

Local Journalism's Possible Future:

Metric Media and its Approach to Community Information Needs

ASA ROYAL & PHILIP M. NAPOLI

JULY, 2021



Executive summary

In recent years, as thousands of local news outlets have folded amidst financial troubles, a number of online news networks have emerged to take their place. But reporting suggests that despite these networks' nimble structure, unfettered as it is by a need for printing presses and local offices, their news product and their adherence to journalistic standards are often lacking.

Metric Media, one such news network, claims to be the largest of the bunch. In this paper, we use data acquired by scraping the front pages of Metric Media's 999 local outlets daily for 2+ months to assess patterns and qualities of the company's news production, in an effort to get a sense of how this network of sites is meeting communities' information needs. We find that:

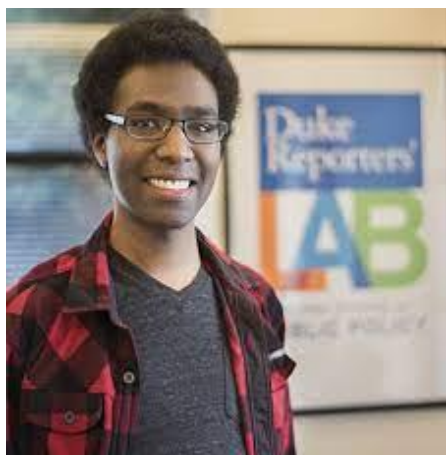
- Although Metric Media operates nearly 1,000 local news outlets, most lack original reporting. **What content there is on the network is sparsely spread and aging quickly.**
 - Over the course of our 78-day observation, nearly two-thirds of outlets did not publish a single human-written article.¹.
 - Over the same period, the median age of front-page stories in the network rose by 70 days.
- **An overwhelming majority of Metric Media's stories appear to be autogenerated.**
 - By our measure, 97% of Metric Media's content is autogenerated, and of that, >98% is never shown on the immediate "front page" of the outlets.
 - Certain strains of autogenerated stories are shown much more often than the rest. These often deal with state and national politics rather than local news.
- **Belying its claims about democratizing local news, Metric Media operates de facto hub outlets in each state;** on an average day, those hubs are the progenitors of ~45% of the content shown across the network of 1,000 sites, despite only making up 5% of the network's population.
- **Metric Media** altered its normal production patterns to give an outsized amount of coverage to stories about electoral fraud in states Donald Trump was contesting during the 2020 general election.
 - In the three most narrowly won states, stories about electoral fraud made up an average of 76% of all human-written content that was written during the 2020 election "contestation period".
 - Shortly after the period ended, overall story production by Metric Media outlets declined by an average of >80% in those states.
 - Compared to Metric media outlets in a control group of states, the amount of fraud coverage and the story production decline among outlets in the battleground states were much more extreme.

Overall, our findings indicate that Metric Media falls short in terms of meeting the information needs of the communities that the sites purport to serve. These findings raise concerns about the Metric Media model of trying to sustain a large number of outlets across the country by supplying them with autogenerated content and the rare human-written story. Though the financial prospects for local newspapers are suffering, automated large-scale national operations appear, in this case, to be a poor

¹ "Human articles" and "human-written content" as used in this paper refers to articles written by Metric Media journalists and excludes press releases Metric Media publications republished

substitute for the capital-intensive inputs of traditional local news. That said, as we note in our conclusion, the Metric Media model may offer some direction forward for those seeking to advance legitimate local news.

Author Bios



Asa Royal is an Associate in Research with the News Measures Research Project at the DeWitt Wallace Center for Media & Democracy in Duke's Sanford School of Public Policy. Asa graduated from Duke in 2019 with a degree in computer science, after interning for the *Tampa Bay Times* and working for the Duke Reporter's Lab. He worked as a software engineer before returning to the DeWitt Wallace Center in 2020.



Philip M. Napoli is the James R. Shepley Professor of Public Policy in the Sanford School of Public Policy at Duke, where he is also the Director of the DeWitt Wallace Center for Media & Democracy and the Senior Associate Dean for Faculty and Research. He is the principal investigator for the News Measures Research Project. His most recent book is [*Social Media and the Public Interest: Media Regulation in the Disinformation Age*](#).

Introduction

Over the past 15 years, more than 2,000 newspapers in the United States have shuttered, leaving the country pockmarked by news deserts (Abernathy 2020). Absent newspaper watchdogs, communities in news deserts have seen rises in local government borrowing costs, increases in government wages (Gao, Lee, and Murphy 2020), decreased competitiveness and engagement in local elections (Schulhofer-Wohl and Garrido 2013; Hayes and Lawless 2017), and the loss of over 36,000 journalism jobs (Abernathy 2020).

As newspapers have lost ground, a number of online local news networks have grown to take their place (Bengani, 2020; Nyhan 2019; Mahone and Napoli 2020). One of the largest, Metric Media, has launched over 1,000 local news sites across the country, with plans to launch an additional 15,000 (Graham 2020).

This influx of local news networks, which comes at a time when the economics of local journalism are growing more challenging, would seem to be a positive development for communities, an indicator of a potentially viable economic model for local journalism amidst the closure of traditional news sources. However, journalistic reports have raised significant concerns about how well outlets in these new online news networks serve communities' information needs. Many of these outlets are plagued by issues such as a lack of transparency in ownership and funding, a lack of original, local reporting, and a willingness to engage in pay-for-play political influence operations (Alba and Nicas 2020; Bengani 2019). Some, including Metric Media, have been cast as “pink slime” journalism (Bengani 2020; Tarkov 2012).

In light of the critiques and concerns these emerging local news networks have generated, their rapid growth highlights the need for greater empirical scrutiny of their operations. Toward that end, this paper presents a case study of Metric Media via a nearly 3-month observation of the home pages of the

company's 999 news outlets². Over the course of that observation period, we scraped and recorded more than 700,000 data points on the daily content composition of each Metric Media outlet's front page, as well as details on the more than 200,000 unique stories shown across the network. With these data, we address the following research questions:

1. Are there any notable patterns in the distribution of Metric Media sites, content, and journalists across the country?
2. How robust is the journalism produced by Metric Media sites in terms of its timeliness, originality, local orientation, and use of human reporting?
3. Is there evidence in Metric Media's reporting patterns of coordinated political influence efforts?

Method

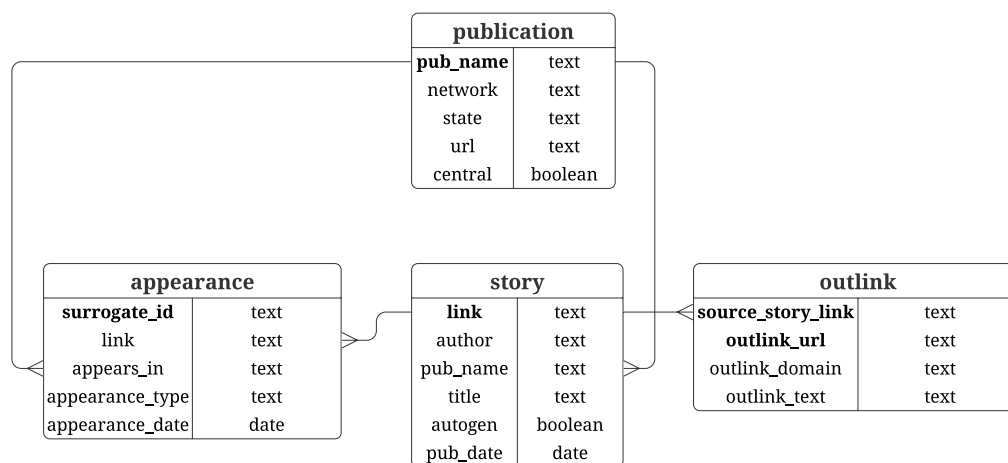
Identifying Metric Media Outlets and establishing a database schema

Our first step in tracking Metric Media's news production was to identify the individual outlets run by the company. To do so, we relied on one of Metric Media's corporate websites, which lists 965 outlets the company operates (Metric Media n.d.-a). We combined that collection of outlets with a network of 34 more Illinois-based news sites (operating under the name LGIS) that researcher Priyanjana Bengani (2019) of Columbia University found used the same servers and infrastructure as some of Metric Media's other outlets and employed identical website templates. At the time Bengani reported on their presence, the LGIS Illinois outlets had not been officially claimed by Metric Media. As of April 2021, one Metric Media corporate page now claims them (Metric Media n.d.-b), another does not (Metric Media n.d.-a), and LGIS' website still claims the sites as their own (Local Government Information Services n.d.).

² On one company page, Metric Media Claims to operate 1300+ "community news sites" but only lists 1,051 (plus nine publications dealing with "industry news"). On another page, the company only lists 967 local outlets. We operated from the latter list combined with another (see "Identifying Metric Media Outlets" below).

We decided to treat the combined networks as one, and with a list of 999 outlets in hand, we manually inspected the sites to determine what kind of story data was worth tracking. The resulting data schema (see Figure 1) was intended to keep track of two views of Metric Media: first, a daily record of all stories appearing on each outlet’s front page (recorded in the *appearance* table); second, a list of each unique story that appeared across the network (recorded in the *story* table).

Figure 1: Database schema



Collecting data (scraping)

To collect data, we wrote code³ that visited each site in the Metric Media Network, observed every front-page story in the immediate viewport of the site, and then recorded details in the relational database described in Figure 1. Our data collection period ran from November 16, 2020 to February 1, 2021. Given the scope of our data schema, we were able to conduct longitudinal analyses across the network, as well as reproduce point-in-time facsimiles of any news outlet’s front page on any date during the observation period.

Auto-generated story labeling

³ Code available at <https://github.com/oneroyalace/metric-media>

Before we began scraping data, we observed that the bulk of stories in Metric Media outlets followed strict patterns, wherein localized statistics were plugged into uniform story templates to produce “local” news. Further observation led us to a simple heuristic for identifying such stories: all content with authorship attributed to “Metric Media News Service” or attributed to a publication by name followed those templates. That heuristic was used to mark a large subset of stories (~207,000, 95% of the total) as autogenerated (“autogen” in the schema)⁴.

Following the observation period, we found roughly 1,000 more templated stories by matching all story titles in the database against a list of regular expressions that spanned titles we saw in the “autogen” set⁵. Many of the stories that escaped detection by our first heuristic were missed because authorship was attributed to a permutation of a publication’s name (e.g., a story in *East Central Georgia News* attributed to the author *EC Georgia News*). After detecting those stories with the regular expressions title match, they too were marked as autogenerated.

Finally, during manual inspection, we noticed two “strains”⁶ of stories that were reproduced according to strict templates but were not caught by our author heuristic or title matching. Those, too, were marked as “autogen” (see “Change in response to political events”).

Further assessment of autogenerated content

To buttress our heuristic-driven autogenerated content labeling, we analyzed^{7,8} intra-strain article similarity, assuming we would find that within template groups, most stories were near replicas. Using the representative list of story titles, for each title, we queried our database for a list of up to 200 articles in the corresponding strain (or 1% of all articles in the strain, whichever figure was greater). For each

⁴ We also tagged ~2,500 stories attributed to “Press Release Submission” as auto generated. As the attribution suggests, these pieces were word-for-word press releases reproduced from outside sources.

⁵ After building the set of regular expressions, about 80/211,000 heuristic-labeled auto generated stories (.04%) remained unmatched. Some of those stories appeared to be auto generated, but others were clearly human-written.

⁶ Used throughout the paper to refer to a group of auto-generated articles that follow a single template

⁷ Code available at <https://github.com/oneroyalace/metric-media>

⁸ 18 recorded story strains representing ~17,000 stories appear to have been taken down from Metric websites and were not included in this analysis.

strain, we randomly picked a “canonical” article from the sample group. Next, we fetched the text of every article in the sample group and compared it to the text of the canonical article using a simple bag-of-words technique: given two articles, we stripped them of proper nouns and stop words⁹, removed duplicate words from them, and considered them as sets of words. We then divided the size of the intersection of both articles — the number of words they shared — by the size of their union — their combined unique word count. The resulting metric (Jaccard Similarity) assigned a score of 1.0 to two documents with m and n unique words if their vocabularies were identical ($intersection/union = [m+n]/[m+n]$) and 0 to two articles if they had no words in common ($0/[m+n]$).

After running comparisons within each strain sample, we found the average strain had a mean similarity score of 0.92 (median 0.976), a result indicating that most auto-generated stories within a strain varied by at most a few words.

A discussion of autogenerated stories follows later in the piece, but we caveat it by admitting that we do not have direct knowledge of how Metric Media’s articles were authored or published. It is true that Metric Media advertises its creation of auto-generated content (“Metric Media” n.d.) and that the articles we labeled as autogenerated were nearly identical within strains, but it is conceivable, even if unlikely, that humans reporters *did* author, for example, the 20,000+ identical stories on local gas prices we observed. It is perhaps even more likely that humans had some level of input on the two story strains we manually labeled as autogenerated (see above), which totaled ~200 stories. Both were attributed to “Staff Reports” and one strain saw 143 permutations of a template published one day and a single new permutation published four days later, a pattern that seems unlikely to have emanated from a machine.

Evaluation of electoral aftermath content

After observing a large amount of content in Metric Media outlets that covered Donald Trump’s unfounded allegations of electoral fraud in the 2020 general election, we undertook a content analysis to

⁹ As identified by the Python library NLTK

find patterns in that coverage. Specifically, we looked to see whether stories about electoral fraud — which Trump alleged had robbed him of victory — were published more often in states whose election results the Trump campaign was challenging, and whether such coverage squeezed out local news in those states during the election “contestation” period. We chose to focus on coverage patterns in the three states won by the narrowest margins in the 2020 election (Georgia, Wisconsin, and Arizona), each of which had its election results challenged by the Trump campaign. In our analysis, we classified every human-written story produced by Metric outlets in those states from election day 2020 onwards as covering one of the following topics: electoral fraud; the presidential election; non-presidential elections; or “other”, a catch-all topic covering anything else, including local news.

To draw comparisons between Metric Media’s content production in the selected trio of contested states and production elsewhere, we randomly picked ten states¹⁰ outside the trio, conducting an identical content analysis on the stories produced by outlets in those states during the same period.

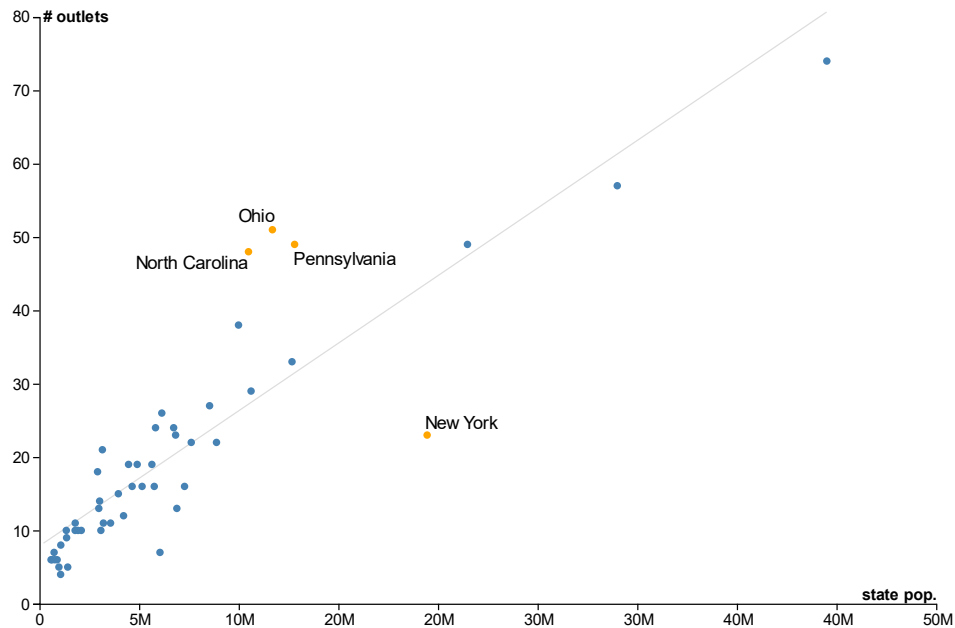
Findings

Distribution of outlets

Metric Media news outlets are roughly distributed across states according to population, though a few notable outliers (North Carolina, Ohio, and Pennsylvania) have significantly more outlets than would be expected and one, New York, has significantly fewer (see Figure 2). The three positive outliers were often considered potential swing states in the 2020 presidential election (FiveThirtyEight 2020; Wollner 2018).

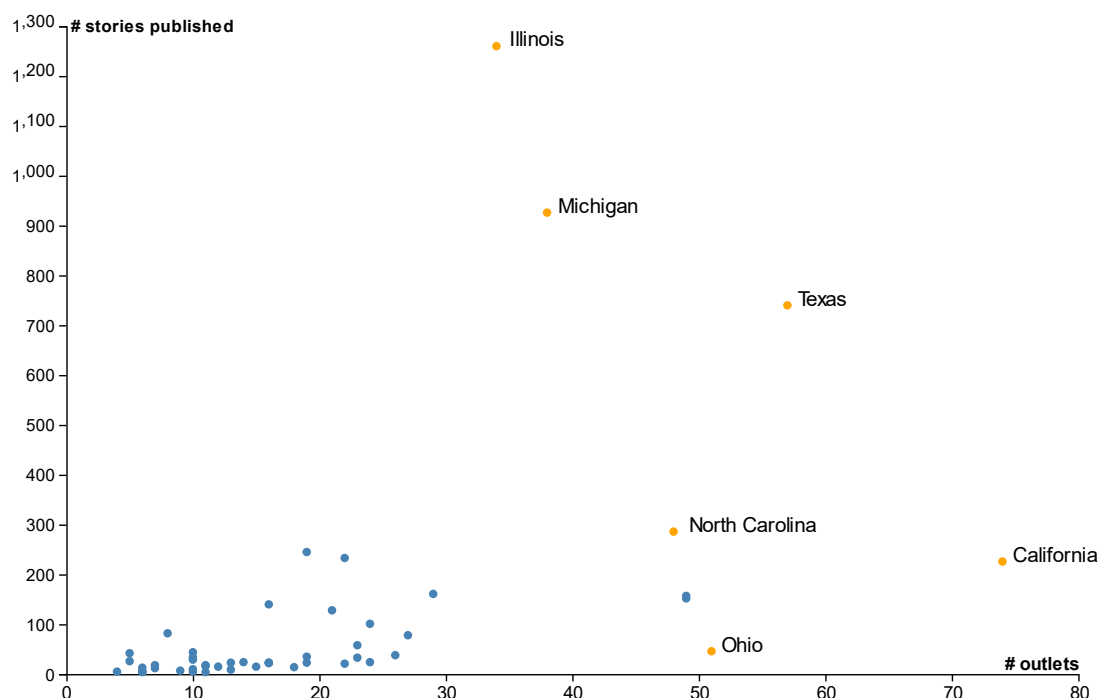
¹⁰ We limited the pool of states to those that had produced at least 10 stories between Election Day and Jan. 8, 2021 (see results for explanation).

Figure 2: Metric Media Outlets Serving State by Population



Surprisingly though, the number of Metric Media outlets in each state does not appear to be closely tied to the total amount of Metric’s journalistic output in that state (see Figure 3). While states producing the most content all have more than 30 outlets, an increased number of outlets beyond that point does not correspond with an increase in human-written journalism. Clearly then, there is a point of diminishing returns (in terms of human-written journalistic output) as the number of outlets in a state increases.

Figure 3: # of human written stories published in a state v. # of outlets



Distribution of authors

During the data collection period, we observed a total of 4,246 non-auto generated stories¹¹ shown across Metric Media’s front pages¹² and RSS feeds. Those stories were attributed to a total of 161 unique human bylines, but the distribution of authorship was highly skewed: the median author published just a single article, while the average author published 26.48 articles. Among the top ten most prodigious authors, the average published 236.5 stories.

Many authors wrote stories for publications in several states: ordered by that metric, the 20th percentile author published across 4 states, the median across 9.5, and the 80th percentile across 16 different states. Authorial output was even more dispersed at the local publication level (a proxy for

¹¹ These included stories with human bylines, as well as stories attributed to “Staff Reports”, “{Outlet Name} Reports”, etc.

¹² Those stories spanned 544 publications across 48 states. 455 publications never had a human bylined article on their front page during the observation period.

measuring how many localities individual authors reported in). There, the 20th percentile author published in 10.4 localities, the median in 30, and the 80th percentile in 75.6.

Story timeliness

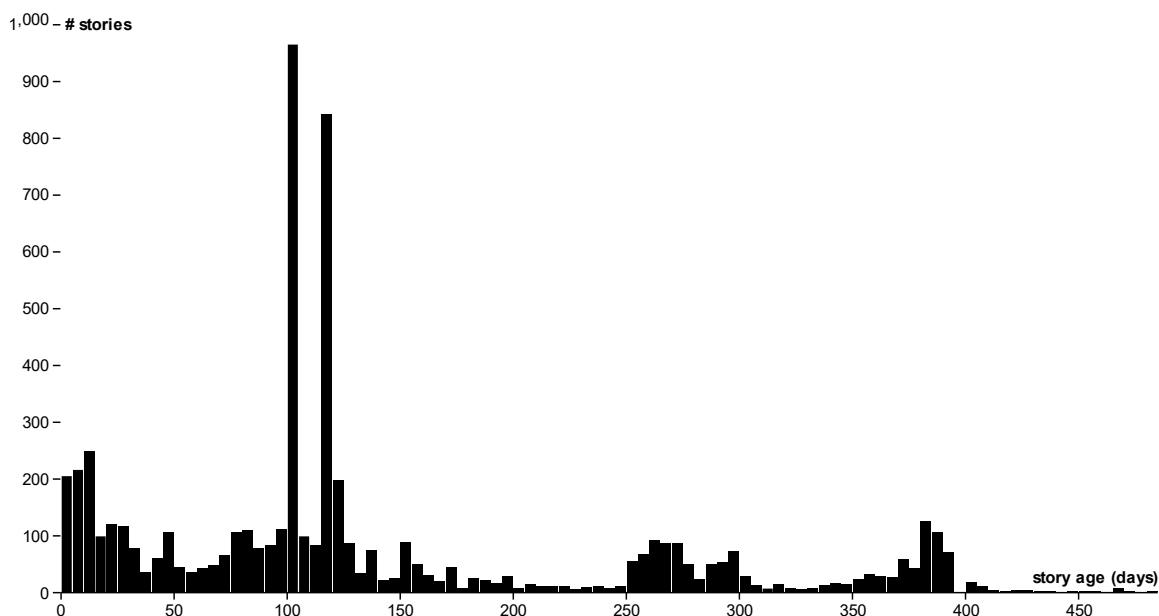
Over the course of the data collection period, we found that the median age of front-page stories across Metric Media’s 999 outlets was 81 days; with each passing day, that figure rose nearly linearly (see Table 1).

Table 1. Median age of front-page stories across all Metric Media outlets.

Date	Median age (days)
2020-11-16	40
2020-12-01	54
2020-12-16	69
2021-01-01	85
2021-01-16	100
2021-02-01	110

The aging nature of Metric Media’s front-page content was reflected in the number of recently published stories (liberally defined as those ≤ 2 weeks old) present on outlets’ front pages. The median publication (in fact, 761/999 publications) had zero “recent” stories, while the average, higher due to outliers (see “hub” outlets, below), had 0.49. Only a tenth of publications (97/999) had three or more “recent” stories on their front pages on the last day of our collection period, and less than 5% (42/999) had three or more stories less than a week old. The age distribution of front-page stories across all outlets on our final day of data collection is shown in Figure 4, below.

Figure 4: Stories on Metric Media front pages grouped by age (2021-02-01)¹³



Auto-generated content

Metric Media's most prolific authors do not appear to be human. A corporate website notes the company publishes “over 5 million news articles every month, including auto-generated data stories”. During our 78-day collection period we recorded the appearance of ~217,000 unique articles, 97% of which we judged to be auto-generated (see above). Those auto-generated stories, essentially mad-libs scripts, used templates to wrap data pulled from sources including U.S. Census manufacturing reports, Board of Elections donor lists, and local Mass schedules. Some story templates, like a strain reporting local gas prices from GasBuddy, a gas price tracker, were reproduced more than 20,000 times during the collection period and transplanted across nearly all outlets. To give a sense of what auto-generated stories looked like, we have included a selection of titles below:

“AT&T, Inc. employee Stacey Johnson Batiste contributes a total of \$2,000 to Democratic Party candidate Kamala D. Harris”

¹³ The two spikes at ~100 days and ~115 days are caused by the glut of outlets featuring stories from the Biden Tax and Cato governor grade autogenerated story strains on their front pages (see below).

“2 professional licenses were issued in Los Angeles during week ending January 16”

“Amid statewide drop in public school enrollment, Ysleta ISD saw a -5.33% decline”

“Cory Diez earns \$22,188 in compensation from Clark County in 2019”

“25 deaths from diabetes mellitus in Oklahoma during week ending November 21”

“Logan CC has an inmate population of at least 1,218 in September 2020”

“Lancaster Elementary School District had 1,618 seventh grade students during 2018-2019 school year”

“Paper and paper products wholesalers report 7.6 percent increase in year-over-year November inventories”

“822 seniors over 79 who were in US test positive for COVID-19 during week ending October 5”

Though most stories (97%) produced by Metric Media outlets were autogenerated, only 1.31% of all autogenerated stories were placed on the front page of any of the network's outlets (see Figure 5). The rest either sat in outlets' RSS feeds before being cycled out a few days later or were stowed below the viewport on the front page (see Figures 4 and 5, below).

Figure 5: Metric Media outlets sorted by fraction of front-page stories that are human written

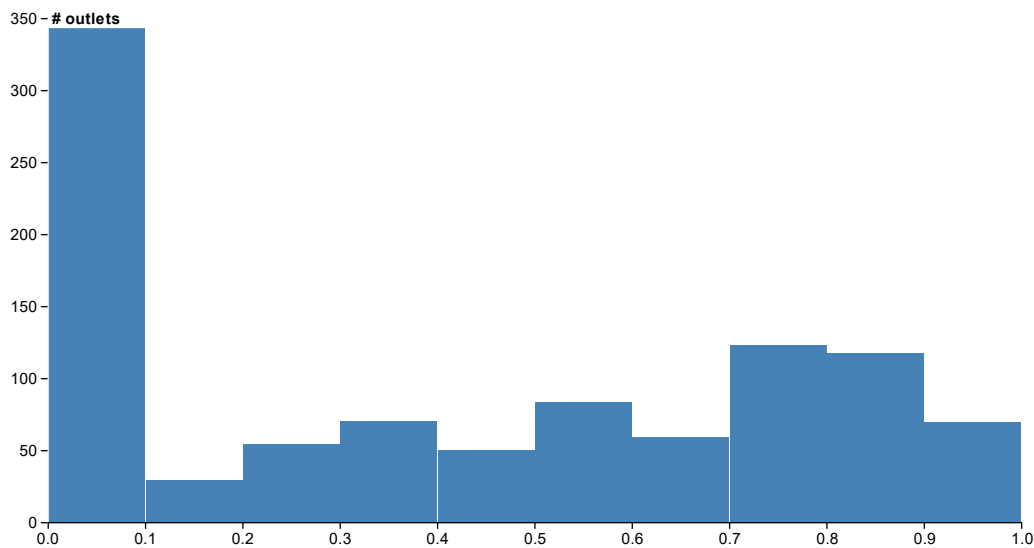


Figure 6: Screenshot of Asheville Reporter front page

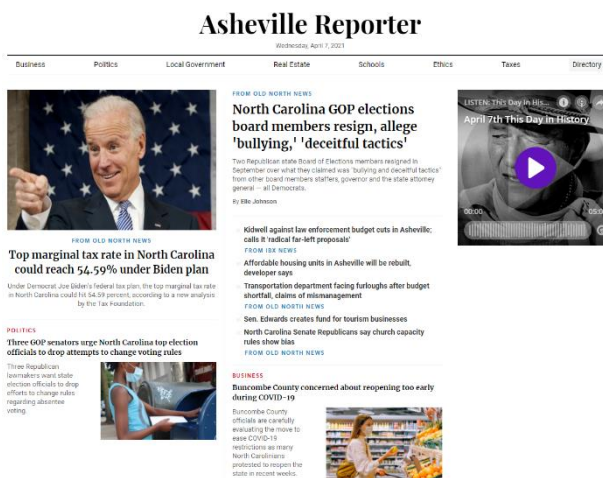
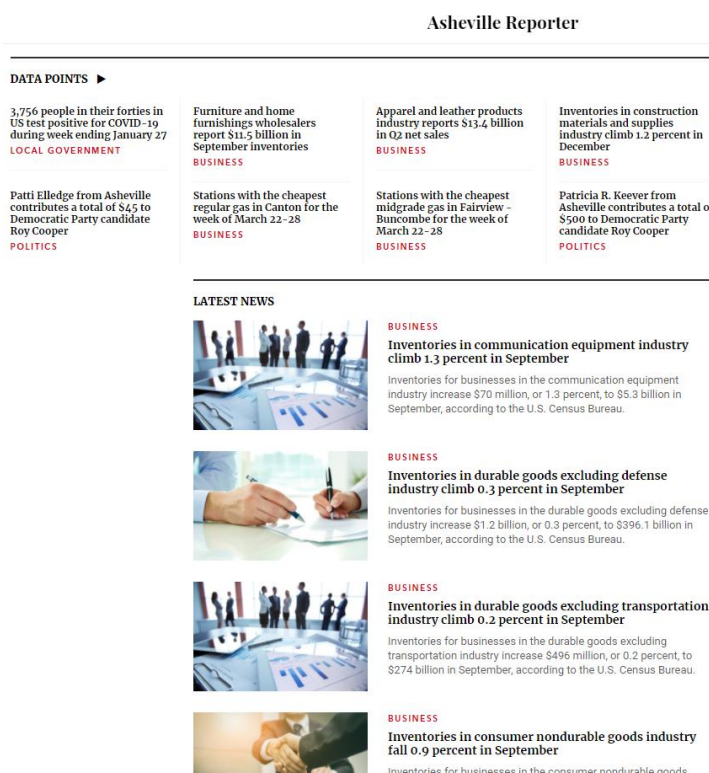


Figure 7: Screenshot of Asheville Reporter front page (scrolled below initial viewport to show autogenerated stories)



Generally, the small share of autogenerated stories that did reach the viewport on Metric's front pages was markedly different than the rest. Instead of featuring tidbits of local information from public datasets, these stories frequently discussed state and national politics. For example, a strain of stories published by state hub outlets titled "Top marginal tax rate in {state} could reach {x}% under Biden

plan” was prominently featured on the front pages of Metric publications in all 50 states (see, e.g. Figure 4); another strain, titled “{State} Gov. {Governor name} gets ‘{A-F}’ grade for fiscal management”, citing ratings by the conservative CATO institute, was published by 41 state hub outlets and ran on the front pages of outlets across those states. Together, the 91 auto-generated stories about Biden's tax plan and CATO governor grades comprised just .43% of all autogenerated pieces, but because of how often they were featured across the network, they made up 64% of all autogenerated story front page appearances.

During the observation period, we observed the elevation of several other autogenerated story strains. On November 18, 2020, shortly after Joe Biden's general election victory over Donald Trump, 143 stories ran across Metric's Georgia outlets, highlighting individual counties' absentee ballot rejection rates. The cookie-cutter stories suggested absentee ballots, the majority of which were believed to be cast in favor of Joe Biden (Doherty et al. 2020), were being rejected at anomalously low rates, a claim echoed by Donald Trump in his quest to discredit Biden's win (Reuters Staff 2020).

Though only ~1% of auto-generated stories were ever featured on the front pages of Metric Media's outlets, the Georgia election pieces had little trouble gaining such traction. A third (47) were featured on Georgia outlets' front pages, and 38 were still present at the end of our collection period, nearly two months after the election.

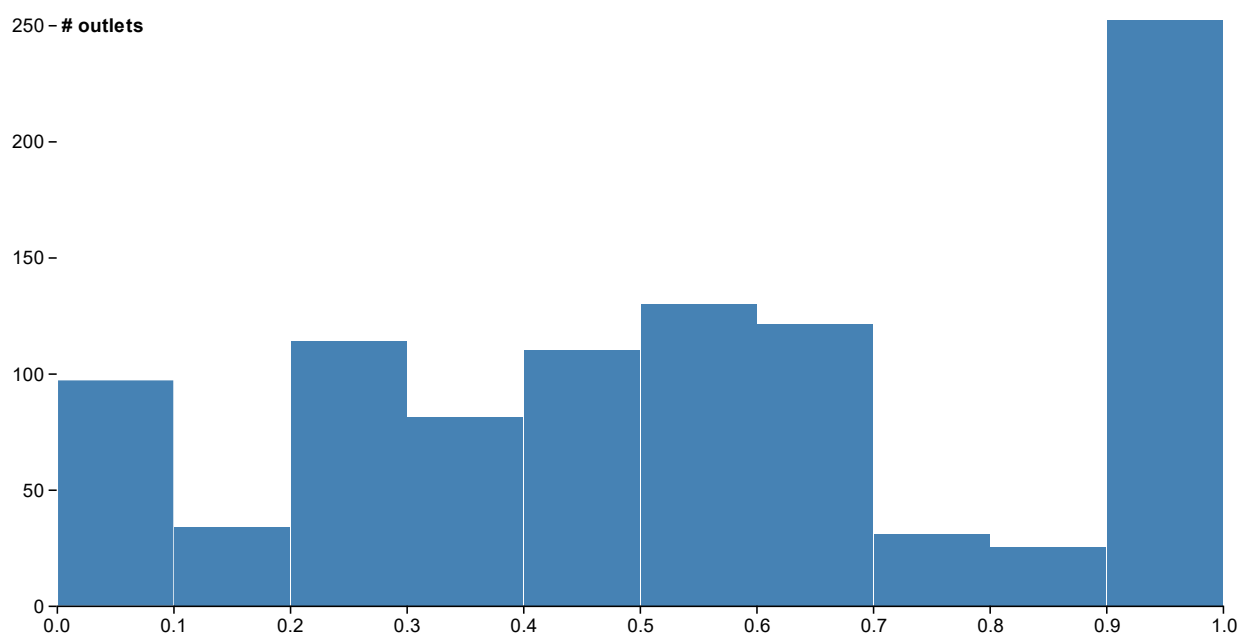
In a similar episode, following a Republican-led vote in the Ohio Senate to roll back the Ohio Department of Health's quarantine authority, Metric Media's Ohio outlets published 33 templated stories, each detailing how individual state senators had voted on the measure.

[Outsourced content and hub outlets](#)

During the data collection period, nearly two-thirds of Metric Media publications (637/999) never published a single human-written article. In place of original journalism, these sites featured articles published by other Metric Media outlets. Such article sharing generally did not take place between

geographically neighboring outlets, but rather flowed from a group of “hub” sites to the rest of the network. This collection of hub outlets consisted of one publication per state, each monikered by a state nickname or descriptor. On an average day, hub outlets were the original publishers of 43% of the stories displayed across every front page in the Metric network despite making up only 5% of the network population. Figure 8, below, shows the distribution of outlets by the percentage of their front-page content “borrowed” from hubs¹⁴.

Figure 8: Metric Media outlets sorted by fraction of self-published, human-written content on front page (2021-02-01)



Hub outlets were often outliers to the trend of aging content discussed above. Whereas only 11% of all outlets had at least one news story less than two weeks old on the front page, 64% of hub outlets did. Even when the front-page content of hub outlets did age, it usually remained much more up to date than those of other “local” publications in their respective states. We illustrate this in Figures 8 and 9, which graph the median story age across “hub” and “spoke” outlets in two representative states (North Carolina and Pennsylvania).

¹⁴ Hub outlets themselves make up fifty members of the rightmost bin.

Figure 9: Change in median front-page story age over time across Metric’s North Carolina outlets. Hub outlet highlighted

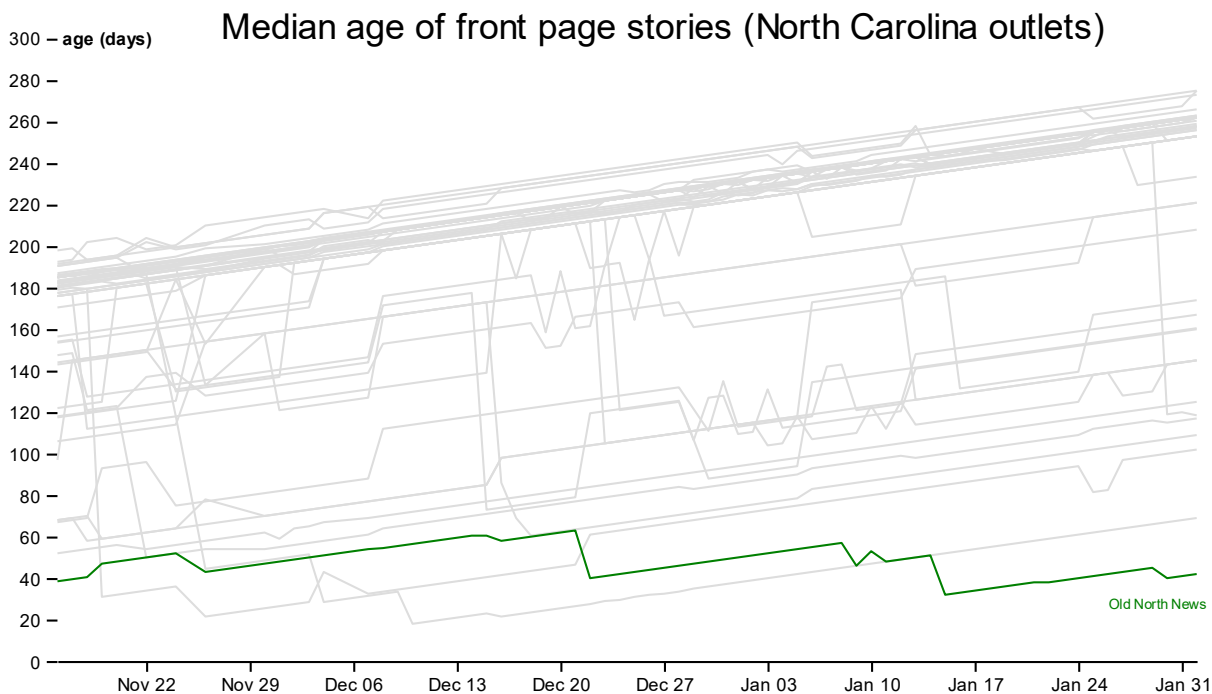
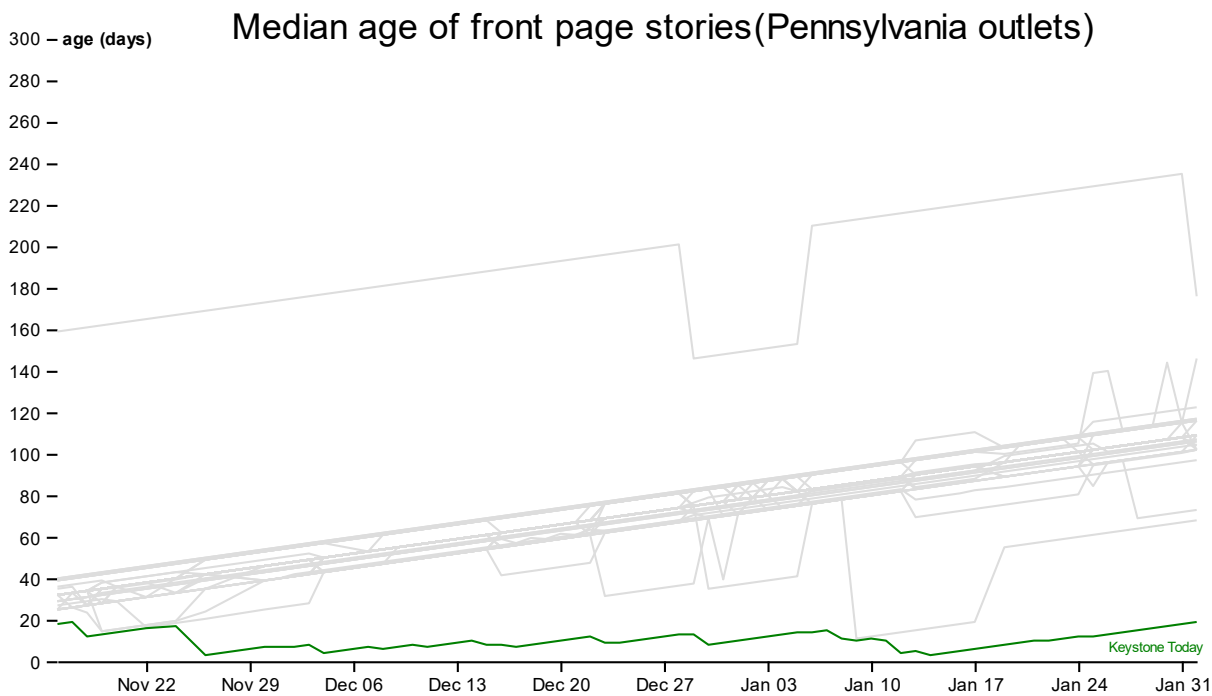


Figure 10: Change in median front-page story age over time across Metric’s Pennsylvania outlets. Hub outlet highlighted



Though we did not conduct a formal content analysis, a cursory observation of central outlets’ front pages makes clear that they deal almost exclusively in state and national political news.

Coverage of voter fraud

Bengani (2019) and Alba and Nicas (2020) have reported that in the past, Metric Media has given outsized positive coverage to conservative talking points and Republican politicians. We explored this possibility of coordinated political influence efforts using as a case study, Metric Media outlets' coverage of parts of the 2020 election. Though our observation period began too late track story production before and during the election, it was well fit to observe reverberations after the election, namely Donald Trump's unfounded challenging of the election results.

The data show that Metric Media news outlets in states with close presidential election margins that the Trump campaign was contesting (Georgia, Wisconsin, and Arizona), began to produce stories about electoral fraud during the election contestation period¹⁵ at a rate much greater than that rates seen in outlets from a randomly selected control group of states¹⁶. Outlets in the trio of “battleground” states also saw much steeper content volume drop-offs after the election contestation period than those from states in the control group, suggesting that overall content production in battleground states had been increased to pump out fraud stories.

For example, during the election contestation period, Metric Media publications in Georgia produced a total of 130¹⁷ human-written stories (94 came from the state's hub outlet, *Peach Tree Times*), of which 102 (78.46%) dealt with electoral fraud. But after the contestation period, from January 9, 2021 to February 1, 2021, the same Georgia outlets produced just 3 human-written stories, reflecting a 94.81% reduction in the number of stories published per day. Similar patterns appeared in Wisconsin and Arizona, in contrast to what we saw in control states (see Table 2, below).

¹⁵ From election day (November 3) 2020 day until January 8, 2021, a day after Congress officially certified the results of the 2020 Presidential Election.

¹⁶ The control group states were Florida, Hawaii, Illinois, Massachusetts, Missouri, Nevada, North Carolina, Pennsylvania, Texas, Virginia, and West Virginia.

¹⁷ In addition, as mentioned above, the Georgia outlets produced a series of 143 auto-generated/templated stories highlighting what they described as aberrantly low absentee ballot rejection rates in the 2020 election.

Table 2. Differences in Metric Media’s Battleground States’ Content.

State	% content covering electoral fraud (Election to Jan 9)	Stories/day (Election to Jan 9)	Stories/day (Jan 9 to Feb 1)	% change in stories/day
Georgia	78.46%	2.41	0.13	-94.81%
Wisconsin	71.15%	0.96	0.17	-82.69%
Arizona	76.00%	1.39	0.50	-64%
Median control	16.22%	0.69	0.29	-38.59%¹⁸

Discussion

Our analysis suggests that Metric Media’s claim to being “the largest producer of local news content in the nation” (Metric Media n.d.-c) is true only in a narrow sense. The company operates nearly a thousand sites and produces a large quantity of content, but to call most of its content “news” or most of its sites “news outlets” conflates the thin veneer of local coverage Metric Media publishes across the country with an operation dedicated to producing substantive local news.

Outlet and author distribution

In a 2012 Poynter Institute article, a reporter summarized a Metric Media executive’s comments on the (lack of) need for locally-placed journalists: “young reporters cycle in and out of the areas they cover without knowing much about them. By that logic, why not give that work to someone outside the area ...?” (Tarkov 2012).

Relying on autogenerated stories and distant reporting means journalists will miss important stories, and their reporting may lack local context. As discussed above, the median Metric Media author¹⁹

¹⁸ The discrepancy between this number and the stories/day figures for the control group is brought on because the control stories/day figures are group medians. Calculating a change from those figures is not meaningful, because it conflates raw production levels across states. The figure we arrived at, -38.59%, was calculated by finding the change in story production within each state, then taking the median of that list.

¹⁹ Filtering out authors who had not published at least 5 pieces.

published stories across 9 states and 30 localities. Given the difficulties of capturing even a single local context, it is highly unlikely that Metric Media's authors, spread so thin, can produce a genuinely locally-oriented news product. Our first research question dealt with the broad distribution of Metric Media's outlets and journalists. In addressing that question, we found that though the company has spread nearly 1,000 outlets across the country, including many into news deserts, the bulk of the outlets lie almost bare, dusted with a shimmer of one-off stories, but lacking substantive community-based reporting.

Robustness of journalism

In our second research question, we asked how robust Metric Media's journalism was in terms of its age, originality, local focus, and use of human reporting.

Age, originality, and local focus

Metric Media presents its outlets as the online equivalents of daily newspapers. But in all but a select few of them — mostly hub outlets catering to state and national political news — original stories are either nonexistent or stale, a conclusion exemplified by the fact that only one third of the 999 outlets we tracked published a single non-autogenerated story during the 78-day observation period.

It is highly unlikely that a local newspaper or its online equivalent can meet the information needs of its readers if it does not publish for 78 days. And to the point of local focus, for most outlets, as discussed above, there is none. Though state hub outlets often share stories with local Metric outlets, such content lending is not a substitute for original local reporting and does nothing to address local information needs.

A case study in Ohio illustrates the gaps left by Metric Media's sparse coverage. In October 2020, after the company took over the 80-year-old independently owned *Mount Vernon News*, a water main in Mount Vernon, Ohio burst, leaving an eight-foot crack in the pipe and spurring a city-wide boil advisory (Pepper 2020). For several days, the Metric Media-owned newspaper failed to cover the water main burst, leaving city leaders scurrying to send out text alerts to constituents. In the words of a Republican city

council member, the paper “failed to alert its citizens” and was “basically just a bunch of press releases” (Hendrix 2020).

Autogenerated content

Most of Metric Media’s auto-generated stories are never shown on the front pages of an outlet. Such content — auto-generated and out of sight — like the nearly 7,000 pieces we recorded about individuals’ professional licenses expiring, made up 97% of all Metric Media stories we observed.

Most of it did not constitute substantive local news. Though some articles had a kernel of interesting data — for example, the percentage of minority students graduating from high schools in a given school district — they lacked any context that might explain the data: they were facts without stories. The news value Metric Media assesses its autogenerated content to offer seems evident in how often the network cites it. Within the text of all human-written Metric Media articles, we observed almost 3,600 total links. 99 pointed to stories by Metric Media outlets, but only one pointed to an auto-generated story. That autogenerated stories account for most of Metric Media’s content suggests that the network’s goals are not fully aligned with the production of robust journalism for local communities.

2020 electoral fraud coverage

As Donald Trump and his Presidential campaign attempted to overturn the results of the 2020 general election on the false pretense of electoral fraud, Metric Media outlets in states the campaign contested gave large amounts of coverage to claims about and battles around electoral fraud. In Georgia, Wisconsin, and Arizona — the three states carried by the narrowest margins in the presidential election — stories discussing electoral fraud made up an average of 76% of all human-written content from Metric outlets. Stories unrelated to the presidential election made up an average of less than 18% of human-written content in each state.

Beyond taking up coverage space and raising doubts about Metric Media claims to be interested in local news, the company's outsized coverage of electoral fraud revealed a predilection for slanted coverage of partisan politics. That finding is not without precedent. A *New York Times* investigation of Metric Media uncovered a pay-to-play political operation in which Republican operatives ordered up articles from "local" Metric sites (Alba and Nicas 2020).

Conclusion

Metric Media operates neither professionally nor economically like a genuine local news network. Most of its outlets do not publish original, human-written content and those that do — primarily state hub outlets — are usually not focused on local news. The freelance journalists Metric employs report across several states and localities, which means they are unlikely to accrue the reservoir of knowledge about any locality that enables journalists to become effective watchdogs or providers of community information. For Metric Media, opening another "outlet" in a new community is as simple and cheap as registering a new web domain, but as our data show, that community will often get nothing more than a templated web page sharing auto generated content and non-local partisan news. An understanding of the shortcomings of the Metric Media approach to local journalism is important in light of both the established size of the network (nearly 1,000 sites nation-wide) and its stated plans to launch an additional 15,000 sites across the country.

Despite the failure of Metric Media to live up to its mission of restoring the local news environment, its model hints at at least one new way to produce substantive local journalism that, while not likely to be taken by the company in question, could be used by others.

Metric's use of common public datasets to fuel the creation of local stories across the nation suggests local news outlets may be able to do the same. Operating with unified resources, individual Metric outlets benefit from a constant stream of organized data at almost no marginal cost. It does not

take much imagination to imagine local news organizations belonging to a consortium or organization set up to provide the same — easily scraped or parsed state/national data scoped to individual localities. Right now, if a local outlet cannot train and pay a reporter or third-party company to gather such information, they may not bother gathering it at all, depriving the community of valuable data. A pooled data solution stands to prevent that or, at the very least, free up reporters' time for other stories.

The distribution of locale-specific data might even be taken a step forward to the distribution of locale-specific auto-generated content. Stories cost money to produce, and while the cookie-cutter nature of Metric's stories is concerning at scale, given that local news readers are unlikely to read the news of other localities, the inexpensive nature of publishing or improving on a mad-libbed story may outweigh the benefits of using reporters' time to write an original story about an anodyne bit of data. Indeed, local news organizations could improve on the Metric Media method by *not* publishing about every anodyne bit of data: it is conceivable that the centralized data pipeline described above might feed data to local news organizations only if the data is either anomalous (e.g., surprisingly low graduation rates at high schools in a city one year) or generally considered newsworthy, like local unemployment rates.

Using auto-generated stories and data briefs might make one of the most attractive bits of Metric's story history — a series of pieces distributed across Ohio outlets detailing how individual state legislators had voted on a contentious bill — something easily achievable at scale for low-resourced local news outlets: each time a local representative took a vote, an alert might go out to the relevant local news organizations. And, in the hands of knowledgeable reporters and editors, those alerts might easily be adapted into stories that, unlike the Metric Media pieces, impart meaningful local context.

In sum, much of the recent research on the state of local journalism has focused on documenting the decline of traditional local news sources and estimating the effects that these declines are having on various aspects of community well-being. This study suggests that we should not only be concerned about the declines in traditional local news sources, but also about some of the new local news sources that are arising in their place.

Bibliography

- Abernathy, Penelope. 2020. "Vanishing Newspapers." *The Expanding News Desert* (blog). 2020. <https://www.usnewsdeserts.com/reports/news-deserts-and-ghost-newspapers-will-local-news-survive/the-news-landscape-in-2020-transformed-and-diminished/vanishing-newspapers/>.
- Alba, Davey, and Jack Nicas. 2020. "As Local News Dies, a Pay-for-Play Network Rises in Its Place." *The New York Times*, October 18, 2020, sec. Technology. <https://www.nytimes.com/2020/10/18/technology/timpone-local-news-metric-media.html>.
- Bengani, Priyanjana. 2019. "Hundreds of 'Pink Slime' Local News Outlets Are Distributing Algorithmic Stories and Conservative Talking Points." *Columbia Journalism Review*. December 18, 2019. https://www.cjr.org/tow_center_reports/hundreds-of-pink-slime-local-news-outlets-are-distributing-algorithmic-stories-conservative-talking-points.php/.
- . 2020. "As Election Looms, a Network of Mysterious 'Pink Slime' Local News Outlets Nearly Triples in Size." *Columbia Journalism Review*. August 4, 2020. <https://www.cjr.org/analysis/as-election-looms-a-network-of-mysterious-pink-slime-local-news-outlets-nearly-triples-in-size.php>.
- Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.
- Doherty, Carroll, Jocelyn Kiley, Nida Asheer, and Calvin Jordan. 2020. "Mail-in Voting, Absentee Ballots and Voter Engagement." *Pew Research Center - U.S. Politics & Policy* (blog). October 9, 2020. <https://www.pewresearch.org/politics/2020/10/09/voter-engagement-and-interest-voting-by-mail-and-in-person/>.
- FiveThirtyEight. 2020. "Ohio President: General Election Polls | FiveThirtyEight." May 14, 2020. <https://web.archive.org/web/20200514221932/https://projects.fivethirtyeight.com/polls/president-general/ohio/>.
- Gao, Pengjie, Chang Lee, and Dermot Murphy. 2020. "Financing Dies in Darkness? The Impact of Newspaper Closures on Public Finance." *Journal of Financial Economics* 135 (2): 445–67. <https://doi.org/10.1016/j.jfineco.2019.06.003>.
- Graham, Jennifer. 2020. "Understanding 'Pink Slime Journalism' and What It Reveals about Conservatives and Liberals." *Deseret News*. September 7, 2020. <https://www.deseret.com/indepth/2020/9/7/21409053/understanding-pink-slime-journalism-and-what-it-reveals-about-conservatives-and-liberals>.
- Hayes, Danny, and Jennifer L. Lawless. 2017. "The Decline of Local News and Its Effects: New Evidence from Longitudinal Data." *The Journal of Politics* 80 (1): 332–36. <https://doi.org/10.1086/694105>.
- Hendrix, Sheridan. 2020. "New Mount Vernon News Owners Linked to Conglomerate of Pay-to-Play Republican News Sites." *The Columbus Dispatch*. November 5, 2020. <https://www.dispatch.com/story/news/2020/11/05/new-mount-vernon-news-owners-linked-network-partisan-news-sites/6039173002/>.
- Local Government Information Services. n.d. "Our Publications | LGIS." Accessed April 7, 2021. https://web.archive.org/web/20210407190035/https://lgis.co/our_publications.
- Mahone, Jessica, and Philip Napoli. 2020. "Hundreds of Hyperpartisan Sites Are Masquerading as Local News. This Map Shows If There's One near You." *Nieman Lab* (blog). July 13, 2020. <https://www.niemanlab.org/2020/07/hundreds-of-hyperpartisan-sites-are-masquerading-as-local-news-this-map-shows-if-theres-one-near-you/>.
- Metric Media. n.d. "About — Metric Media Foundation." Accessed March 15, 2021a. <https://web.archive.org/web/20200303005207/https://www.metricmedia.org/about>.
- . n.d. "Metric Media." *Metric Media News*. Accessed April 6, 2021b. <https://web.archive.org/web/20201101012102/https://metricmedianews.com/>.
- . n.d. "Metric Media." Accessed April 28, 2021c. <https://web.archive.org/web/20210407191207/https%3A%2F%2Fwww.metricmedia.org%2F>.

- “———.” n.d. Accessed April 7, 2021d.
<https://web.archive.org/web/20210407191207/https://www.metricmedia.org/>.
- . n.d. “Metric Media News.” Metric Media News. Accessed April 6, 2021e.
<https://web.archive.org/web/20201101012102/https://metricmedianews.com/>.
- Nyhan, Brendan. 2019. “Americans Trust Local News. That Belief Is Being Exploited.” *The New York Times*, October 31, 2019, sec. The Upshot. <https://www.nytimes.com/2019/10/31/upshot/fake-local-news.html>.
- Pepper, Grant. 2020. “City Officials: Aging Infrastructure Likely Caused Water Line Break.” KnoxPages. October 16, 2020. https://www.knoxpages.com/news/city-officials-aging-infrastructure-likely-caused-water-line-break/article_607c2df0-0f9f-11eb-8561-07e774d3927e.html.
- Reuters Staff. 2020. “Fact Check: Georgia Rejected Ballots Did Not Go from 4% to ‘Almost Zero’ in 2020.” *Reuters*, November 23, 2020. <https://www.reuters.com/article/uk-factcheck-georgia-rejected-ballots-fo-idUSKBN2832CM>.
- Schulhofer-Wohl, Sam, and Miguel Garrido. 2013. “Do Newspapers Matter? Short-Run and Long-Run Evidence From the Closure of The Cincinnati Post.” *Journal of Media Economics* 26 (2): 60–81. <https://doi.org/10.1080/08997764.2013.785553>.
- Tarkov, Anna. 2012. “Journatic Worker Takes ‘This American Life’ inside Outsourced Journalism | Poynter.” Poynter. June 30, 2012.
<https://web.archive.org/web/20140702081758/http://www.poynter.org/latest-news/top-stories/179555/journatic-staffer-takes-this-american-life-inside-outsourced-journalism/>.
- Wollner, Adam. 2018. “‘The New Swing States:’ Presidential Battleground Map Shifts Heading into 2020.” McClatchy Washington Bureau. December 12, 2018.
<https://www.mcclatchydc.com/news/politics-government/election/campaigns/article222917745.html>.