23. Ways of Doing Data Journalism

Sarah Cohen

Abstract

This chapter explores the various ways that data journalism has evolved and the different forms it takes, from traditional investigative reporting to news apps and visualizations.

Keywords: investigative journalism, news applications, data visualization, explanatory journalism, precision journalism

data (*dey-tah*): a body of facts or information; individual facts, statistics or items of information (“Data,” n.d.)

journalism: the occupation of reporting, writing, editing, photographing, or broadcasting news or of conducting any news organization as a business (“Journalism,” n.d.)

If you’re reading this handbook, you’ve decided that you want to learn a little about the trade that’s become known as data journalism. But what, exactly, does that mean in an age of open data portals, dazzling visualizations and freedom of information battles around the world?

A dictionary definition of the two words doesn’t help much—put together, it suggests that data journalism is an occupation of producing news made up of facts or information. Data journalism has come to mean virtually any act of journalism that touches electronically held records and statistics—in other words, virtually all of journalism.

That’s why a lot of the people in the field don’t think of themselves as data journalists—they’re more likely to consider themselves explanatory writers, graphic or visual journalists, reporters, audience analysts, or news application developers—all more precise names for the many tribes of this growing field. That’s not enough, so add in anything in a newsroom that requires the use

of numbers, or anything that requires computer programming. What was once a garage band has now grown big enough to make up an orchestra.

Data journalism is not very new. In fact, if you think of “data” as some sort of systematic collection, then some of the earliest data journalism in the United States dates back to the mid-1800s, when Frank Leslie, publisher of *Frank Leslie’s Illustrated Newspaper*, hired detectives to follow dairy carts around New York City to document mislabelled and contaminated milk. Scott Klein (2016), a managing editor for the non-profit investigative site *ProPublica*, has documented a fascinating history of data journalism also dating to the 1800s, in which newspapers taught readers how to understand a bar chart. Chris Anderson also explores different genealogies of data journalism in the 1910s, 1960s and 2010s in his chapter in this volume.

With these histories, taxonomies of different branches of data journalism can help students and practitioners clarify their career preferences and the skills needed to make them successful. These different ways of doing data journalism are presented here in an approximate chronology of the development of the field.

Empirical Journalism, or Data in Service of Stories

Maurice Tamman of Reuters coined the term “empirical journalism” as a way to combine two data journalism traditions. Precision journalism, developed in the 1960s by Philip Meyer, sought to use social science methods in stories. His work ranged from conducting a survey of rioters in Detroit to directing the data collection and analysis of an investigation into racial bias in Philadelphia courts. He laid the groundwork for investigations for a generation. Empirical journalism can also encompass what became known as computer-assisted reporting in the 1990s, a genre led by Eliot Jaspín in Providence, Rhode Island. In this branch, reporters seek out documentary evidence in electronic form—or create it when they must—to investigate a tip or a story idea.

More recently, these reporters have begun using artificial intelligence and machine learning to assist in finding or simplifying story development. They can be used to help answer simple questions, such as the sex of a patient harmed by medical devices when the government tried to hide that detail. Or they can be used to identify difficult patterns, such as Peter Aldhous’ analysis of spy planes for *Buzzfeed* (Aldhous, 2017; Woodman, 2019).

These reporters are almost pure newsgatherers—their goal is not to produce a visualization nor to tell stories with data. Instead, they use records to explore a potential story. Their work is integral to the reporting project,

often driving the development of an investigation. They are usually less involved in the presentation aspects of a story.

Arguably the newest entry into this world of “data journalism” could be the growing impact of visual and open-source investigations worldwide. This genre, which derives from intelligence and human rights research, expands our notion of “data” into videos, crowdsourced social media and other digital artefacts. While it’s less dependent on coding, it fits solidly in the tradition of data journalism by uncovering—through original research—what others would like to hold secret.

One of the most famous examples, *Anatomy of a Killing* from BBC’s Africa Eye documentary strand, uncovers where, precisely, the assassination of a family occurred in Cameroon, when it happened, and helps identify who was involved—after the Cameroonian government denied it as “fake news” (BBC News, 2018). The team used tools ranging from Google Earth to identify the outline of a mountain ridge to Facebook for documenting the clothing worn by the killers.

Data Visualization

Looking at the winners of the international Data Journalism Awards would lead a reader to think that visualization is the key to any data journalism.¹ If statistics are currency, visualization is the price of admission to the club.

Visualizations can be an important part of a data journalist’s toolbox. But they require a toolkit that comes from the design and art world as much as the data, statistics and reporting worlds. Alberto Cairo, one of the most famous visual journalists working in academia today, came from the infographics world of magazines and newspapers. His work focuses on telling stories through visualization—a storytelling role as much as a newsgathering one.

News Applications

At *ProPublica*, most major investigations start or end with a news application—a site or feature that provides access to local or individual data through an engaging and insightful interface. *ProPublica* has become known for its news apps, and engineers who began their careers in coding have evolved into journalists who use code, rather than words, to tell stories.

1 See Loosen’s chapter in this volume.

ProPublica's Ken Schwenke, a developer by training who has worked in newsrooms including the *Los Angeles Times* and *The New York Times*, became one of the nation's leading journalists covering hate crimes in the United States as part of the site's Documenting Hate project, which revolved around stories crowdsourced through *ProPublica*'s news application.

Data Stories

The term “data journalism” came of age as reporters, statisticians and other experts began writing about data as a form of journalism in itself. Simon Rogers, the creator of *The Guardian*'s Datablog, popularized the genre. *FiveThirtyEight*, *Vox* and, later, *The New York Times*' Upshot became this branch's standard bearers. Each viewed their role a little differently, but they converged on the idea that statistics and analysis are newsworthy on their own.

Some became best known for their political forecasts, placing odds on US presidential races. Others became known for finding quirky data sets that provide a glimpse into the public's psyche. One example of this is the 2014 map of baseball preferences derived from Facebook preferences in the US Table stakes. The entry point for this genre is a data set, and expertise in a subject matter is the way these practitioners distinguish themselves from the rest of the field. In fact, Nate Silver and others who defined this genre came not from a journalism background, but from the worlds of statistics and political science.

Amanda Cox, the editor of *The New York Times*' Upshot, has said she sees the site's role as occupying the space between known hard facts and the unknowable—journalism that provides insight from expert analysis of available data that rides the border between pure fact and pure opinion (Cox, personal communication, n.d.).

Investigating Algorithms

An emerging field of data journalism is really journalism about technology—the “algorithmic accountability” field, a term coined by Nicholas Diakopoulos at Northwestern University.² Reporters Julia Angwin and Jeff Larson left *ProPublica* to pursue this specialty by founding The Markup, a

2 For more on this field, see Diakopoulos' and Elmer's chapters in this book.

site that Angwin says will hold technology companies accountable for the results that their machine learning and artificial intelligence algorithms create in our society, from decisions on jail sentences to the prices charged based on a consumer's zip code.

This reporting has already prompted YouTube to review its recommendation engines to reduce its tendency to move viewers into increasingly violent videos. It has held Facebook to account for its potentially discriminatory housing ads, and has identified price discrimination in online stores based on a user's location (Dwoskin, 2019).

Works Cited

- Aldhous, P. (2017, August 8). We trained a computer to search for hidden spy planes. This is what it found. *BuzzFeed News*. <https://www.buzzfeednews.com/article/peteraldhous/hidden-spy-planes>
- BBC News. (2018, September 23). *Anatomy of a killing*. BBC Africa Eye. <https://www.youtube.com/watch?v=4G9S-eoLgX4>
- Data. (n.d.). In *Dictionary.com*. Retrieved May 20, 2020, from <https://www.dictionary.com/browse/data>
- Dwoskin, E. (2019, January 25). YouTube is changing its algorithms to stop recommending conspiracies. *The Washington Post*. <https://www.washingtonpost.com/technology/2019/01/25/youtube-is-changing-its-algorithms-stop-recommending-conspiracies/>
- Journalism. (n.d.). In *Dictionary.com*. Retrieved May 20, 2020, from <https://www.dictionary.com/browse/journalism>
- Klein, S. (2016, March 16). Infographics in the time of cholera. *ProPublica*. <https://www.propublica.org/nerds/infographics-in-the-time-of-cholera>
- Woodman, S. (2019, October 22). Using the power of machines to complete impossible reporting tasks. ICIJ. <https://www.icij.org/blog/2019/10/using-the-power-of-machines-to-complete-impossible-reporting-tasks/>

About the Author

Sarah Cohen is the Knight Chair in Journalism at the Walter Cronkite School of Journalism and previously worked as the editor of a data journalism team at *The New York Times*.