

027

Journalists and the media

Alex Howard and Eva Constantaras

Key points

- Data journalists have a key role to play as public interest watchdogs. The label “data journalism” can cover a wide range of practices from using data science to find stories to storytelling with data and creating interactive content and visualisations for articles.
- The costs and complexity of effective data journalism, combined with the time pressures common in reporting, make for a difficult business case. Finding sustainable models of data journalism is even more urgent at a time when traditional media outlets face competition from online journalists.
- The promise of “automated journalism” based on open data is largely unfulfilled; however, if more media houses focus on making open data an essential source, we may see more examples of automation tools in the future.
- This is a critical moment for public trust, and there is no clear template for how data journalism can contribute to society’s response.

Introduction

The social impact of open data depends upon its use and reuse by different actors across society. For decades, journalists and media outlets have served and informed the public as a key infomediary for government statistics, academic and scientific research, business information, and their own analysis.

In the 21st century, data journalism is now an emerging field of practice around the world, but there is still no universally shared definition. Some practitioners consider data journalism to be thoughtful storytelling with data as the centrepiece. Others say it is including statistical data in stories, visualisations, interactive elements, or the application of data science to journalism.



Journalists who acquire, clean, analyse, report on, and, in some cases, create open data, now play a key role as public interest watchdogs. Reporters can transform the raw data into new insights, facilitate public engagement in the democratic process, inform consumers, and hold powerful institutions accountable. Journalists have made opaque financial transactions more visible, catalysing regulatory reform, such as with the International Consortium of Investigative Journalists (ICIJ)¹ collaborative reporting on massive leaks of data. Reporters at the *Washington Post* and *The Guardian* have created a database of undocumented killings by law enforcement,² raising awareness of racial injustice and shaming government agencies into keeping better statistics. In “Life in the camps”,³ Reuters journalists mapped the abhorrent living conditions of refugees, revealing the failures of global diplomacy.

History

The computer-assisted reporting of the last century, when the tools and methodologies of social science were first applied to journalism using mainframe and then desktop computers, has been transformed by the rapid public adoption of broadband internet connections, smartphones, open source software and frameworks, and the unprecedented scale of data generation and publication. While the data-driven journalism of 2019 is distinguished from what has come before by computers and the internet, journalists have been reporting on data as a source, the most basic definition for “data journalism”, for centuries, printing statistics and tables in pamphlets, tabloids, and newspapers long before “open data” was an idea.

Around the world, a growing number of media organisations are downloading, analysing, and reporting on open data from multiple sources in ways that inform, engage, and empower the public. Over the past decade, pioneers in the global transparency movement have adopted and adapted principles and practices from the open source software world, where “showing your work” and collaborating around a shared code are important signals for both trust and transparency.

At their best, journalists are as appropriately sceptical of the open data published by corporations and governments online as they would be of the accounts from human sources.

Expanding horizons

The geographic reach of data journalism has grown over the past decade, driven by foundational investments, industry adoption, and grassroots organisations training reporters on the ground. The data journalism community, traditionally centred around the annual National Institute for Computer-Assisted Reporting⁴ (NICAR) conference run by investigative reporters and editors, now hosts a much more cross-border conversation.

The Global Editors Network⁵ and their annual Data Journalism Awards, the European Investigative Journalism & Dataharvest Conference,⁶ the Global Investigative Journalism Conference,⁷ and a plethora of smaller country-based data journalism events organised under the auspices of Hacks/Hackers or Schools of Data, are convening communities of data journalism around the world.

The growing diversity in the practice has opened new “scrapper” models for data journalism with a more explicit public interest mission in countries where small teams of data journalists expose blatant inequality and corruption.

Enduring challenges to the field

Like journalism that relies upon humans as a source, there is no global consensus that the purpose of data journalism is to inform the public about governance issues, reveal corruption, or hold powerful institutions accountable. Despite the extraordinary work produced by publications around the world, some recent research about the state of data journalism paints a grim picture of the field.

Google News Lab’s *Data journalism in 2017*⁸ found that data skills remain specialised and time-consuming to develop. Time pressures prevent aspiring journalists from producing the data-driven stories they believe need reporting. While some interactive features that engage readers have proven wildly popular, the business case for media houses to invest in data teams is still a hard sell in boardrooms.

The editorial decisions of organisations also determine where the lens of data-driven reporting is directed. When researchers at the University of Hamburg evaluated the projects that were nominated for the Global Editors Network’s annual Data Journalism Awards, they found that most of the coverage is on political rather than social issues, and that most of the data used was disclosed by official sources rather than collected by reporting teams.

The failure of data journalism to become just journalism

The data journalism community has exploded in the last decade with more and more media houses now including at least one data journalist on their staff. Yet six years after Simon Rogers, founder of *The Guardian DataBlog*,⁹ declared “Anyone can do it. Data journalism is the new punk”;¹⁰ everyone is *not* doing it. While every reporter now uses a computer and a smartphone in their reporting, the “nerds” have not taken over the newsroom.

Despite the potential of data journalism to harness open data to provide critical context, insight, and actionable information in times of information crisis, some media outlets are turning away because it is difficult, slow, and expensive.¹¹ Even after massive projects like the Panama Papers¹² prompted political upheaval and an examination of global financial oversight, some publishers remain reluctant to take the risk of investing in their own data team.

The unfulfilled promise of automation

A universal complaint among data journalists is the lack of quality, reliable, timely, and disaggregated open data, which, in turn, inhibits opportunities for its use in local journalism. Stakeholders emphasise that government failures to facilitate access to structured, high-quality open data make it onerous for journalists to keep datasets updated or to build relevant apps to monitor spending, performance, and government services. They also highlight governments’ use



of loopholes and grey areas in access to information laws being used to hide data, a strategy also employed by businesses. Still, in many countries, the data journalism community has not made use of the limited statistics that are made public, and only a few journalists have requested data through official Freedom of Information Act processes.

There have, however, been some encouraging developments. The Organized Crime and Reporting Project, for instance, provides an Investigative Dashboard¹³ that centralises open data resources and helps reporters avoid repetitive scraping and cleaning. Reuters' Lynx Insights¹⁴ guides reporters to relevant data sources by "surfacing trends, facts, and anomalies in data, that reporters can then use to accelerate the production of their existing stories or to spot new ones". If more media houses focus on making open data an essential source, we may see more automation tools in the future.

Making the business case for data journalism

The sustainability of data journalism is all the more urgent as more legacy media outlets struggle for survival in the digital age. Around the world, only a few newsrooms have made an effort to ensure reporters understand how the business model for their newsroom works. Some outlets need to produce in-depth, value-added investigations that are not available elsewhere in order to attract and retain subscribers. Media companies that rely upon advertising need to create compelling interactives, infographics, and visualisations that attract and retain large audiences.

Economic challenges within the industry have also led to competing priorities within newsrooms with editors emphasising urgency of delivery and data journalists seeking more time to produce complex investigations with statistical rigour. In that context, there is also an enduring debate about whether public interest journalism should move to a non-profit model entirely. If more publishers regarded data, and staff with the skill to report on it, as more of an asset and not as a resource drain, data journalism could fulfil an essential role in the transparency and accountability cycle.

While some outlets, such as ProPublica,¹⁵ have established data stores to try to monetise the data that they have obtained and cleaned, this is still a nascent strategy. Other outlets are experimenting with offering personalised news apps with mainstream appeal to subsidise their work on longer-term investigations.

Many of the current challenges faced by stakeholders across the journalism world, including the financial pressures that have resulted from the massive disruption of the business model for traditional publishers and the dominance of tech companies in the online advertising markets, stem from a crucial failure to make data an essential, strategic resource for all the journalism produced by outlets.

Inclusion and accessibility for data-driven democracy

With the rise of nationalist, populist political movements in many countries, an increasing number of members of the global journalism community are exploring how to produce more accessible journalism by, about, for, and *with* marginalised communities.

OpenNews¹⁶ has taken the lead in this work, proposing that “a diverse community of peers working, learning, and solving problems together can create the stronger, more representative ecosystem that journalism needs to thrive”. These efforts include creating opportunities for journalists of colour and women within an industry that traditionally has been primarily white and male. It also means having a greater focus on supporting local media in rural areas to produce data-driven content relevant to their communities. ProPublica’s Local Reporting Network¹⁷ and the Bureau Local¹⁸ are both relevant efforts to accelerate such content production.

In some developing countries, both legacy print and startup news organisations are experimenting with harnessing data to make injustice, inequality, and discrimination measurable, visible, and solvable. Data teams at La Nación Data¹⁹ in Argentina, IndiaSpend²⁰ in India, and Nation Newsplex²¹ in Kenya are devoting themselves almost exclusively to analysing open data to explain policy issues that affect ordinary citizens. Even much bemoaned technical obstacles like weak content management systems, low internet speeds, mobile display limitations, and low-quality data sources have not stopped these journalists from pushing for better socioeconomic policies and evidence-based decision-making.

Nation Newsplex: Inclusivity and an alternative lens

The Nation Newsplex, the data team of the *Daily Nation*, the highest circulation daily newspaper in Kenya, used data to focus its election coverage on policy not politics during the last presidential elections. Nation Newsplex has a rare combination: a public interest data-driven mandate and an audience with a diverse background. Whether fact-checking economic issues related to small business growth,²² foreign direct investment,²³ and unemployment;²⁴ examining public services related to health²⁵ and food security;²⁶ or analysing compensation for internally displaced people,²⁷ the focus of reporting and analysis was on the welfare of citizens, not on the politicians involved or the populist rhetoric that dominated mainstream coverage.

Another common frustration expressed by data journalists in both Western and non-Western newsrooms is the dearth of professional growth opportunities and clear advancement paths. This may play a role in dissuading professionals from more socioeconomically diverse backgrounds from pursuing careers in the media, which, in turn, has negative effects on the industry and leads to errors or blind spots in the journalism itself.

While reporters with data skills continue to be in high demand in many newsrooms, peer-to-peer learning and online courses remain insufficient to meet the growing demand for more advanced data journalism skills. The soaring popularity of the annual NICAR conference in recent years, which features many hands-on workshops, speaks to the demand for skills development. Unfortunately, entry-level data journalists complain of being pigeonholed into roles such as simply creating visualisations for more senior reporters or managing websites instead of producing their own data journalism. Mid-career professionals bemoan the lack of a clear career ladder as many senior-level editors are typically selected from the traditional newsroom.



Data-driven beat reporting

In one of the poorest areas of Pakistan, where self-censorship is rampant, an online publication, *Balochistan Voices*, used data to take on a corrupt, inefficient government.²⁸ In “Balochistan government builds roads while people sink further into poverty,”²⁹ Adnan Amir explored Balochistan’s corrupt procurement process. He showed how siphoning off public funds had left Balochistan’s citizens sicker, poorer, and less educated than their neighbours.

In a follow up story, “In Balochistan, five times more people die in highway accidents than suicide blasts,”³⁰ published in late February, he offered the government an easy way to save face and improve access to healthcare. Sure enough, the National Highway Authority signed a Memorandum of Understanding³¹ on 18 March to establish five trauma centres along Balochistan’s highways.

Reusability and archiving

An underreported challenge to measuring the evolution of data journalism over the last ten years is the lack of consistent archiving. Interactive journalism, created using a plethora of proprietary tools, technologies, and data formats across the industry, has been inconsistently stored and maintained by publishers, unlike the print and microfiche archives of traditional journalism.

Many data-driven interactives now no longer work or have even disappeared entirely. Archiving news apps and visualisations presents a particular challenge, because they depend on tools and systems that may be managed by external entities. Without a consistent conservation strategy, this first draft of history will be lost.³²

In many newsrooms, data units are simply too short-staffed to address pressing issues like archiving or data reusability. The lack of developers or data scientists in the newsroom to clean and maintain key datasets, preserve projects, and ensure data reusability prevents many media houses from investing in the very systems that would make their journalism faster, more efficient, and more impactful.

Journalists have sought to harness open data to inform the public about institutionalised discrimination, inequality, and oppression in an attempt to uncover issues that were invisible before, and now, may become invisible again if journalists stop counting the causes, the victims, and the wasted funds, or if the projects disappear from the internet entirely. Consistent coverage driven by independent editors combined with effective archiving has the potential to foster systemic change.

Data reusability

In the United States, ProPublica was founded to “expose abuses of power and betrayals of the public trust by government, business, and other institutions, using the moral force of investigative journalism to spur reform through the sustained spotlighting of wrongdoing.”³³

The independent non-profit’s digital-first journalism not only uses data as a source in its narrative reporting, but publishes the data with the stories and the code that drives the news applications that make that data more understandable.

ProPublica takes government information disclosed directly to the public online, sometimes as the result of a successful public records lawsuit and court-ordered disclosure under the federal Freedom of Information Act, cleans and structures it, and then publishes the open data, newsapps,³⁴ code, and stories online.

Notably, ProPublica partners with other news organisations and the public online in collaborative reporting projects that then produce open data, like collecting political ads on social media platforms.

A decade later, the social impact of the release of public records as open government data by federal agencies can be measured by the social impact of ProPublica’s data-driven reporting.

ProPublica’s Local Reporting Network,³⁵ launched in early 2018, seeks to address the dearth of local investigative reporting by using its expertise in open data and investigative reporting to strengthen and amplify local reporting. Investigations lowered the barrier for accessing large national and state-level datasets on issues such as conflicts of interest, housing, mental healthcare, criminal justice, and workplace safety.

The new project has a special focus on state government accountability reporting. Recent research by Poynter finds that the great majority of Americans trust their local news sources.³⁶ Making open data sources relevant to local communities is an emerging strategy for reinvigorating democratic participation at the local level through journalism.

Conclusion

In 2019, the journalism industry around the world is grappling with multiple challenges after quicksilver societal changes in how information is being created, shared, and discovered by humans everywhere.

Public trust in mass media is sinking to historic lows in the United States,³⁷ fuelled by the spectre of an American president decrying “fake news” and describing the press holding his administration accountable as the “enemy of the people”. To close that deficit, editors and producers will need to “optimize for public trust”, as New York University professor Jay Rosen has proposed,³⁸ prioritising verification, openness, listening, engaging, co-production, and diversity.



These principles echo many of the conversations and frustrations faced by data journalists who are navigating and shaping the many ways in which communities are consuming and engaging with the news.

Technology companies like Facebook, Google, and Amazon now dominate the markets for online advertising and information discovery, sharing, and distribution by operating vast digital platforms for billions of users around the world. Media companies face a dramatically different landscape for their work and its publication. News organisations need to find readers, viewers, and subscribers, and get their attention to inform and engage them. National, state, and city governments are publishing open data, usually driven by only a nominal desire to inform the public about public business.

It's a difficult historic moment for journalists and their publishers who face these challenges at a time when the industry itself has been so diminished. Tens of thousands of journalists have lost their jobs, leaving local governments without partners for their online disclosures. At the same time, far too many cities and states have engaged in “transparency theatre”, engaging in openwashing with low quality data disclosures. Officials cloak themselves in the rhetoric of “open data” while denying freedom of information requests from the press.

Worse still, less democratic states have shown how open data programmes and policies can exist without open government, where disclosures are focused on driving economic impact or even weaponised against marginalised populations. The levels of press freedom, strength of access to information laws, and the capacity of the government institutions that exist in a given country heavily influence whether journalism based upon open data will lead to positive societal changes.

The transparency and accountability that governments publishing open data cite as a key goal will only occur when journalists do not face hazards from seeking information under the law, publishing reports that reveal corruption, or bearing witness to human rights violations or abuses of power. In that context, the local journalists who report on that data should be viewed as the community's guardians, ensuring that the rights of the governed are upheld by informing the public about the public's business.

Further reading

- Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2017). The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*, 4 April. <https://journals.sagepub.com/doi/10.1177/1464884917700667>
- Baack, S. (2018). Practically engaged: The entanglements between data journalism and civic tech. *Digital Journalism*, 6(6), 673–692. <https://www.tandfonline.com/doi/full/10.1080/21670811.2017.1375382>
- Constantaras, E., Asad, J., Asma Usman, A., & Adil, M. (2018). *The promise and challenge of data journalism: Lessons from Pakistan. Reports and surveys*. Internews. https://www.internews.org/sites/default/files/2018-04/Internews_Pakistan_data_journalism_casestudy_public.pdf

Gray, J., Chambers, L., & Bounegru, L. (Eds.). (2012). *The data journalism handbook*. First edition. Sebastopol, CA: O'Reilly Media.

Howard, A.B. (2014). *The art and science of data-driven journalism*. New York, NY: Tow Center for Digital Journalism Publications.

Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2017). Data-driven reporting: An ongoing (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*, 12 October. <https://journals.sagepub.com/doi/abs/10.1177/1464884917735691>

About the authors

Alex Howard is an independent writer, digital governance analyst, and open government advocate based in Washington, DC. He is the former deputy director of the Sunlight Foundation and the author of *The art and science of data-driven journalism* from the Tow Center for Digital Journalism at Columbia University and other publications. Follow Alex on Twitter at <https://twitter.com/digiphile> or on his blog on government information technology at <https://e-pluribusunum.org/>.

Eva Constantaras is an investigative data journalist specialised in establishing data units in mainstream media in developing countries. Her teams operate in Kenya, Afghanistan, Pakistan, and Central America, covering public interest topics ranging from extractive industries and access to justice to reproductive rights and food security. You can follow Eva on Twitter at <https://twitter.com/evaconstantaras>.

How to cite this chapter

Howard, A. & Constantaras, E. (2019). Open data stakeholders: Journalists and the media. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: Histories and horizons* (pp. 395–405). Cape Town and Ottawa: African Minds and International Development Research Centre. <http://stateofopendata.od4d.net>



This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. It was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada.



Endnotes

- 1 <https://www.icij.org/>
- 2 Tate, J., Jenkins, J., Rich, S., & Muyskens, J. (2018). Fatal force: 2018 police shootings database. *Washington Post*. <https://www.washingtonpost.com/graphics/2018/national/police-shootings-2018/>
- 3 Cai, W. & Scarr, S. (2017). Life in the camps. *Reuters*, 4 December. <http://fingfx.thomsonreuters.com/gfx/rngs/MYANMAR-ROHINGYA/010051VB46G/index.html>
- 4 <https://www.ire.org/nicar/>
- 5 <https://www.globaleditorsnetwork.org/>
- 6 <https://europeanjournalists.org/blog/2018/02/20/eijc18-dataharvest-in-may-2018/>
- 7 <https://gijn.org/global-conference-2/>
- 8 Rogers, S., Schwabish, J., & Bowers, D. (2017). *Data journalism in 2017: The current state and challenges facing the field today*. San Francisco, CA: Google News Lab. <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>
- 9 <https://www.theguardian.com/data>
- 10 Rogers, S. (2012). Anyone can do it. Data journalism is the new punk. *The Guardian*, 24 May. <https://www.theguardian.com/news/datablog/2012/may/24/data-journalism-punk>
- 11 Rogers, S., Schwabish, J., & Bowers, D. (2017). *Data journalism in 2017: The current state and challenges facing the field today*. San Francisco, CA: Google News Lab. <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>
- 12 ICIJ (International Consortium of Investigative Journalists). (2018). The Panama Papers: Exposing the rogue offshore finance industry. <https://www.icij.org/investigations/panama-papers/>
- 13 <https://investigativedashboard.org/>
- 14 Bilton, R. (2018). Reuters' new automation tool wants to help reporters spot the hidden stories in their data (but won't take their jobs). *NiemanLab*, 12 March. <http://www.niemanlab.org/2018/03/reuters-new-automation-tool-wants-to-help-reporters-spot-the-hidden-stories-in-their-data-but-wont-take-their-jobs/>
- 15 <https://www.propublica.org/>
- 16 <https://opennews.org/>
- 17 ProPublica. (2017). ProPublica local reporting network selects seven newsrooms across US. <https://www.propublica.org/atpropublica/propublica-local-reporting-network-selects-seven-newsrooms-across-u-s>
- 18 <https://www.thebureauinvestigates.com/local>
- 19 <https://www.lanacion.com.ar/data>
- 20 <https://www.indiaspend.com/>
- 21 <https://www.nation.co.ke/newsplex>
- 22 <https://www.nation.co.ke/newsplex/state-of-nation-2016-fact-check/2718262-3852980-j0gsfw/index.html>
- 23 <https://www.nation.co.ke/newsplex/foreign-investment/2718262-3990224-ehbaunz/index.html>
- 24 <https://www.nation.co.ke/newsplex/mudavadi-high-unemployment/2718262-3995274-pbpw17/index.html>
- 25 <https://www.nation.co.ke/newsplex/Thika-Kiambu-hospitals-Uhuru-Kenyatta/2718262-3986010-ocrxmg/index.html>
- 26 <https://www.nation.co.ke/newsplex/weather-maize/2718262-3995278-pu22a7/index.html>
- 27 <https://www.nation.co.ke/newsplex/idp-compensation-funds/2718262-3975950-nc8j92/index.html>
- 28 Aamir, A. (2017). Balochistan schools: Failing grade despite funding spike. *Balochistan Voices*, 26 May. <http://balochistanvoices.com/2017/05/balochistan-schools-failing-grade-despite-funding-spike/>

- 29 Aamir, A. (2017). Balochistan government builds roads while people sink further into poverty. *Balochistan Voices*, 26 December. <http://balochistanvoices.com/2017/12/balochistan-government-builds-roads-people-sink-poverty/>
- 30 Aamir, A. (2018). In Balochistan, five times more people die in highway accidents than suicide blasts. *Balochistan Voices*, 26 February. <http://balochistanvoices.com/2018/02/balochistan-five-times-people-die-highway-accidents-suicide-blasts/>
- 31 <https://www.documentcloud.org/documents/4436283-NHA-MOU.html>
- 32 Boss, K. & Broussard, M. (2017). Challenges of archiving and preserving born-digital news applications. *IFLA Journal*, 43(2), 150–157. <https://doi.org/10.1177%2F0340035216686355>
- 33 ProPublica. (n.d.). About us. <https://www.propublica.org/about/>
- 34 <https://www.propublica.org/newsapps/>
- 35 <https://www.propublica.org/local-reporting-network/>
- 36 Guess, A., Nyhan, B., & Reifler, J. (2018). *All media trust is local? Findings from the 2018 Poynter Media Trust Survey*. St. Petersburg, FL: Poynter Institute. <http://www-personal.umich.edu/~bnyhan/media-trust-report-2018.pdf>
- 37 Swift, A. (2016). Americans' trust in mass media sinks to new low. *Gallup*, 14 September. <https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx>
- 38 Rosen, J. (2018). Optimizing journalism for trust. *Medium: The Correspondent* [Blog post], 14 April. <https://medium.com/de-correspondent/optimizing-journalism-for-trust-1c67e81c123>