This is a repository copy of *Data Journalism in the United States*.

Article:

Data Journalism in the United States: Beyond the "Usual Suspects"

Katherine Fink and C.W. Anderson

ABSTRACT: Understanding the phenomenon of data journalism requires an examination of this emerging practice not just within organizations themselves, but across them, at the -inter-institutional level. Using a semi-structured interview approach, we begin to map the emerging computational journalistic field. We find considerable variety among data journalists in terms of their educational backgrounds, skills, tools, and goals. However, many of them face similar struggles, such as trying to define their roles within their organizations and managing scarce resources. Our cross-organizational approach allows for comparisons with similar studies in Belgium, Sweden and Norway. The common thread in these studies is that the practice of data journalism is stratified. Divisions exist in some countries between resource-rich and resource-poor organizations, and in other countries between the realm of discourse and the realm of practice.

Introduction

Data-journalism, it appears, is everywhere. At least, it is everywhere if one looks primarily at the in-progress academic literature and at the online buzz over new developments in digital news production. Whether and how data-journalism actually exists as a thing in the world, on the other hand, is a different and less understood question. In an attempt to probe the data journalism phenomenon in a more nuanced fashion, this article deliberately goes beyond the organization-specific study and considers the role of data-journalism on a more inter-institutional level (Benson 2006). We begin this article by outlining our methodology, paying particular attention to the manner in which we diversified our object of analysis beyond the so-called "usual newsroom suspects." After elaborating on the results of our interviews in some detail, we compare the results of these findings to a few other nation-wide surveys of data journalists in Sweden, Belgium, and Norway. We hope that this analysis will allow us to go beyond the elaboration of the state of data journalism in a single country and begin to sketch a process through which we might subject this emerging cross-institutional nexus of data journalistic practice to a comparative framework.

Analytical Framework

The last several years have seen an explosion in data journalism-oriented scholarship, making it possible to cluster current and past research on this topic into three major strands. The first strand, which for a long time included the majority of computational journalism studies, is geared primarily toward professional journalists and addressed practical concerns (for example

Cohen S, Li C, Yang J, et al. 2011; Daniel A, Flew T and Spurgeon C 2010; Nguyen 2009; Hamilton JT and Turner F 2009).A second strand maps the infrastructures and arenas that enable the connection between computer scientists and journalists including research focusing on the relationship between data journalism and the rhetoric and institutional structures of the open source software movement (Lewis, S. C., & Usher, N. 2013), the New York Times technology department (Royal, C. 2010; Weber, W., & Rall, H. 2013), the organizational roles played by data journalists in Chicago (Parasie, S., & Dagiral, E. 2012), and Ananny's work on "press-public collaboration as infrastructure (Ananny 2012). A third strand historicizes current developments, examining the links between computational journalism and older forms of data-oriented newswork, such as Computer Assisted Reporting (CAR) (Parasie, S., & Dagiral, E. 2012; Powers, M. 2012). There is even a forthcoming special issue of Digital Journalism, edited by Lewis (2014), which aims to capture the current state of the research.

This study aims to stand alongside these recent scholarly works, but makes two important additions to the extant scholarship as it now stands. First, it aims for breadth rather than depth, forsaking single case-studies or ethnographic research in favor a more wide-scale semi-structured interview approach (though still one limited to organizations within the United States; for methodological specifics, see below). This approach stands in the tradition of recent computational journalism studies of Belgium (De Maeyer et. al. 2014), Norway (Karlsen, J., & Stavelin, E. 2013), and Sweden (Appelgren & Nygren 2014) and we turn to a more comparative analysis of these different studies and our own at the conclusion of this paper. Second, the purpose of this tendency towards breadth is to direct analytical attention to what we think of as

an emerging computational journalistic field, with a focus on the fractures, fissures, and power-dynamics at work within that field, as well as the way that this field is shaped by other institutional clusters in adjacent spaces. This effort is highly provisional, and of course does not resemble anything close to a traditional field analysis; for starters data journalism is very much a field in development and has not yet solidified into anything resembling a classic Bourdieuean structure with formal poles of cultural, economic, or temporal capital (Benson 2006). In addition, we lack the space here to elaborate on all the other inter-field structures—the professional spaces that help socialize data journalists, the foundations and think tanks that fund various data journalism projects, etc—which make up the core of an actually existing journalistic field. What we do do, however, is try to compare the practice of data journalism at multiple large, medium, and small-sized news organizations, and by doing so gain a larger understanding of the inter-organizational tensions, rifts, and stratifications that are part of any widespread multi-institutional social apparatus. This is a modest goal, we admit, but it is an important one.

Methods

One question immediately arises at this stage: if we wish to understand "data journalism" across multiple news organizations, how do we define what data journalism even **is**? Data journalism is ultimately a deeply contested and simultaneously diffuse term, and thus would seem to impose analytical difficulties for those who wish to study it. Two options are available at this stage: the first is to rigorously define what we mean by data journalism and only study those workers who conform to our definition. The second option- the one we ultimately chose- is to begin with a

3

wide cross section of news organization and let the workers within those organization define what **they themselves** mean by doing data journalism. This technique, of course, is similar though not identical to the kind of "grounded theory" approach advocated by Glasser and Strauss (1967) in which initial empirical material is used to define initial analytical frameworks which are then tested again via a return to empirical material. While we go into greater detail about this in the paragraphs below, this open analytical approach relies on news workers either self-defining as data journalists or pointing us to those who fit their already preconceived definitions.

Thus, in order to focus on as broad a range as possible of computational journalism practices, we conducted semi-structured interviews with 23 data journalists who worked at U.S. newspapers and online-only news sites. In order to determine whether data journalism was practiced differently at newspapers of different sizes, we attempted to contact data journalists from publications with large, medium, and small circulations. We chose the ten largest newspapers according to their weekday circulations, according to the 2012 report of the Audit Bureau of Circulation (now known as the Alliance for Audited Media). Since newspaper circulation in the U.S. follows a "long tail" pattern, with the largest three (Wall Street Journal, USA Today, and New York Times) having particularly large circulations compared to other newspapers, we adjusted our selection processes for the medium and small circulation newspapers in order to ensure that newspapers in mid-sized cities would be represented. To choose the medium and small circulation newspapers, we excluded the top ten newspapers and split the rest into two groups, divided by the median circulation of the 365 daily newspapers that remained on the ABC list. The medium circulation sample consisted of the median ten newspapers from the higher-

ranked group; ABC ranked these newspapers 27-36. The small circulation sample consisted of the median ten newspapers from the lower-ranked group; these newspapers ranked 150-159.

We attempted to contact data journalists at twelve online-only news sites. We chose these sites based on 2011 data from Nielsen and comScore on the top news sites based on average unique monthly visitors. We excluded news sites that were operated by newspapers that were already represented in our sample as well as sites that were based outside the United States.

The 23 interviews we conducted resulted from attempts to contact data journalists at the 42 news organizations we selected. Because data journalism is produced differently at different organizations, finding people to interview was not always a straightforward task. Methods of finding appropriate contacts included consulting news organization staff pages to find employees with titles that included words like "data" or "digital." Other methods of finding appropriate contacts included Google searches of the news organization's name and "data" or "digital." In some cases, these searches would lead to web pages that were dedicated to the organization's data projects. In other cases, such searches would lead to stories that mentioned the use of data or included data visualizations like maps or charts. We would contact reporters whose names appeared in bylines or were otherwise connected to these projects. Finally, when those methods did not work, we would email an editor. Editors who responded sometimes answered questions themselves; other times, they referred us to journalists they identified as being primary producers of data stories. Our search led to journalists whose job responsibilities could include data procurement, statistical analysis, graphic design, computer programming, and advising colleagues. The interviews that resulted were with six data journalists from large circulation

newspapers, seven from medium circulation newspapers, six from small circulation newspapers, and four from online news sites.

Interviews took place via telephone and lasted an average of 60 minutes. Journalists were asked five open-ended questions, which led to more specific follow-up questions. They were asked how their professional experience and educational backgrounds related to data journalism. They were also asked about how data journalism fit into the work of their respective organizations. Journalists were asked about their processes for reporting and producing data stories, including how they acquired data and from which sources, and how they decided how to present data—for instance, with narrative description, or with visual elements such as charts, tables, and maps. They were asked how they measured the quality of data journalism, and what constraints they experienced in their work. We arranged their responses into general themes upon the first round of review, then categorized and refined placement upon further analysis.

Results: Enabling Factors

Data journalists had a variety of skills, roles in their organizations, and personal values, and those differences shaped the type of work they produced. Data journalists also suggested that a lack of resources could limit the work they wanted to do. Those limited resources included time, tools, manpower, and the financial means and expertise to fight data requests that were denied. We discovered that there were some fairly profound differences between the way that data journalism was practiced at larger, more resource rich news organizations and the compromises required to practice data journalism at smaller newspapers. One of our most important findings

involved the prevalence of the National Institute for Computer-Assisted Reporting (NICAR), the University of Missouri, and Investigative Reporters & Editors (IRE) in the organizational background of many people working in the data journalism field. We uncovered a diverse yet thematically unified set of organizational roles and skills. Finally, we want to draw attention to the important finding that many medium and small circulation newspapers had trouble keeping data journalists on staff, because those journalists tended to leave for larger organizations.

Skills

Data journalists varied widely in their hands-on skills and educational backgrounds. There is, as of yet, no readily generalizable "data journalism" career path, though many reporters pointed to background exposure to organizations like NICAR and IRE as particularly important. Not surprisingly, all the data journalists we interviewed believed that all reporters should possess at least some facility with data.

Many data journalists began as politics or business reporters and gradually picked up data skills as they became useful to particular stories. One reporter, for example, said he began learning how to use Excel in the 1980s so that he could organize property records that he obtained for a crime series. Other data journalists did not begin their careers as reporters. One had a doctoral degree in political science. Another had a master's in library and information science. Others were graphic designers or computer scientists. Their varied backgrounds were reflected in their varied job titles. The titles of data journalists we interviewed included "Database Editor," "Interactive News Editor," "Infographic Design Editor," and "Computer Assisted Reporting

Specialist." Other journalists we interviewed did not have job titles that suggested data-related responsibilities. One was a city hall reporter. Another was an assistant editor.

Roughly half of data journalists we interviewed mentioned a connection with the University of Missouri, IRE, and/or the joint project of those two institutions, NICAR. Although they were not asked specifically about any of these institutions, 12 out of 23 data journalists mentioned an affiliation with one or more of them. Four data journalists had attended Missouri's journalism school. They and other interviewees had attended NICAR training workshops or conventions, or subscribed to the organization's email list. Several data journalists said that they believed NICAR had changed over time. One data journalist said that when she attended her first NICAR conference in 1999, she was mostly among reporters, especially investigative reporters. She believed that recent attendees were more likely to have backgrounds in the Web, information technology, and graphics. Another data journalist said he saw a split in NICAR between "geeky reporters looking at data who tend to be older," and "a younger group of students who have more of a background in computers and games… they're much more interested in functionality, their technical skills are more advanced, and it's harder for the older group to mix with them." But while some data journalists believed NICAR was becoming more technical, one said he thought it was getting less technical due to an influx of newcomers. He also said he was reluctant to share his own experiences on the email list because he was worried that someone would steal his ideas.

The skill sets of data journalists influenced the type of work they did. One journalist said that her data stories tended to feature interactive databases, because her web producers knew how to make them. Another journalist said that her newspaper had gone for months without any data

visualizations, such as charts or graphs, because it had no graphic designer. Another data journalist said that she wanted to do more data-mapping and interactive graphics, "but am stretched a little thin for time." Another data reporter said that his stories often had interactive elements but not infographics, because those were the responsibility of the graphics department.

The data journalists we interviewed believed that all reporters should have data-related skills. One reporter at a news organization that had no dedicated data journalist said that data should be "part of every journalist's toolkit" because it can help identify uncovered trends. A data journalist at a small newspaper said that understanding data was crucial for reporters because "when you know how it's done, you're better able to question the results." One editor said he expects all of his future reporter hires to, at minimum, be comfortable working with spreadsheets. "A lot of people got into journalism because math wasn't their thing," but he said that was no longer acceptable. One data journalist said she told reporters they increasingly needed to see stories as "a question rather than a noun… define your story in a way that requires you to quantify something."

Although there is widespread diversity in the backgrounds and skill sets of the data journalists we interviewed, a few overarching similarities were apparent. We find a similar mixture of diversity and homogeneity when we turn to an overview of data journalists' organizational roles.

Organizational Roles

There was an extreme degree of heterogeneity when it came to the organizational roles of data journalists at small, medium, large, and online only news organizations. They often tended to be isolated or working alone at small and medium sized newspapers, and part of overarching teams at larger news organizations. At small organizations, in particular, data journalists often found their time and attention divided across a variety of tasks that were more generally "technical" in nature (Powers 2012).

Data journalists could be leaders or low-ranking employees. Data journalists who had editorial roles often lobbied for particular projects and tried to encourage other reporters to have a "data state of mind,"—in other words, to think creatively about ways that they could incorporate data into their stories. On the other hand, data journalists who were lower-level employees sometimes felt isolated from the rest of the newsroom. Data journalists could fit into a variety of physical and virtual locations within newsrooms. A data journalist said that his co-workers tended to see him as the "geek in the corner." Some data journalists were considered to be part of investigative reporting teams. Other data journalists were grouped with their organization's online, IT, or graphics departments.

Data journalists at large newspapers tended to work as part of a team. Members of those teams often had a mix of formal education in statistics, computer science, and graphic design. One large paper had an interactives desk, graphics desk, and computer-assisted reporting team, all of which worked together as well as with reporters. Another large newspaper's data team consisted of computer-assisted reporters, a graphics desk, and programmers. One data journalist said his newspaper had formed a "data desk" that included six reporters. Another data journalist said he

imagined his newspaper as a "factory for reporting," consisting of several assembly lines in the form of newspaper sections or beats. The data team moved among the assembly lines as necessary.

Some journalists we interviewed did their own data reporting, while others played more of a support role. One such journalist said he saw himself as a "teacher and helper" to his co-workers. Data journalists could help reporters crunch numbers, build databases or design graphics. Other times, data journalists worked largely autonomously, finding story ideas, doing their own reporting, and creating their own visuals.

Data journalists often had other duties besides data projects. One data journalist said she spent a lot of time on "simple, tedious tasks" such as updating her newspaper's voter guide. Another data journalist said she wore "many many hats in the newsroom," and her other duties included "making a lot of lists for print, answering phones, and event planning." Another data journalist said she had become her newspaper's public records expert because she bore the responsibility of fighting for data when government agencies refused to release it, or charged too much for it—which she said happened often.

<u>Personal Values</u>

The work of data journalists often reflected their personal values about storytelling and privacy. Those values included consideration of the most effective ways to present data, the differences between data that is difficult to obtain versus that which is easy to obtain, the value of "one-off"

data journalism projects versus continually updated data reporting interactives, and various privacy considerations. The fact that there is little consensus in how data journalists think about privacy, in particular, points to this as an area in which journalism schools and professional training may play an important role. Finally, the number of data journalists who actively embraced the use of reader metrics in driving story choices and editorial decisions turned out to be surprisingly small.

While some journalists liked to plot data on maps, one said he actually believed that maps were used too often. "I used to be a total map nerd. Now I ask myself, 'how can I not make this a map?'" he said. Some journalists said databases sometimes stood on their own, while others would never create a database or data graphic without a story alongside. Some data journalists posted whole databases whenever possible, while others said they only did so when they believed readers that would want to search for specific records—for instance, school test scores. One journalist said the advantage to publishing whole databases was that it could lend credibility to a story. By seeing all the data, he reasoned, readers could understand the way his organization had analyzed it.

Some data journalists said they purposely sought data that was hard to get. One reason was that hard-to-get data tended to be more controversial. Data that took work to uncover was the "holy grail… I'm more interested in what people don't want to give me," one journalist said. Another journalist praised a public records expert at his newspaper who excelled at "wrenching [data] from the grasp of government."

Data journalists were more interested in one-time projects than updating prior stories when new data came out. News organizations may not update a story about hospital infection rates, for example, even though new data are released annually. "In a perfect world, everything would be durable," one data journalist said, but he said his organization did not have the manpower to keep older databases updated. Another data journalist said his organization once tried to update old databases but stopped because they were difficult to monetize. One data journalist, however, felt strongly that news organizations should provide ongoing data on the communities they serve. She compared the concept of continually updated data pages to the annual community directories that many newspapers once distributed to local residents.

The personal beliefs of data journalists about privacy could determine how they presented stories, or whether they presented them at all. Only one data journalist we interviewed believed that all public records that his organization acquired should be made available to readers. Others said that they believed there times data should be withheld in the interest of protecting people's privacy. Examples of data they believed should remain private included medical records, voter registrations, and birth and death certificates.

Sometimes, journalists believed it was acceptable to publish sensitive data if it were aggregated. Presenting aggregated data can inform news audiences about public issues while minimizing the risk of revealing individual identities. One journalist cited a story his organization did on geographical differences in 911 response times. Although he had the exact location of each 911 call, the journalist mapped the calls by neighborhood rather than by individual address, thus not revealing which specific residences were responsible for each call. Another newspaper published

Section 8 data in aggregated form following a months-long legal battle with a local housing agency that claimed releasing the data would violate the privacy of Section 8 recipients. The agency ultimately released only data for apartment buildings that had ten or more Section 8 units, in the interest of protecting individual Section 8 recipients.

Which data are considered to be sensitive may vary by organization and geographical region. Some journalists regularly reported government salaries. But one journalist recalled a backlash when he planned to publish salary data. In the region his news organization serves, "asking people how much money they make is not a nice thing to do," he said. Local officials tried to stop his organization from publishing the data, and his newspaper received several complaints afterward. Not as controversial was the publishing of criminal data. Two journalists said their organizations published mugshots or other arrest-related data.

Journalists often cited a public interest standard as a guide for determining whether to publish sensitive data. One journalist defended publishing arrest records as a way to let readers know about crimes that had been committed in their area, and to keep them informed about the activities of the local jail. Another data journalist said he tried to encourage "watchdogging" rather than "snooping." In the words of another data journalist, an important question to ask was "is this voyeurism or is this journalism?" As they acknowledged, however, news organizations benefit when data appeal to the voyeuristic instincts of readers, since those instincts can lead to more online traffic.

In fact, data journalists knew little about how their audiences interacted with their stories. Most journalists we interviewed said they were generally aware of which stories generated heavy online traffic. But they were skeptical of the metrics they saw. "People click on cats, after all," said one journalist. She said online traffic for data stories could rise or fall dramatically based on factors that were unrelated to news value, such as the placement and size of stories, and how well they were promoted. She said she also preferred tracking which stories were shared, rather than which generated the most pageviews. Another data journalist said pageviews did not necessarily correlate to the value of a dataset. Election data, he said, might only interest ten percent of his readers, but he believed that they found that data to be highly important. A data journalist who said he was "not obsessed" with pageviews acknowledged that he might still use them to evaluate whether a particular project was successful. Another data journalist said that if a data project generated a lot of pageviews, he would be more likely to consider updating the project when new data became available.

Pageview metrics also told journalists little about the impact of data elements within stories. If a database were embedded into a page that also included a text-based story, for example, journalists might not know if users searched the database, or noticed it at all. One data journalist said improving the tracking of online behavior was a top priority for him. Another journalist said she had no way to know how users interacted with data, but she did know that having data tended to increase the time they spent on a page.

Three out of the 23 journalists we interviewed said they felt pressured to make story choices based on what they thought would drive online traffic. One journalist said his newsroom was

highly focused on generating pageviews. All reporters at his organization got a daily email about how much traffic their stories generated, and he felt conflicted about what to do with that information. "I don't want to be TMZ," he said, referring to the gossip news site. On the other hand, he said, he wanted to write stories that many people would want to read. All the same, he said he felt strongly that his organization should be reporting more data stories on local government—a topic that never drew a lot of online traffic. The second data journalist said he always tailored his story choices according to which he thought would generate the most pageviews, "because that's what the bosses want." The third data journalist said she was expected to produce high-ranking databases, and that over time she developed a good sense of which topics drove the most traffic: education and crime.

Results: Constraints on Data Journalism

While the previous discussion examined inter-institution-level factors that tended to facilitate the production of data journalism, the following section discusses the external factors that often act as constraints on the production of data-driven news stories. They include a lack of time, a lack of technological tools, a lack of manpower, and a lack of legal resources. Even more than the section on enabling factors, differences between large and small / medium-sized news organizations were particularly important.

Lack of Time

Most data journalists indicated that a lack of time could influence the stories they chose to do. Journalists were more likely to use data that were easy to procure, were presumed to be credible, and required minimal cleaning and formatting. U.S. Census data were popular—one data journalist called the release of new Census numbers "Data Christmas." Other popular datasets related to education, such as test scores, teacher ratings and school budgets. Several data journalists said they created searchable databases of government salaries. Budgetary, unemployment, and campaign finance data were also commonly mentioned, as well as health care data such as hospital ratings. Journalists at medium and small circulation newspapers were more likely to mention data stories involving crime. Those stories often included maps of crimes that had been reported to police.

These data were popular because they were readily available and easily digestible. Government agencies like the U.S. Census bureau regularly posted data online, and in formats that allowed journalists to work with them easily. Journalists said they usually spent more time cleaning data than analyzing it, so datasets that were easily readable and had few errors were more appealing. The cleanest datasets, they said, tended to come from large, public institutions.

Working with private data could be more time-consuming. Public records laws required government data to be available, although the specifics varied by state. Private data were less accessible, and could come at a greater price. Government agencies tended to limit what they charged to the cost of labor and materials. Private companies could charge whatever they wanted. Another reason data journalists avoided private data was that they saw it as requiring more vetting. Public data, like other types of government sources, had the advantage of

presumed credibility. Journalists feared that third-party data providers had a particular agenda they were trying to push that was not necessarily in the public interest. This concern could be amplified by the fact that private companies did not always disclose all of their raw data or the specifics of their methodologies.

Examples of private datasets used by journalists included those produced by the company DataQuick, which specializes in foreclosure and other real estate data. The company Kantar Media had what one data journalist considered to be the best data on political advertisements. One small newspaper's website featured a widget from the private company GasBuddy, which featured a map that had real-time crowdsourced data on gas prices.

Lack of Tools

Some data journalists felt limited by the tools that were available to them. Larger newspapers were more likely to have developers on staff. Those developers used Python, Ruby on Rails, JavaScript, HTML, or other computing languages to tailor software to individual projects. Data journalists at smaller news organizations were less likely to have programming backgrounds, although some of them were interested in developing more expertise in that area. Data journalists at smaller news organizations were more likely to use third-party software like MySQL, Access and Excel. They also used data visualization programs like Caspio and Tableau. One data journalist at a small circulation newspaper said he once used the mapping program ArcView, but his newspaper let its subscription expire due to a lack of funds. Another data journalist said she was learning a particular type of mapping software because the company that created it happened

to be located nearby. Data journalists at smaller news organizations were also more likely to mention that they used free tools like Google's Fusion Tables, Maps and Docs.

The tools journalists had available could influence whether stories were presented as interactive databases, graphs, maps, or whether they had any visual elements at all. Data journalists at large newspapers were less likely to identify limitations associated with their tools, since they created many tools themselves. Smaller newspapers, "if they want to go beyond their CMS [Content Management System], they can't," one data journalist said. Other journalists affirmed that they felt limited by their CMSes, as well as software that they saw as a less-than-ideal fit for the type of work they wanted to do. "I use Caspio, but I hate it," one data journalist said, explaining that the databases he used tended to be too large for the software to handle.

Lack of Manpower

Our research leaves little doubt that the economic downturn at many American news organizations has had a deleterious impact on the production of data journalism. Indeed, while the last decade has seen an overall increase in the prominence of data in news, we were left to wonder how things might have been different if these changes had been made in less economically disastrous times. One editor at a small circulation newspaper, for instance, said he did more data journalism a decade ago. Since then, most of his operations had been "trimmed to a bare minimum" due to the newspaper's debt. Another journalist who was the sole data reporter at his organization said the number of stories he wrote had increased—but only because of a decrease in the number of investigative, in-depth stories that required more time for data analysis

and presentation. Another data journalist said that when she was hired, she was part of a team that included four full-time and two part-time researchers. She was the only one from the team who still worked there.

Some medium and small circulation newspapers had trouble keeping data journalists on staff, because those journalists tended to leave for larger organizations. Some data journalists at larger newspapers said they also had employee turnover problems. Newspapers that lost their data journalists sometimes left those positions unfilled for months, either due to a lack of qualified applicants or because waiting to hire a replacement saved money. One editor who had had trouble hiring a data journalist expressed hope that reporters who already worked there would take an interest in data reporting. "We have a young guy now who's interested in Google Maps," he said.

Lack of Legal Resources

Most data journalists said they encountered at least occasional problems getting the data they wanted from government sources. "Just because they're dumping stuff doesn't mean they're happy about it," one data journalist said. Larger news organizations had attorneys who could advise data journalists about their rights and file appropriate paperwork when agencies were uncooperative. Smaller organizations often lacked the resources to put up a fight. One data journalist said that if he wanted to challenge an agency that refused to release data, "we know that we have to go into it with a bluff." He said his newspaper does not have the resources for an extended legal battle, so "if we fail, we complain," but then drop it. "I've tried to steer reporters

away from things that are too time intensive," he said. Another data journalist said such battles are fought on a case-by-case basis, after weighing whether the story was important enough to invest the time and money necessary.

Some data journalists saw public records battles as partly their responsibility. One data journalist said she learned the law over time because she saw "great misperceptions" among government employees about their obligations regarding the release of public data. One journalist also said government agencies repeatedly tried to overcharge her for labor because they would base their costs on the amount of time it took to download the datasets she wanted. She argued that was excessive because no human labor was involved while the download was taking place. "If it's over 50 dollars, I question it," she said.

The cooperation and tech savvy of public officials could determine which data stories were done. One data journalist said the former head of her city's housing agency granted generous access to a database of vacant and abandoned properties, which led to a large story on safety concerns near those properties. Another news organization helped its county jail establish an RSS feed of its daily bookings, in exchange for allowing a constant feed of that data to appear on the newspaper's website. One editor said a local police chief agreed to release a specific type of crime data on the condition that a reporter teach him how to work with it.

On the other hand, some data journalists said they often received paper copies of records, despite their requests for electronic formats. One data journalist said a public official who was reluctant to release data deliberately provided it in a difficult-to-read format. The data journalist said the

official's secretary told him that she had been instructed to scan the data, photocopy it, and fax it to herself before mailing it to him. (The journalist said it ended up not being much of a problem—he used optical character recognition software to convert the data into electronic form.) Another data journalist said he believed government agencies sometimes released far more data than was requested in order to obscure the information journalists most wanted.

Although our sample size was limited, our results suggest a much more optimistic future for data journalism at large circulation newspapers and online organizations than at medium and small newspapers. Reporters at large circulation papers and online organizations said the amount of data journalism produced there had increased or remained the same during their tenure, while journalists at smaller newspapers said the opposite. Larger organizations were more likely to do data work that involved a division of labor, with computer-assisted reporters, graphic designers, statisticians and programmers working on teams. Smaller organizations were more likely to have "one-man bands" who acquired data skills as needed or due to their own initiative. When those journalists left for greener pastures, as they often did, data journalism efforts at those organizations could have to be rebuilt, or might stall completely. Larger organizations also had a greater ability to develop their own data tools, which they could improve and customize over time. Those improvements allowed them to take on more ambitious projects over time. Smaller organizations were more susceptible to the limitations of third-party tools. In an industry that remains economically strained, data journalism can be seen as a luxury that only the most elite news organizations can afford to do well.

Discussion and Conclusion

This paper, while primarily serving \ as a stand-alone overview of data journalism practices in the United States, can also be said to belong to an emergent tradition of articles on data and news at the national level. Recently published studies have looked at data journalism in Sweden (Appelgren & Nygren 2014), Norway (Karlsen, J., & Stavelin, E. 2013) and Belgium (De Maeyer et. al 2014). While this may be a somewhat eclectic cross-section of national systems, and while the eclecticism and lack of a shared definition of data journalism does not allow for systematic scientific comparison across countries, it is indeed helpful that there is an growing set of nationally-focused data journalism studies that allow us to better understand developments in the United States. To understand the United States, in other words, it is necessary to compare the U.S. to other places. In our conclusion, want to bring these Belgian, American, Swedish, and Norwegian data journalism practices into dialog with each other around two issues: first, methodologically, how do they conceptualize the population of potential data journalistic actors in their country. And second, what are the barriers, if any, to data journalism being practiced more widely across a national media ecosystem?

Our interviews used a deliberately stratified interview sample, with the primary goal being to let working journalists and editors define how they envisioned the emerging cross-institutional field of data journalism,, and with a secondary goal of talking to editors and reporters at **large**, **medium**, and **small** size news organizations. In large part this focus on newsroom size was due to the fact that the federalist structure of the United States (both in terms of where political decisions are made and where the vast majority of local news comes from) means that small and

medium sized news organizations are usually local, and local news is particularly important for citizens' political knowledge and in democratic decision making more generally (Downie and Schudson 2009). In the study of Norway, on the other hand, all organizations "have their base in Oslo or Bergen (or both)." The lack of local news organizations, the authors note, may be an artifact of the snowball sampling methodology or, importantly, "the possibility that very few local newsrooms practice computational journalism on a regular basis" (Karlsen, J., & Stavelin, E. 2013, 38). The Sweden survey included a sample of regional newspaper organizations (37 percent), though the results are not stratified by size. The authors of the Belgium study, finally, talked to national editors and newsroom managers, as well as journalists working at both national and regional newspapers (De Maeyer et al. 2014, 18).

We have already noted that local news (and the health and journalistic behavior of small or regional news organizations) is particularly important in the United States; it is obviously less important in many other countries with a less federalized political structure or less of a history of local or regional newspapers. In French-speaking Belgium, for example, there is little difference between national and regional news organizations because the country itself is so small! But even in all three of these studies that focus largely (if not entirely) on major news players, the health of data journalism was mixed. Only a few newsrooms in Norway practice data journalism. In Sweden, data journalism is fairly uncommon. In Belgium, finally, the excited rhetoric about data journalism has not been match by a performative reality: there is much talking, but less doing.

It would be easy to frame the United States as the exception to these trends, and in many ways, it

is. The largest news organizations in the U.S. (along with the Guardian in the United Kingdom)

are doing truly pioneering, even revolutionary, computational journalistic work. But the contrast

becomes less acute when we lower our gaze to the second and third tiers of newswork. And when

we do that, we shift our focus to organizational routines and resources, which might help add

substance to a comparative tendency to focus on grand national differences in the language of

journalistic history or culture.

So if data journalism is rare in Belgium, Sweden, and Norway, and rare in some places in the

United States, why might that be? Importantly, all four studies pointed to lack of time and lack of

resources as major culprits. Many of our respondents at smaller papers even went so far as to say

that they used to have more time for data journalistic work, in large part because of shrinking

resources and therefore less time. In Sweden, one respondent noted simply that "time is scarce."

In Norway, "the limiting factors are not technological infrastructure, but time and goodwill"

(Karlsen, J., & Stavelin, E. 2013, 39). In Belgium, finally:

> Time, or the lack thereof, emerges as one of the main barriers to the practice of data
> journalism. Some respondents frame time as something that the organization refuses to
> give to journalists (HR2) because it has other priorities, or admit that their practice of data
> journalism is confined to their free time (J2, J1). One journalist who has successfully
> engaged in the production of data journalism projects emphasizes as a key enabler how he
> was able to convince his hierarchy to give him some time (J4). (De Maeyer et. al 2014,
> 12).

One final similarity in our findings relates and helps nuance this strictly resource oriented focus

on time and organizational capacity. All four studies noted that the conception of data journalism

was extremely vague, both rhetorically and organizationally. While often framed as a pragmatic benefit by news workers (vagueness allows for spontaneity and organizational flexibility) this lack or clarity also comes with a cost. In managerial terms, uncertainty about organizational roles can lead to a world where data journalists are often also social media managers, fixers of broken technical devices, and all around "helpers out" in newsrooms. In a world where resources are plentiful, this might not be such a problematic scenario. But in a situation of shrinking resources, this lack of clarity can result in journalists who used to be data journalists suddenly becoming a great many other things at once.

And so, our study fits somewhat uneasily into a general set of findings that see the production of data journalism as highly stratified and existing in some places and not in others. Stratified between resource rich and resource poor organizations in the United States and possibly Norway, and between the realm of discourse and the realm of practice in Belgium. Even if we avoid overt framing conceptions such as "organizational field," it should be clear that any relationally conceived inter-institutional organizational structure is going to possess these differences is resources and cultural capital. This study begins to put flesh on these relational bones.

Obviously one of the opportunities for further research in this area would be to continue to broaden our analytical lens to include other countries, particularly those outside of North America and Europe. Hallin and Mancini's (2004) work on comparative media systems, while controversial, offers one possibility of a larger thematic framework to compare different understandings and practices of data journalism. Further research opportunities in this area include a more formal comparison of data journalism practices at organizations of varying sizes--

perhaps through a survey instrument or through interviews with a larger number of organizations. Our sample size was too small to draw definitive conclusions about differences based on size, but our interviews suggested that the state of data journalism at the lower circulation newspapers was precarious. Data projects there came as the result of a lucky hire, or at the initiative of journalists who took it upon themselves to learn data skills in their free time. Meanwhile, data journalism at the larger newspapers and online-only organizations appeared to be thriving. If the gap between data journalism resources is as wide as our preliminary research suggests, this would add to an already considerable list of concerns about the future of newspapers in all but the largest metropolitan areas in the United States.

Appendix: Organizations Where Data Journalists Worked

Wall Street Journal

New York Times

USA Today/Gannett Digital

Los Angeles Times

San Jose Mercury News

Chicago Sun-Times

Portland Oregonian

Seattle Times

Detroit Free Press

San Diego Union-Tribune

St. Paul Pioneer Press

Schenectady Gazette

Lincoln Journal Star

Santa Rosa Press Democrat

Huffington Post

NPR

Slate

Boston.com

3 additional small circulation newspapers

Works Cited

Anderson, C.W. 2012. "Towards a Sociology of Computational and Algorithmic Journalism." New Media and Society 15(7): 1005-1021.

Ananny 2013. "Press-Public Collaboration as Infrastructure: Tracing News Organizations and Programming Publics in Application Programming Interfaces. American Behavioral Scientist. 57(5): 623-642

Appelgren, Ester, and Gunnar Nygren. 2014. "Data Journalism in Sweden: Introducing New Methods and Genres of Journalism Into Old Organizations." Digital Journalism (0):0 1-16.

Benson, Rodney. 2006. "News Media As a 'Journalistic Field': What Bourdieu Adds to New Institutionalism, and Vice Versa." Political Communication 23(2): 187-202.

Cohen, Sarah, Chengkai Li, Jun Yang, and Cong Yu. 2011. "Computational Journalism: A Call to Arms to Database Researchers." In Proceedings of the 5th Biennial Conference on Innovative Data Systems Research. Sailorman, CA: ACM.

Daniel A, T Flew, and C Spurgeon (2010) The promise of computational journalism. In Proceedings of the 2010 Australian and New Zealand Communication Association. Canberra: ANZCA

Downie, Leonard and Michael Schudson. 2009. "The Reconstruction of American Journalism." Columbia Journalism Review 19.

Hallin, Daniel C., and Paolo Mancini. 2004. Comparing media systems: Three models of media and politics. Cambridge University Press.

Hamilton, James T., and Fred Turner. 2009. "Accountability Through Algorithm: Developing the Field of Computational Journalism." Last accessed November 12, 2013, available at: http://dewitt.sanford.duke.edu/images/uploads/About_3_Research_B_cj_1_finalreport.pdf

Karlsen, Joakim, and Eirik Stavelin. 2014. "Computational Journalism in Norwegian Newsrooms." Journalism Practice 8(1): 34-48.

Lewis, Seth C. 2012. "From Journalism to Information: The Transformation of the Knight Foundation and News Innovation." Mass Communication and Society 15(3): 309-334.

Lewis, Seth C., and Nikki Usher. 2013. "Open source and journalism: toward new frameworks for imagining news innovation." Media, Culture, and Society 35(5): 602-619.

Ngyuen D (2010) Scraping for journalism: A guide for collecting data. Available at:

http://www.propublica.org/nerds/item/doc-dollars-guides-collecting-the-data.

Parasie, Sylvain, and Eric Dagiral. 2012."Data-Driven Journalism and the Public Good: 'Computer-Assisted-Reporters' and 'Programmer-Journalists' In Chicago." New Media and Society 15(6): 853-871.

Powers, Matthew. 2012. "'In forms that are familiar and yet-to-be invented': American journalism and the discourse of technologically specific work." Journal of Communication Inquiry 36(1): 24-43.

Royal, Cindy. 2010. "The Journalist As Programmer: A Case Study Of The New York Times Interactive News Technology Department." Presented at The International Symposium For Online Journalism, Austin, TX.

Schudson, Michael. 2005. "Four Approaches to The Sociology of News." In Curran J and Gurevitch M (eds.) Mass Media and Society, (4th ed) London: Hodder Arnold.

Weber, Wibke & Hannes Rall. 2012. "Data Visualization in Online Journalism and its Implications For the Production Process." Paper for the 16 International Conference on Information Visualization. Retrieved via http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber =6295837.