Motivation behind launching Factly*

In the last 10 years of my experience in this space, I came across a lot of government (public) data. A lot of that data is extremely important to the public. But because this data was not easily accessible and even in cases where it was, it was not easily understood. Hence public data was anything but 'public'. Misinformation and rumour also frequently trend on social media, but little of it is substantiated. Wrong data/information or a morphed picture that might be sensational has a greater chance of being shared by many people on social media than genuine information. This was the motivation behind Factly, to make public data more meaningful to the public and encourage them to look for facts or genuine data.

All our data comes from government sources such as government websites, answers to questions asked in parliament, government reports and for some, we use RTI. We zero in on a topic and then find relevant data. We also did not want to stop at making data alone meaningful; we wanted to make government information in general more meaningful to the public. We plan to explain policies/laws that are relevant to people in simple language that can be understood by everyone. We are already doing that to a certain extent.

*How is it going to be helpful to people?*

We have three aims when producing stories based on official data. The first is to make people more knowledgeable about issues with relevant data and information. That is what Factly is doing right now by packaging data in easily understandable stories, well designed infographics and easy to understand visualisations. The second is to mobilise people using this knowledge; and the third is to engage with them. We are currently at the first stage and are working towards introducing tools and other features that can take us to the second and third stages. There are many cases where our data stories have been very useful to people. By mobilise, we essentially wish to inspire people to take action or engage with the system based on the data or information. It could be a local issue to do with how their funds were spent or a state or national issue. We do not want to lead movements ourselves, but be a force that can inspire the activism and subsequent engagement with the system.

For example, we explained how petrol and LPG are priced, which was a big mystery to many. We also explained surrounding planning and land purchases so that the public could easily understand them. Examples of these stories can be found at <https://factly.in/petrol-price-breakup-in-india-infographic> and <https://factly.in/9-point-check-list-for-buying-agricultural-land-in-telangana-ap>.

*Future plans?*

We are working on building some data tools and information products. We hope to be ready with the first ones by the end of this year. With all these, we want to expand our base, encourage more people to engage and mobilise for causes based on this data. We are also looking at options to take this offline for those who do not have access to the internet by partnering with a few non-profits that work on the ground. For example we would like to make templates available for stories such as the performance of MPs (see this link <https://factly.in/sachin-tendulkar-mp-progress-report-infographic>); or local spending on village budgets so that anyone can select their MP or village and download copies of the same. These can be printed and be used for offline work. We wish to do similar work in other areas such as governments and schools and tie up with NGOs that work in those local areas.

**IndiaSpend - Govindraj Ethiraj**

IndiaSpend is the country's first data journalism initiative whose vision is to improve the quality of public discourse by using data to write stories in areas of public interest. In March 2014, the same team also launched [www.factchecker.in](http://www.factchecker.in), a dedicated fact checking initiative that examines statements and assertions made by individuals and organisations in public life for both accuracy and context. Both initiatives have a strong social media presence. IndiaSpend's articles are now distributed to India's leading newspapers, magazines, television stations, online dailies and wire services and area usually cited in at least a dozen major media platforms daily. IndiaSpend/FactChecker are registered as non-profit



organisations. They analyse government policy on issues such as the economy, education, health, agriculture and security. The team have reported on a number of important stories, such examples include:

- Why child rapes have soared 151 per cent in five years: This article looks at the sharp increase in registered child rapes in India and the states that have had the highest cases. In some nine of ten cases, victims knew their attackers. The article is available here: <http://www.indiaspend.com/cover-story/why-child-rapes-have-soared-151-in-5-years-25891>
- How 46 million Indians are being slowly poisoned: This article explores why millions of Indians are exposed to contaminated water, which could lead to serious health issues such as crippling skeletal damage, kidney degeneration, cirrhosis of the liver and cardiac arrest. The article is available here: <http://www.indiaspend.com/cover-story/how-46-million-indians-are-being-slowly-poisoned-40865>

### **DataMeet India - *Nisha Thompson***

DataMeet is a community that started in 2011 on a Google group, and now has more than 1,000 people throughout India coming together to discuss data issues and civic-minded topics. DataMeet has 6 chapters - Bangalore, New Delhi, Mumbai, Ahmedabad, Pune, and Hyderabad - who meet monthly offline to discuss, learn, share experiences and skills and also organise events.

DataMeet hosted the first Bangalore Open Data Camp in India in 2012, and since then the yearly Open Data Camp has been a great way to bring together people in the data civic space. Since then Hyderabad and New Delhi have had Open Data Camps to organise communities in their cities. These camps are great venues to discuss large ideas and problems encountered in different sectors. At the 2014 Bangalore camps that focused on elections, the community came together and shared Assembly and Parliamentary boundary shape files. The 2015 camp focused on education, and it brought civil society and government together to try to work out issues around education data in India.

DataMeet also has worked with the government on how to open more data in India. After the passage of the National Data Sharing and Accessibility Policy in 2012, we gave them feedback on standards and also implementation issues. When Data.Gov.In - India's official open data portal - launched, we worked with the officials to make sure they were aware of what high priority datasets were. For example we had requested that the Census be available and currently it is available on the portal.

Part of the open data movement is to raise data literacy. DataMeet has hosted and supported data training events for journalists, with partners like Oorvani Foundation and The Hoot. We also got involved in data expeditions on urban data along with Hyderabad Urban Labs, and other events with Field of View and

IIIT Bangalore. These events help introduce data concepts and skills to people who want to learn more and use data more often.

### **Conclusion?**

Clearly 'better late than never' is an apt description for India with respect to data journalism. Many of the initiatives that have started are making their presence felt in their various areas of expertise. However, this is just the start and as people start recognising the importance of data, it will soon be the sought after area for decades to come.

### **References**

- Data Meet Bangalore, available at <http://datameet.org/chapters/bangalore/>, accessed 16 August 2015
- Factly India, available at <https://factly.in/team/>, accessed 16 August 2015
- Government of India (2012) 'National data sharing and accessibility policy – 2012 approved', available at <http://pib.nic.in/newsite/erelease.aspx?relid=80197>, accessed 16 August 2015
- Health Analytics India, available at <http://healthanalyticsindia.com/>, accessed 16 August 2015
- How India Lives, available at <http://howindialives.com/>, accessed 16 August 2015
- How India Lives (2015) 'Which are the most stolen cars, bikes and scooters in Delhi?', available at <http://howindialives.com/testchart/dlicrthft/>, accessed 16 August 2015
- India Spend, available at <http://www.indiaspend.com/>, accessed 16 August 2015
- National Informatics Centre Data Portal India, available at <http://www.nic.in/projects/data-portal-india>, accessed 16 August 2015
- Parihar, Isha (2015), 'On the road to Open Data: glimpses of the discourse in India' *World Bank Blog*, 17 February, available at <http://blogs.worldbank.org/ic4d/road-open-data-glimpses-discourse-india>, accessed 16 August 2015
- Singh, Shreya (2013) 'The beginnings of India's open data movement', available at <http://www.igovernment.in/opinion/20131/the-beginnings-of-indias-open-data-movement#sthash.eUJxO8NM.dpuf>, accessed 16 August 2015