

Buzzing about data journalism

By Raymond Joseph

Anyone who is the least interested in new developments in media cannot have missed the buzz being generated around “data journalism” on Twitter, G+, Facebook and any other forums where journalists hang out.

Data journalism, along with the burgeoning of social media platforms with useful reporting applications, is helping change the face of media in this second decade of the 21st century. Leading the field internationally are publications such as *The New York Times* and *The Guardian*, while more and more journalists in the United States and Europe are cottoning on to the possibilities that this ability to access and interrogate data in ways not possible before helps deepen reporting and enhance storytelling abilities, both on and offline.

Driven by demand, universities, colleges and other institutions that teach journalism are increasingly offering courses in data journalism, incorporating web scraping and basic coding, as part of the curriculum.

But in South Africa things are moving at a snail’s pace, with only Wits University offering formal training in this field, although Google and Wikiscraper have also done some ad hoc training in the uses of their tools.

HacksHackers, an international organisation with chapters in most major cities around the world that brings journalists (hacks) and coders (hackers) together, has stepped into the breach to try and inculcate a culture of data journalism in South Africa’s newsrooms. Chapters have already been founded in Johannesburg and Cape Town, creating a space for journalists, coders, data experts and other likeminded people to network

and collaborate on data-driven projects. Several hackathons – one- or two-day gatherings at which coders, journalists and activists, have collaborated to develop useful apps and tools – have already been held.

But few South African journalists, other than a handful of pioneers, like Media 24’s award-winning Investigations Unit led by part-journo-part-geek Andrew Trench, are using it as part of their reporting toolkit on a regular basis.

Part of the problem is a generally technophobic South African journalism community that still struggles to get their heads around anything more than the basics of the growing array of social media platforms that are changing the way journalism is practised.

To be clear, no-one is suggesting that all South African journalists should strive to become expert coders and programmers in order to practise data journalism. My mantra as convenor of HacksHackers Cape Town is a very simple one: you don’t need to be a mechanic to drive a motor car; you just need to know a mechanic to service and repair your vehicle. I, for one, cannot hack code, but I have learned enough to be able to communicate what I want to someone who can, when working on data-driven reporting projects.

And it is in that collaboration between people with different skills sets that the real power of data journalism lies. It has upped the game in newsrooms where data journalism is practised, with journalists, coders and data wranglers (geeks who scrape websites and then clean “liberated” data sets) and graphics artists collaborating on ways of telling stories in interesting

And it is in that collaboration between people with different skills sets that the real power of data journalism lies. It has upped the game in newsrooms

Another major hurdle to data journalism is the almost universal lack of access to public data paid for from the public purse, which is all too often claimed as being copyright protected.

and innovative ways, including the creative use of infographics.

Another major hurdle to data journalism is the almost universal lack of access to public data paid for from the public purse, which is all too often claimed as being copyright protected. And even when it is available, it is inevitably not accessible in a user-friendly, machine readable format like Excel or CSV. More often than not it is only available as user-unfriendly PDFs where, data journalists will tell you only half jokingly, data goes to die. While it is possible to extract data from PDFs, it is time-consuming and, in the absence of access to the original raw data, only yields what the person who made it chose to include.

Writing on *Daily Maverick*, Adi Eyal, a leading Open Data advocate who heads up the CodeforSA initiative, says: "It seems somewhat absurd that publicly funded institutions in South Africa should be allowed to copyright data produced using public funds." He cites the example of the publicly-funded Municipal Demarcation Board, that last year received R38.5-million from the National Treasury.

"Unfortunately there is a lot of commitment to our data and it is copyright to us and we cannot allow you to use it commercially at all," Eyal was told after he asked whether he could download data from their site and use it for commercial purposes.

There are other examples of public-funded data that is closely guarded and difficult to access, like the City of Cape Town, which manages a very rich dataset on municipal valuations, and the National Assembly, which hosts an "intriguing database" on gifts received by parliamentarians and other elected officials. But, "we are explicitly excluded from using this data for anything but personal use," says Eyal.

When I approached the City of Cape Town some time back about getting access to data on toilets, especially the number of people still using the infamous bucket system, I didn't get very far as they were unsure of its accuracy and were reluctant to hand it over.

This is a far cry from the United States where the White House recently released an executive order signed by President Barak Obama committing all federal government data to be open and machine-readable.

"Making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation," reads the White House executive order issued in May 2013.

I constantly find myself frustrated at the excuses I hear in conversations with South African editors and journalists when I try to persuade them why data journalism is important. Too often it's a case of "we don't have the resources or manpower" or variations on those themes, rather than "how can we make this happen?"

So the reality for now is that until media houses invest in equipment, training and hiring people with the right skills sets, the mainstreaming of data journalism in South African will remain a distant dream and the domain of a handful of determined pioneers.



Raymond Joseph is a freelance journalist, journalism trainer and media consultant, specialising in content offerings and "fixing" under-performing news rooms. After a decades-long career in print journalism he has made the transition to the digital world and also convenes HacksHackers Cape Town. Contact him at rayjoe@iafrica.com and on Twitter as @rayjoe

USEFUL LINKS

- How to get started in data journalism: www.goo.gl/mDh4f
- Data Journalism Handbook: <http://datajournalismhandbook.org>

FREE DATA TOOLS

- Easy to use open source data scraper: <http://www.haystaxdata.org/>
- An excellent, free and simple to use dataviz tool: <http://datawrapper.de/>
- This searchable resource lists NGOs and NPOs in South Africa by sector, with contact details, and was created in a day at a HacksHackers Cape Town hackathon: <http://carolune.org/c4d/index.php>
- This photo editing tool was created at a hackathon in New York: <http://evanw.github.io/webgl-filter/>

SOME GREAT EXAMPLES OF WHAT CAN BE DONE WITH DATA VISUALISATIONS

- World arms trade: <http://workshop.chromeexperiments.com/projects/armsglobe/>
- Charting the wind: <http://www.economist.com/blogs/graphicdetail/2012/08/charting-wind?fsrc=scn/tw/dc/windmap>
- How Twitter reacted to the resignation of former Egyptian president Hosni Mubarak: <http://gephi.org/2011/the-egyptian-revolution-on-twitter/>
- Examples of data stories from *The New York Times* and *The Guardian*: <http://www.marijerooze.nl/thesis/graphics/?paper=All>
- Finalist stories for the Data Journalism Awards 2013: <http://www.globaleditorsnetwork.org/dja/>

EXAMPLES OF SA DATA STORIES/ VISUALISATIONS

- Journalism salary survey undertaken by HacksHackers Cape Town to illustrate data mining, cleaning and visualisation: <http://salarysurvey.hackshackers.co.za/>
- Visualisation: battle for the 2016 municipal elections <http://za.okfn.org/2013/03/23/battle-for-local-municipalities/>
- Visualisation: battleground wards in Cape Town <http://za.okfn.org/2013/03/22/battleground-wards-in-cape-town/>

OPEN DATA RESOURCES (COLLECTION OF CLEAN, LIBERATED DATA SETS)

- African open data: <http://opendataforafrica.org/>
- Liberated South African data: <http://africaopendata.org/group/south-africa>