1:02 AM

DATA JOURNALISM IN 2017:

THE CURRENT STATE AND CHALLENGES FACING THE FIELD TODAY

Google News Lab

Authors

Simon Rogers, Google News Lab, Data Editor
Jonathan Schwabish, PolicyViz
Danielle Bowers, Google News Lab, Research and Development

Acknowledgments

WHAT'S INSIDE:

Abstract

For many journalists, data is now a way of life. In fact, according to one of the first comprehensive studies of the field undertaken by the Google News Lab and PolicyViz, **42% of journalists surveyed said they use data regularly to tell stories. And 51% of news organizations have a dedicated data journalist on staff.**

In a world where news consumption habits are evolving, trust in media is changing, and misinformation online is rising, how has the use of data for storytelling changed? What challenges do data journalists face? And what can we do to better support this important work?

Through a series of in-depth qualitative interviews and an online survey, we found that many survey respondents—editors, reporters, digital experts, and designers—want their organizations to be using more data and employing it more effectively to tell stories. But there are barriers that limit the use of data in newsrooms.

- **53% of the sample saw cleaning, processing, and analyzing data as a speciality skill that requires extensive training,** and not something all journalists have been able to pick up easily.
- Survey respondents also discussed the time pressures they face as well as **bottlenecks in the editorial process** as a result of the limited bandwidth from dedicated data journalists. We found that **49% of data stories are created in a day or less.**
- Our research also found that **data visualization tools are not keeping up with the pace of innovation.** As a result, newsrooms are building their own solutions: **a fifth of data journalists use in-house tools and software.**
- For some newsrooms, there is an **unclear return on investment** as the production of data journalism can take significant time and resources.

Despite these challenges, data journalism in 2017 has become more mainstream than at any other time in its history. Newsrooms are producing incredible work that helps explain the world around us. Pieces like NPR's "Fact Check: Trump and Clinton Debate for the First Time" or Berliner Morgenpost's "It Was Not Always the East" enrich stories and give meaning to the numbers. The Globe and Mail's "Unfounded" showcases the power of investigative reporting.

Progress comes with unique challenges—and opportunities—for data journalists and those who support them. We hope this research serves as a foundation for discussion and an impetus for action to better meet the needs of data journalists around the world.

1.

Introduction and Background

Over the past two decades,

the amount of data available for research, communication, and analysis has increased at unprecedented rates. For instance, Google alone now processes more than 3 billion searches per day, and there are now more than 195,000 datasets available on data.gov.

Technological innovation has not only made more sources of data available, it has also made it easier to access, analyze, and share that data. Journalists are reshaping their organizations to take better advantage of these new sources of information and serve their mission of providing news consumers with a richer, more accurate understanding of the world.

The Google News Lab partnered with PolicyViz to conduct a series of in-depth qualitative interviews and an online survey to better understand the state of data journalism today and the challenges moving forward.

WE USED THREE PRIMARY QUESTIONS TO GUIDE THE STUDY:

1 | How do journalists define "data journalism"?

2 | What types of stories are produced using data?

3 | What main challenges does data journalism face?

We conducted 56 in-depth interviews with newsroom decision makers, data visualization experts, data journalists, and video journalists in the United States, the United Kingdom, France, and Germany. This was combined with a large-scale survey of over 900 newsrooms in those four countries.

We hope the study helps shed light on what support is needed from the broader ecosystem to ensure data journalism's long-term health.

BACKGROUND: *THE DEBATE OVER THE TERM "DATA JOURNALISM"*

Wellcome Library, London

Data has been used in journalism for hundreds of years, and actually predates newspapers. In the early 1600s, there was a weekly accounting of deaths and births in London that sold for a few pennies.

The more recent increase in the amount of data readily available to journalists alongside more sophisticated data, analytic, and visualization tools makes defining "data journalism" a challenge.

Historically, there have been varying interpretations of the field. Some definitions focus on data journalism's purpose: the combination of the traditional "nose for news" with the sheer scale and range of digital information now available. Others emphasize the processes that help to produce it: "gathering, cleaning,

organizing, analyzing, visualizing, and publishing data to support the creation of acts of journalism"
(Howard 2014). Further complicating the definition is the evolution of other subdisciplines such as
computer-assisted reporting and computational journalism.

**Our findings provide a classification that helps to establish a common understanding of data journalism
based on the types of news produced from it.**

- *STORIES THAT ARE ENRICHED BY DATA*
  (e.g., *"Fact Check: Trump and Clinton Debate for the First Time," NPR*). These are the traditional news
  stories in which relevant data is used to verify the underlying reporting.

- *STORIES THAT USE DATA TO INVESTIGATE*
  (e.g., *"Unfounded," The Globe and Mail*). In these news stories, the journalist exposes information or
  surfaces a story hidden in the data. These types of stories may take longer to produce and may rely on
  a combination of skill sets, ranging from algorithm creation to sophisticated data visualizations.

- *STORIES THAT EXPLAIN DATA*
  (e.g., *"It Was Not Always the East," Berliner Morgenpost*). With more data readily available to the public,
  journalists have made use of increasingly sophisticated data tools to provide meaning behind the
  numbers to readers.

2.
Data and Methodology

As part of the Google News Lab's mission to collaborate with journalists and entrepreneurs to build the future of media, this study was commissioned to better understand data journalism and the challenges the discipline faces today. The initial goal of the study was to explore how Google products such as My Maps, Street View, Earth, Maps APIs, Trends, Fusion Tables, and YouTube could better meet the needs of journalists. The original samples were thus built on a set of contacts provided by Google. An additional set of journalists and newsrooms was added to generate a more representative sample of the field, but the sample is likely over-weighted towards users of Google products. After digesting the findings, the News Lab felt the insights could also be leveraged by the broader industry and used them to develop this paper.

The conclusions of this study are drawn from a two-phase survey. In the first phase, in-depth interviews were conducted with 56 journalists, and the second phase consisted of a large, quantitative polling survey with more than 900 journalists and editors.

THE IN-DEPTH INTERVIEW SAMPLE

Between January 3 and February 28, 2017, 56 journalists in the United Kingdom, the United States, France, and Germany were interviewed. Initially, 33 journalists were provided by Google; another 23 were selected from a sample of other organizations in the four countries. Overall, the sample covered four main job types in each of the four countries: decision makers (e.g., editorial departments), data visualization experts (e.g., digital graphics teams), data journalists, and video journalists. All of the respondents to the qualitative interviews use Google products. Selections for the in-depth interview sample (and, to a lesser extent, the online poll) were based on the classification system described above.

Interviews were conducted by phone and ranged from about 30 minutes to 90 minutes, with the average interview lasting about 60 minutes. Interviews were language specific—English in the US and UK, German in Germany, and French in France. European survey participants were offered a £100/€120 donation to a charity of their choice, and US groups were offered a précis of the report (differences in governing laws about compensation from surveys dictated the different treatment of the organizations).

THE POLLING SAMPLE

To supplement the qualitative findings, we also conducted an online poll with more than 900 people working in the media industry. The poll was conducted between March 7 and March 23, 2017. About 90 poll respondents were sourced from contacts provided by Google, and the remaining were selected from a media panel operated by Omnisis, an online survey firm headquartered in Manchester, UK. Screening questions were used to ensure the right professional background and experience of the appropriate Google products; again, the sample likely oversampled Google product users.

The entire sample was stratified into smaller groups to answer specific questions about specific Google products. Individuals in the stratified groups were required to have experience in the relevant Google product subject areas. The specific Google products include Google Trends (378 respondents), Google Maps (426 respondents), and YouTube (409 respondents).

Over 40% of the overall online sample in the US came from organizations that are a hybrid of traditional and digital news organizations. By comparison, that same group accounted for 23% in the UK, 26% in France, and 32% in Germany. Traditional print organizations accounted for 15% of the US sample, but only 8% in the UK, 11% in France, and 10% in Germany (see Table 1).

TABLE 1. DISTRIBUTION OF ONLINE POLLING SAMPLE BY TYPE OF ORGANIZATION & COUNTRY (PERCENT)					
TYPE	**TOTAL**	**UK**	**US**	**FR**	**DE**
Online or digital news organization	34	41	20	36	41
Traditional print news organization	11	8	15	11	10
Broadcasting organization	15	20	15	16	9
A hybrid of traditional and digital news organization	32	23	43	26	32
Journalism or advertising technology company	6	8	1	11	4
Other	3	0	6	0	3

Basic demographic information for each respondent—job type, tenure in journalism, and age—revealed that the age distribution of the French sample skewed younger than those of the other three countries. About 28% of respondents in the French sample were between 18 and 24 years old, compared with 17% in the UK, 12% in the US, and 24% in Germany. The French sample was therefore weighted to reflect a slightly older sample; all estimates provided below use this weighted sample.

3.
Findings

> " *I kind of, to a large extent, reject the idea of data journalism as a thing. I don't think it's a separate thing, I think it's just like if you seriously want to be a journalist, you really should have some ability to work with data because if you can't, I think you're stumbling around in the dark a lot of the time.*"
>
> — Online newsroom based in the US

PART I. STATE OF DATA JOURNALISM IN 2017

DATA JOURNALISM IS "JUST JOURNALISM": *THE DEFINITION*

One of the concepts explored in the in-depth interviews was respondents' sense of the term "data journalist." During the in-depth interviews, some respondents said data journalists know what data can be used for, what data formats are available, and how data can be structured. Others said data journalists work with data and have data analytic, code, and visualization skills as well as traditional journalism skills. However, a few rejected the term altogether.

We also heard that being a data journalist is "more of a 'state of mind'" and that a data journalist is "essentially the same as any other reporter except that the primary source is data as opposed to interviewing a person—you interview a dataset."

Despite the disagreement over what constitutes a data journalist, the majority of our respondents felt that data was critical to the modern newsroom. More than four out of every five survey respondents (in the Google Trends stratified sample of about 400 respondents) agree that all journalists need to be comfortable reporting on data. For these respondents, data journalism is a response to changing societal norms, changing trust in the media, and challenges with "fake news." Specifically, our survey respondents felt that data:

1. *Helps reduce complexity and give readers a chance to make sense of the world around them.* "The world's a complex place but by visualizing and seeing patterns we can make it seem less complicated." – Data visualizer, USA

2. *Keeps society rooted in facts.* "If you don't have the ability to ask questions of raw data, you're just in a world of, 'He said, she said,' and that's not where I want to be as a journalist." – International, Online, USA

3. *Improves reputations of newsrooms with advertisers, increasing revenue potential with innovative data journalism and visualization.* My job is to make content shine on new platforms / new technologies." – National, Broadcast, USA

AS A RESULT OF THESE DRIVERS, WE FOUND THAT DATA IS USED ALMOST AS OFTEN AS QUOTES

Q *HOW OFTEN, IF AT ALL, DO YOU PERSONALLY USE EACH OF THE FOLLOWING IN YOUR STORIES?* (SAMPLE MEDIA PANEL)

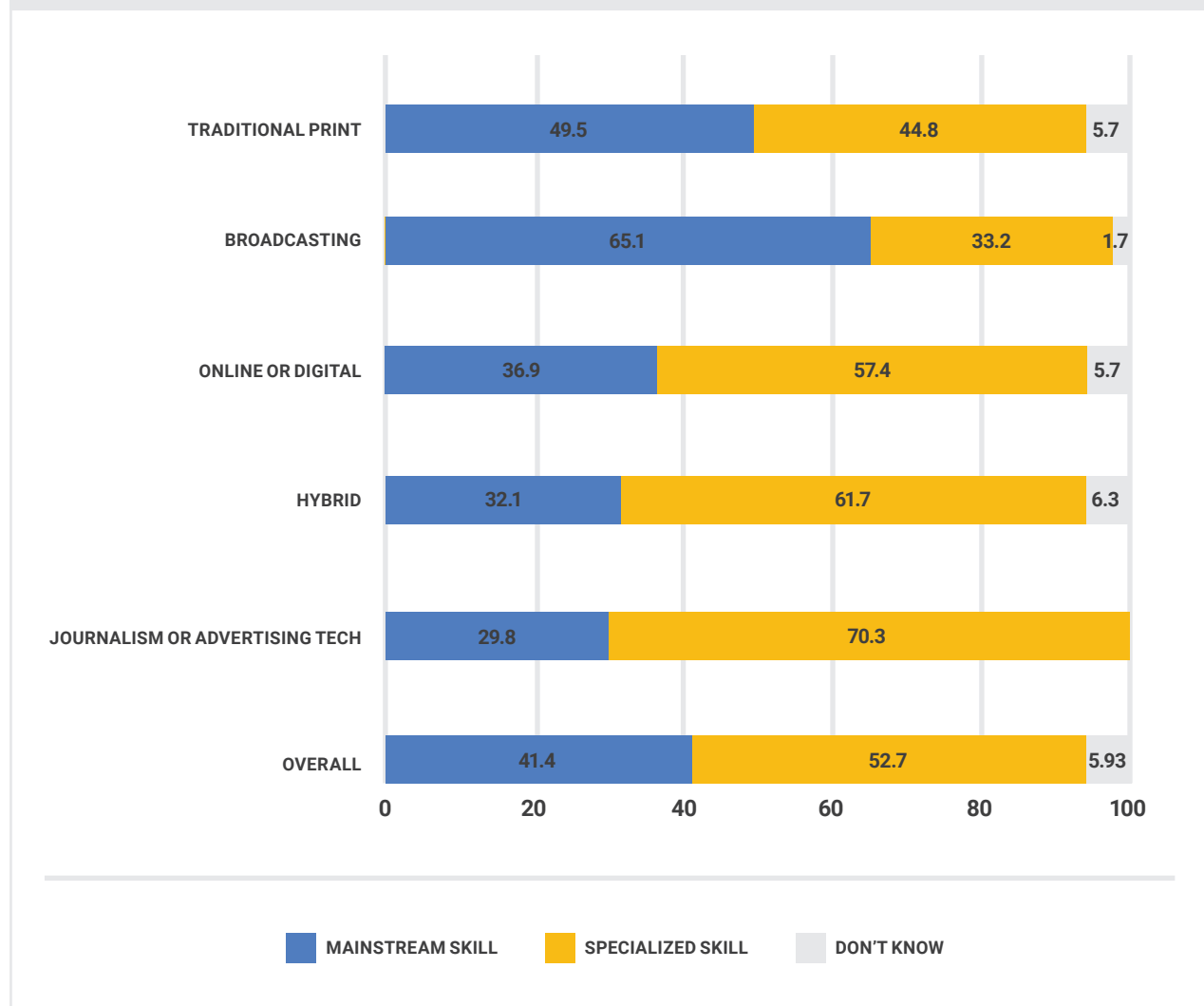**42%** USE DATA 2+ TIMES PER WEEK

14% USE DATA DAILY

50% USE QUOTES 2+ TIMES A WEEK

DATA JOURNALISM IS "JUST JOURNALISM": THE SKILLS

Many journalists did draw a distinction between reporting on data and analyzing data. While reporting on data may be an expectation of all journalists, the ability to find, clean, and analyze data is much more likely to be seen as a specialized skill. About 53% of sample respondents see analyzing datasets as a specialized skill rather than a mainstream skill most journalists should have. Perhaps unsurprisingly, that percentage varies across different types of newsrooms: 33% of people in broadcasting believe it's a specialized skill while nearly 60% in online or digital newsrooms agree that it is a specialized skill.

**FIGURE 1. ABOUT 53% OF ALL SAMPLE RESPONDENTS SEE ANALYZING DATA SETS AS A SPECIALIZED SKILL**

	Mainstream Skill	Specialized Skill	Don't Know
TRADITIONAL PRINT	49.5	44.8	5.7
BROADCASTING	65.1	33.2	1.7
ONLINE OR DIGITAL	36.9	57.4	5.7
HYBRID	32.1	61.7	6.3
JOURNALISM OR ADVERTISING TECH	29.8	70.3	
OVERALL	41.4	52.7	5.93

Data visualization skills are viewed as even more specialized. About 83% of respondents agree that visualizing data to tell a story is a specialized skill. Here, there was less variation across news organization types, though there was more variation across countries: more than 90% of people in newsrooms in the US agree this is a specialized skill, but only 70% of people in German newsrooms agree.

Slightly more than half of all respondents said their news organizations have a dedicated data journalist, but that share rose to 60% in digital-only organizations (Figure 2).

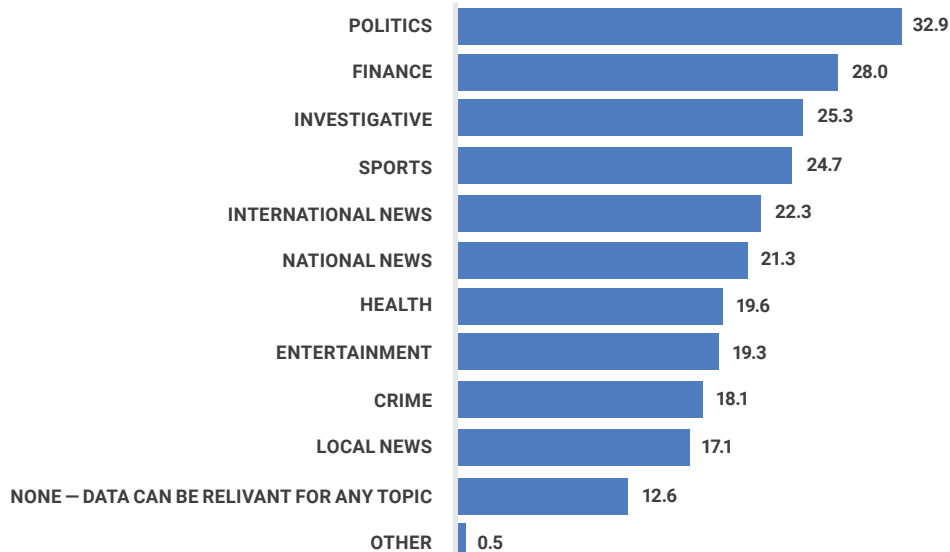FIGURE 2. 60% OF DIGITAL ONLY NEWSROOMS REPORTED HAVING "DEDICATED DATA JOURNALISTS"
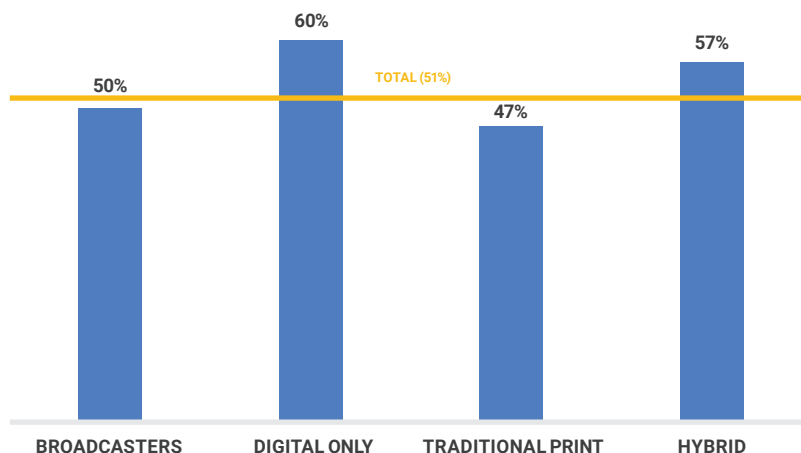
There was slight variation among organizations in different countries. In the US, about 46% of survey respondents reported having dedicated data journalists in the newsroom, compared with 52% in the UK, 56% in France, and 52% in Germany. That being said, US respondents were also slightly more likely not to respond to the question or to respond that they didn't know the answer.

DATA JOURNALISM IS "JUST JOURNALISM": THE TOPICS

About a third of respondents stated that stories on politics are the most relevant to data visualization (Figure 3). Finance, investigative pieces, and sports followed. Nearly 13% of people chose the option for "None—data can be relevant for any topic."

**FIGURE 3. POLITICS AND FINANCE LEAD THE WAY IN
THE KIDS OF STORIES CONSIDERED MOST RELEVANT TO DATA VISUALIZATION**

POLITICS	32.9
FINANCE	28.0
INVESTIGATIVE	25.3
SPORTS	24.7
INTERNATIONAL NEWS	22.3
NATIONAL NEWS	21.3
HEALTH	19.6
ENTERTAINMENT	19.3
CRIME	18.1
LOCAL NEWS	17.1
NONE — DATA CAN BE RELIVANT FOR ANY TOPIC	12.6
OTHER	0.5

DATA JOURNALISTS CATEGORIZE THEIR WORK INTO THREE MAIN BUCKETS

We asked our survey respondents to define their work. From those conversations, three main themes emerged:

1. **First, stories that are enriched by data** (e.g., "Fact Check: Trump and Clinton Debate for the First Time," NPR). These are the traditional news stories in which relevant data is used to verify the underlying reporting.

2. **Second, investigative data journalism.** In these types of news stories, the journalist exposes information that is hidden in the data (e.g., "Unfounded," The Globe and Mail). These types of stories may take longer to produce and may rely on a combination of skill sets from data to algorithm creation to sophisticated data visualizations.

3. **Third, data that needs explaining** (e.g., "It Was Not Always the East," Berliner Morgenpost). With the increase in publicly available and better, more powerful data tools, journalists take responsibility for reducing the complexity of data and providing meaning to readers. This bucket also includes projects that allow readers to explore large datasets themselves (e.g., "Dollar for Docs," ProPublica).

STORIES THAT ARE **ENRICHED BY DATA**

In a "post-truth era," journalists use data to enrich, verify, or illustrate existing story ideas, providing evidence for a story or point of view. Examples might include polling data or economic trends data that is used to help buttress the news story or argument. This kind of data journalism is not fundamental to a story—the story can exist without it—but the data rather enriches and fortifies it.evidence for a story or point of view. Examples might include polling data or economic trends data that are used to help buttress the news story or argument. This kind of data journalism is not fundamental to a story—the story can exist without it—but the data rather enriches and fortifies it.

> **Data is an expected part of any package with any main news story—as part of the package there could be a comment piece, an analysis piece and some data. For example, Trump getting inaugurated—showing his policies in seven charts."**
>
> – National, Hybrid, UK

INVESTIGATIVE DATA JOURNALISM

More in-depth investigations of large data sources are another type of news story produced using data. In this case, the data journalist might explore a dataset to find stories. Because such stories are often buried in larger datasets, these types of stories require a more specialized skill set, and may also require a combination of skills across the newsroom, including data collection, analysis, and visualization, plus the domain-specific expertise to fully understand the data and its potential impact.

Investigative journalism stories may also provide the data to the reader in its entirety to enable the reader to further explore it on his or her own. One example, ProPublica's Treatment Tracker (2014), told the story of individual doctor Medicare billing records from a large dataset provided by the Centers for Medicare and Medicaid Services in 2012. ProPublica's related news article, "Top Billing: Meet the Docs Who Charge Medicare Top Dollar for Office Visits," was accompanied by an interactive data tool that enabled users to explore the dataset on their own.

> **Our role as journalists is...trying to expose information that people want to remain hidden. Trying to make sense of a complex world for people."**
>
> – International, Online, USA

STORIES THAT *EXPLAIN DATA*

Finally, with increased availability of data and better data tools, data that needs explaining is a big driver of data journalism. In such stories, the data journalist becomes the gatekeeper to reducing complexity and providing meaning to the data. "It [data journalism] represents a new role for journalists as a bridge and guide between those in power who have the data—and are rubbish at explaining it—and the public who desperately want to understand the data and access it but need help. We can be that bridge." (Simon Rogers, Google News Lab) The data increasingly being made available is often relevant and meaningful to people only if it is explained and contextualized by people with expertise on a specific topic.

It should also be noted that using data to elucidate the facts does not always help the reader connect with the content on a visceral level. Showing a statistic or number may not be the key to helping a reader fully understand and appreciate the story (Rogers, 2013; Schwabish, 2016). As one of our interviewees noted, "Journalistic work still remains the same: the journalists must still go out to interact with people, to detect the story behind the story. Data alone cannot be used to tell stories about individuals; the emotional story can only be found when journalists go out into the community."

> ❝ *Every amateur can retell a story. Journalism must offer more, reduce the complexity of data so that ordinary users can understand it.*
>
> – Freelancer, Germany

> ❝ *Data is becoming more and more important—governments are releasing larger and more complex datasets. It's reflected in a need to analyze and drill down and tell people what it means—to open up that information you need this skill set.*"
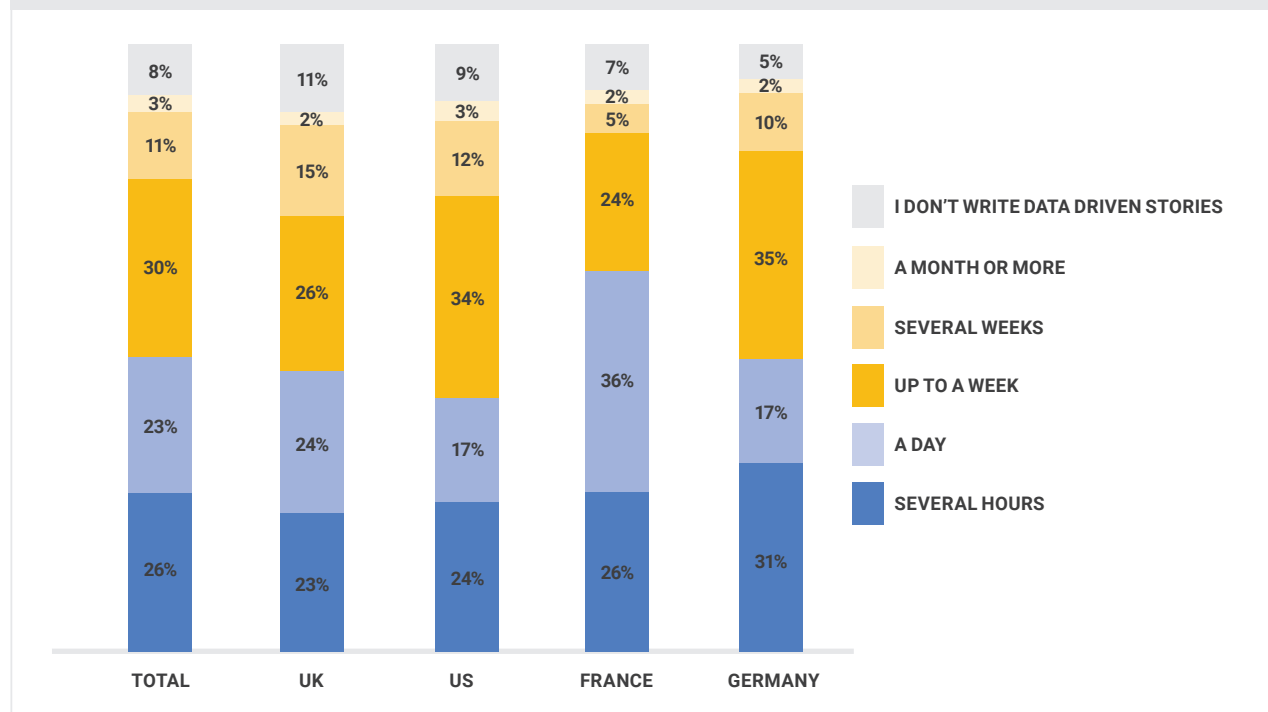>
> – Hybrid, UK

PART II. BARRIERS AND CHALLENGES

TIME PRESSURE

The types of data-driven stories discussed above are sometimes, but not always, affected by time pressures in the newsroom. Overall, about half (49%) of all respondents reported taking a day or less to create a data-driven story; about 44% reported taking up to a week or more; another 8% reported that they don't write data-driven stories.

These shares vary only slightly by country. Most respondents reported that about a quarter of data-driven news stories are created in several hours or a day. Between one-quarter and one-third reported spending about a week on certain stories, while typically about 10% of respondents reported spending several weeks on a project. Newsrooms in France appear to be slightly different here and work on a faster timetable: nearly two-thirds (62%) of their data-driven stories are created in a day or less compared with 41% in the US, 47% in the UK, and 48% in Germany (Figure 4).

FIGURE 4. NEARLY TWO-THIRDS OF DATA-DRIVEN STORIES IN FRANCE ARE CREATED IN A DAY OR LESS

The time pressures vary depending on the type of data story being told as well. Investigative data stories, for example, might take much longer than a story that is enriched by data. Timing is also certainly impacted by the primary platform of the news organization (i.e., print versus online versus television) and other news of the day. Certainly, some news stories have sufficient lead time and organizational backing that resources and time can be invested. For example, the reporting for The New York Times' "Snow Fall: The Avalanche at Tunnel Creek" (December 2012) was put together over 6 months with a team of nearly 20 people (e.g., journalists, designers, and programmers) devoting much of the month to the project. Derek Willis explains that "'Snow Fall' also took longer than many because much of the work required was new. Subsequent work that built on the processes developed through 'Snow Fall' benefited from the lessons of it."

Breaking and fast-turnaround news, however, is still at the core of journalism. With digital platforms and the advent of the 24-hour news cycle, breaking news has now taken on an entirely different meaning. Furthermore, readers' ability to check (and demand) news on the go—especially on mobile devices—is changing how journalists respond to breaking news. The need to obtain, analyze, visualize, and publish data quickly and accurately compounds the time pressure to produce data-driven breaking news stories.

DATA INCONSISTENCIES

Journalists are selective about which datasets they use. Data from official authorities is often the most sought after because it comes from an official source, and is presumably viewed as having an additional layer of credibility. More than half of all respondents reported that publicly available or independent data is "very useful" in their jobs. Independent data from credible sources such as federal governments may also be more highly valued because journalists likely feel they need to spend less time vetting the quality of the data. It should be noted, however, that even those data sources often require careful checking and sourcing, the absence of which could lead to erroneous or misleading reporting. Furthermore, data from public authorities is not always available or, if it is available, it may be available only in difficult-to-use formats such as unformatted data or PDFs. Here is a summary of the various elements we found that are used to determine the credibility of a dataset.

- *GEOGRAPHICAL BREAKDOWN*–*the more granularity, the better*
- *TIME PERIOD COVERED*–*consistency in coverage signals credibility*
- *NUMBER OF CITATIONS AND BY WHOM*–*the more articles written using this data, the better*
- *FORMAT – PDF, EXCEL, ETC.*–*formats that can be easily manipulated save time*
- *ROW AND COLUMN DETAILS*–*the more detail, the better*
- *AUTHOR DETAILS*–*name and contact information is helpful for verification purposes*

Conversely, data collected by private companies (like Google) was reported as being viewed with some degree of cynicism. Respondents expressed suspicion that data might be distorted in some unknown way by the profit motive or expressed concern that it might not be verifiable.

> **"** *Number one, it's got to be…respected, independent, impartial data. Verified, all that stuff. If it isn't, then that's going to be a barrier."*
>
> – International, Hybrid, UK

> **"** *The important data comes from official authorities. Typically, government bodies, local authorities issue quarterly reports or annual reports."*
>
> – National, Online, Germany

> **"** *Increasing fields with more and more data available—we will have to be extra careful on how the data is built. For example, a startup proposed us a lot of data on airlines, but we couldn't control the quality of their aggregation so we said no. We wouldn't have been able to detect mistakes."*
>
> – Regional, Hybrid, France

SKILL SHORTAGES

There is a clear need for organizations to use more data and employ data more effectively in their storytelling. 83% of Google Trends respondents would like to see their organizations use more data in storytelling, and 84% would like to see their organizations use data more effectively. But the technical skills to access, clean, and analyze data remain a general barrier to using data. One respondent noted that the barrier is really cultural: "Not all journalists have the command of math to effectively analyze data. How do you look at outliers? Why do we use the medium vs the mean vs the average? The solution is training and providing tools that people can use." (Online, International, USA)

The challenge is that newsrooms are not finding success at properly training and promoting data journalists within their organizations. ProPublica developer Derek Willis explains why:

> "In most places, the structure is like a clothes hanger: a single person at the top with a large number of reporters. Most other editing structures have upper-level editors that might supervise the work of three to five reporters, leaving enough time to more effectively support their work."

Willis points out that this type of structure also limits promotion opportunities, resulting in few obvious paths for career growth for data journalists, despite the value placed on the skills both by editors and journalists alike. Currently, newsrooms tend to promote reporters rather than data journalists or digital experts to manage data journalism efforts in the organization. One explanation is that there is a limited senior stratum of experienced hands that can foster and guarantee the development of the skills necessary for this field. The lack of senior leadership makes data journalism, in general, less stable and more susceptible to the departure of key personnel, and also makes widespread adoption or understanding of data journalism more difficult. ProPublica's Assistant Managing Editor, Scott Klein, believes "the 'it's just journalism' mentality can be counterproductive to fixing this challenge because it supports the conclusion that a traditional editor can 'edit data' but not vice versa."

VISUALIZATION TOOL LIMITATIONS

Using data to communicate a story also entails having the tools to visualize it. Survey results suggest that data visualization mapping tools are not keeping up with the pace of the newsroom, detracting from the effective use of data in an environment where time pressures can be intense and skills not up to par.

Many newsrooms are now relying on in-house tools because they offer customizable branding opportunities, reliability, no licensing fees, and the efficiency of training in-house (Figure 5). Among the sample of people asked about mapping tools (a total of 426 respondents), about 13% of respondents are mostly or entirely using external tools. One-fifth of the sample are using entirely in-house software and graphics, while another two-thirds of the sample are using some mix of in-house and external tools.

FIGURE 5. A MAJORITY OF RESPONDENTS USE AT LEAST SOME TYPE OF IN-HOUSE GRAPHICS TOOL
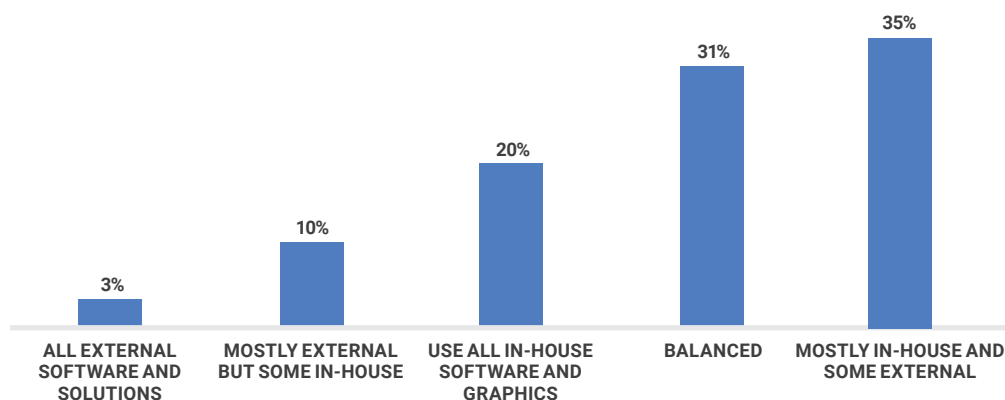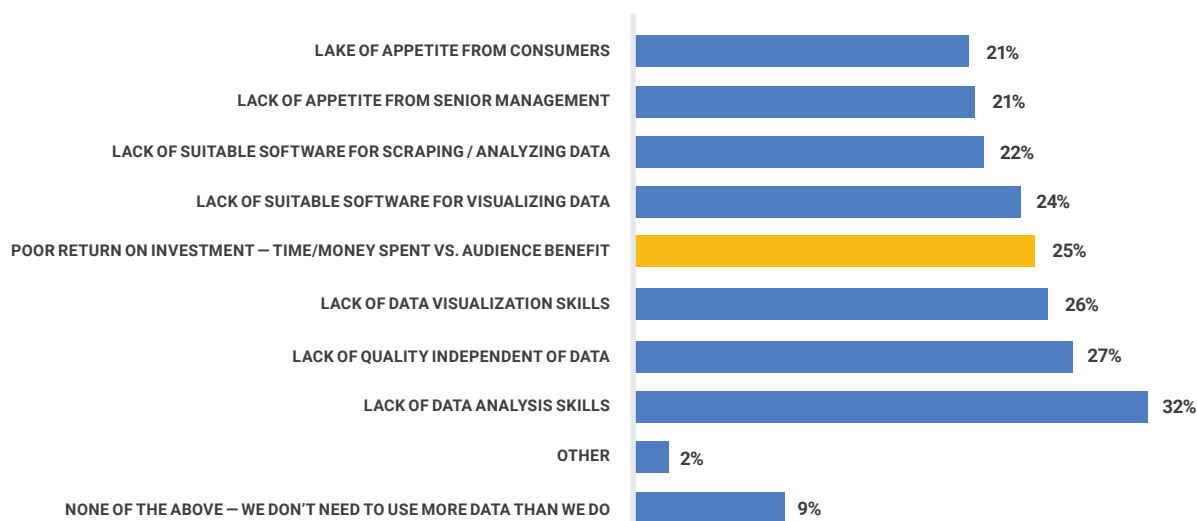

UNCLEAR RETURN ON INVESTMENT (ROI)

The increased availability of data is undeniable, and the responsibility journalists feel to use data to root their coverage in facts is clear. Journalists expressed a great deal of confidence that investigative journalism in particular is valuable to their audiences, but also noted that the time it takes to create and the skill sets required can be very costly. Our findings from the qualitative portion of this study suggest that ROI is clearer to newsrooms with dedicated teams of data journalists than to smaller organizations with fewer dedicated data experts.

Many respondents stated that it takes a lot of time to clean and analyze the data just to understand whether it may be useful. Moreover, the time commitment involved in finding, downloading, cleaning, preparing, analyzing, and visualizing data varies wildly from project to project, creating further uncertainty regarding the ROI. A quarter of respondents reported that poor ROI is a significant barrier to using more data in the newsroom (Figure 6).

FIGURE 6. ONE-QUARTER OF RESPONDENTS NOTE THAT POOR ROI IS A BARRIER TO USING MORE DATA IN THE NEWSROOM

Barrier	Percentage
LAKE OF APPETITE FROM CONSUMERS	21%
LACK OF APPETITE FROM SENIOR MANAGEMENT	21%
LACK OF SUITABLE SOFTWARE FOR SCRAPING / ANALYZING DATA	22%
LACK OF SUITABLE SOFTWARE FOR VISUALIZING DATA	24%
POOR RETURN ON INVESTMENT — TIME/MONEY SPENT VS. AUDIENCE BENEFIT	25%
LACK OF DATA VISUALIZATION SKILLS	26%
LACK OF QUALITY INDEPENDENT OF DATA	27%
LACK OF DATA ANALYSIS SKILLS	32%
OTHER	2%
NONE OF THE ABOVE — WE DON'T NEED TO USE MORE DATA THAN WE DO	9%

"*It's not clear that very complex data pieces, which take a lot of time, generate more audience. Sometimes quick and easy data uses like, say, on taxes work best. There is reluctance from the management to invest in complex pieces.*"

– National, Hybrid, France

"*ROI on data? It's difficult to say, it's a job that demands a lot of time. On the web it's not always the subject that works the best; it's rather niche, so the economical model isn't clear.*"

– Regional, Online, France

"*It's good ROI, in terms of audiences. It's more of a problem in terms of time—it can be two days of work to get that audience, and we don't often have two days.*"

– Regional, Hybrid, France

4.
Conclusions

In many ways, a concrete definition of "data journalism" is perhaps not necessary anymore—we are all awash in data and there is an implicit, if not explicit, expectation that news organizations will use data in their reporting. That data visualization has become a buzzword and something that all news organizations focus on and promote further serves to highlight the importance newsrooms have placed on using data in news stories.

The role of the journalist is also changing. Being able to report on data is now expected, but being able to collect, analyze, and visualize data is is still largely viewed as a specialized skill. Not all newsrooms have the resources to employ a dedicated data journalist or data journalism team, but most organizations are exploring various methods to use more data in their reporting. Below is a snapshot of the various organizational structures we came across:

- *LARGE ORGANIZATIONS WITH ESTABLISHED TEAMS.* These organizations now have established, coordinated data journalism teams. Examples: The New York Times, The Guardian, the Los Angeles Times Data Desk, De Zeit, ProPublica.

- *LARGE ORGANIZATIONS WITH SMALL OR DISPARATE TEAMS.* These organizations have data visualization and journalism teams split across several areas and output types such as print and web. Example: The Washington Post's Wonkblog.

- *DATA-FOCUSED ORGANIZATIONS.* These groups tend to have strong in-house visual and data teams. Examples: 538 and The New York Times' The Upshot.

- ***SMALL DATA TEAMS IN BROADCAST ORGANIZATIONS.*** These organizations have established on-air graphics teams but small data teams. All have websites, but those teams tend to work separately. Examples: CNN, ABC News, Univision News.

- ***SMALL DATA TEAMS IN SMALL BUT AGILE ORGANIZATIONS.*** These groups tend to bring together traditional journalists and data journalists. Journalists often have more autonomy and the ability to post content to the sites. Examples: Vox, Mother Jones, Mic.

- ***LONE RANGERS.*** These are data journalists working in organizations of all sizes who are not part of a large data team. Examples: Mona Chalabi, The Guardian.

The future of data journalism will hinge on a variety of factors, from the sustainability of newsrooms to innovation in technology and the evolution of skill sets to extract value and report new insights to audiences. One thing, though, is certain: the days when data journalism was a "new" field, of use only to a select few, are over. Progress comes with unique challenges—and opportunities—for data journalists and those who support them. We hope this research serves as a foundation for discussion and an impetus for action to better meet the needs of data journalists around the world.

References

Anderson, C. W., Bell, E., & Shirky, C. (2012). *Post-industrial journalism: Adapting to the present.* New York: Tow Center for Digital Journalism.

Appelgren, E., & Nygren, G. (2014). Data journalism in Sweden: Introducing new methods and genres of journalism into "old" organizations. *Digital Journalism, 2*(3), 394–405.

Bradshaw, P. (2012). What is data journalism? In Gray, J., Bounegru, L., & Chambers, L. (Eds.), *The data journalism handbook: How journalists can use data to improve the news* (xx–xx). Sebastopol: O'Reilly.

Cairo, A. (2017). *Nerd journalism: How data and digital technology transformed news graphics* (Unpublished doctoral dissertation). Universitat Oberta de Catalunya, Barcelona, Spain.

Coddington, M. (2015). Clarifying journalism's quantitative turn. *Digital Journalism, 3*(3), 331–348.

De Maeyer, J., et al. (2015). Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism, 3*(3), 432–446.

Diakopoulos, N. (2012). *Cultivating the landscape of innovation in computational journalism.* CUNY Graduate School of Journalism—Tow-Knight Center for Entrepreneurial Journalism.

Dick, M. (2014). Interactive infographics and news values. *Digital Journalism, 2*(4), 490–506.

Fink, K., & Anderson, C. W. (2015). Data journalism in the United States: Beyond the "usual suspects." *Journalism Studies, 16*(4), 467–481.

Gray, J., Bounegru, L., & Chambers, L. (Eds.) (2012). *The data journalism handbook: How journalists can use data to improve the news.* Sebastopol: O'Reilly.

Hamilton, J. T., & Turner, F. (2009). Accountability through algorithm: Developing the field of computational journalism. *Developing the Field of Computational Journalism.* Retrieved from http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf

Howard, A. (2014). *The art and science of data-driven journalism.* New York: Tow Center for Digital Journalism. Retrieved from http://towcenter.org/research/the-art-and-science-of-data-driven-journalism/.

Karlsen, J., & Stavelin, E. (2014). Computational journalism in Norwegian newsrooms. *Journalism Practice, 8*(1), 34–48.

Meyer, P. (2002). *Precision journalism: A reporter's introduction to social science methods.* Lanham: Rowman & Littlefield Publishers Inc.

Nawi, D. E. (2001, April 8). SOAPBOX; "I know it when I see it." *New York Times.* Retrieved from http://www.nytimes.com/2001/04/08/nyregion/soapbox-i-know-it-when-i-see-it.html?mcubz=0.

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society, 15*(6), 853–871.

Petre, C. (2013, October 30). A quantitative turn in journalism? Web blog post, Tow Center for Digital Journalism. Retrieved from http://towcenter.org/a-quantitative-turn-in-journalism/.

Rogers, S. (2013). *Facts are sacred: The power of data.* London: Faber & Faber.

Royal, C. (2010). The journalist as programmer: A case study of *The New York Times* Interactive News Technology Department. *International Symposium on Online Journalism,* The University of Texas at Austin.

Schwabish, Jonathan. (2016). *Better Presentations: A Guide for Scholars, Researchers, and Wonks.* New York: Columbia University Press.

Smit, G., de Haan, Y., & Buijs, L. (2014). Visualizing news: Make it work. *Digital Journalism, 2*(3), 344–354.

Stray, J. (2011). A computational journalism reading list. Retrieved from http://jonathanstray.com/a-computational-journalism-reading-list.

[1] These percentages are essentially unchanged across countries, across types of organizations, and whether the organization is national or local.