

Bridging the Divide Between Local News and Online Platforms

**Discussing the challenge of
surfacing local news online**

**Schema-Online Local News Discussion Series
July 2021**

An initiative of

**CREDIBILITY
COALITION**

Nieman

Contents

Introduction.....	1
Defining Local.....	3
Signaling Localness and Originality	4
Breaking News: Speed Versus Quality.....	6
Struggling to Surface Local News Online.....	6
The Search for Solutions	8
Giving Control to Local Audiences.....	9
Toward a New Model	9
Conclusion	11
Resources	12
Questions?	13

Bridging the Divide Between Local News and Online Platforms

By the Credibility Coalition and the Nieman Foundation for Journalism at Harvard

In early 2021, the [Credibility Coalition](#) (which receives funding support from Craig Newmark Philanthropies and other organizations), in partnership with the [Nieman Foundation for Journalism at Harvard](#) convened a series of conversations aimed at both establishing a common vocabulary for greater understanding between local news outlets and online platforms and examining the way local news is surfaced online. The project, which received financial support from the Google News Initiative, is part of a wider effort to empower journalists to better understand — and platforms to better execute — the promotion of quality local journalism on news delivery platforms.

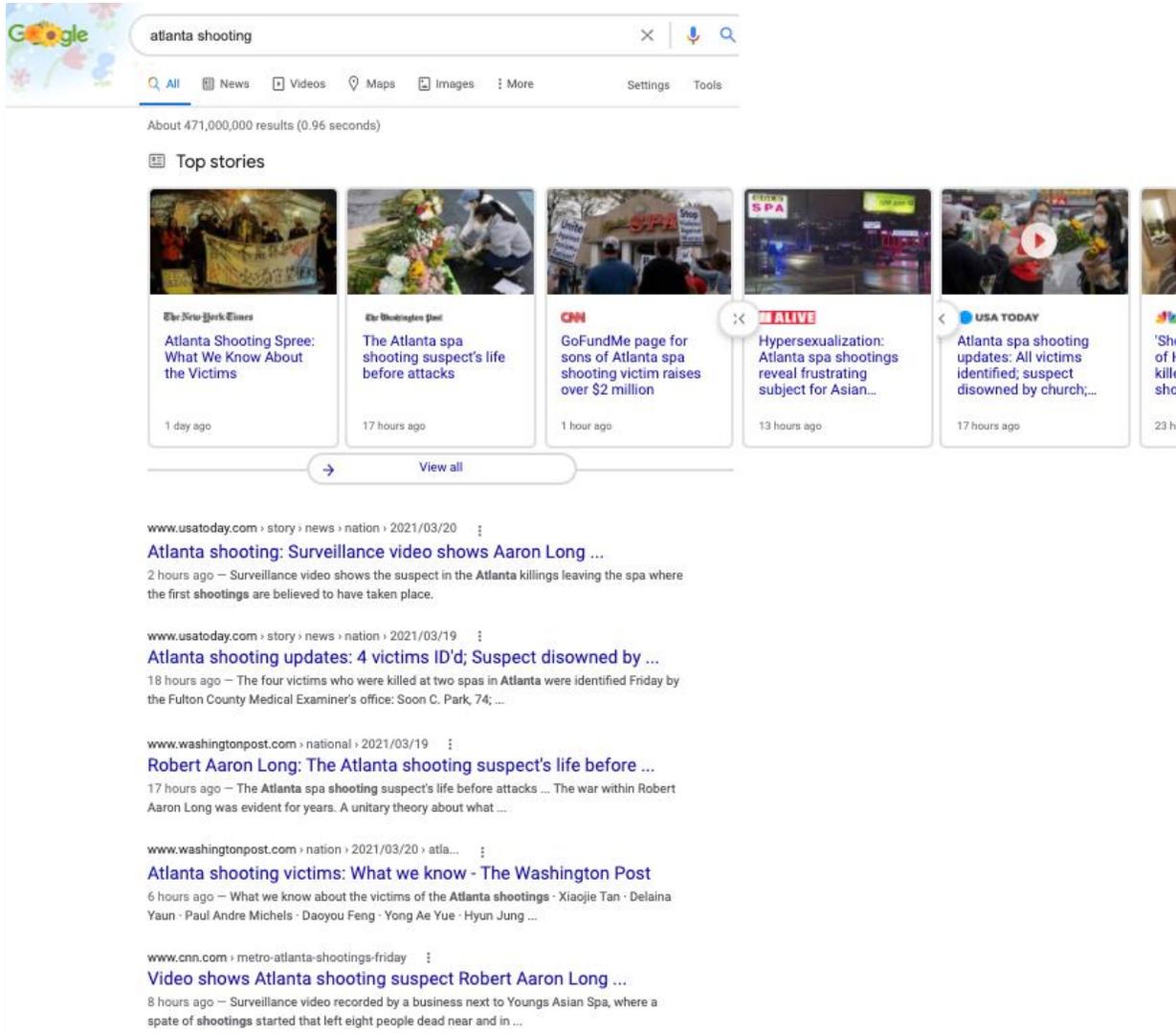
The Schema-Online Local News discussion group consisted of journalists from a variety of outlets (with majority representation from local news operations), and representatives of several of the major tech platforms. Over the course of three meetings, the group examined ways that journalists might better adapt local content to effectively communicate with platform technologies. At the same time, they looked at ways platforms might improve systems for identifying local news and tweaking their algorithms to place local stories more prominently in front of readers.

Introduction

News outlets today have the unprecedented ability to share their stories with readers not only locally, but throughout their regions and across the world. Despite the expanded reach the internet provides, many local news organizations struggle to achieve visibility in an online environment awash with content.

The algorithms that technology platforms such as Google, Facebook, and Microsoft use to determine which news stories are presented in search results or news apps often favor national publications or large news organizations, including wire services, over smaller outlets. Reporters for local news organizations have noted with increasing frustration instances where their newsrooms have broken stories only to find them “ranked,” appearing much lower in online searches than those that come later from larger competitors. Many newsroom leaders note that this happens even in instances where readers might best be served by search results that lead

them to the most thorough, on-the-ground local reporting that is closest to the story. They have also shared examples where, even if a larger outlet gives a local publication proper credit, the national outlet version of the story appears much higher in search results, and the associated online traffic rarely finds its way back to the original news source.



An example where national coverage of a local event appears higher in search results than reporting by local outlets. On March 16, 2021, a series of mass shootings occurred at three massage businesses in the metropolitan area of Atlanta, Georgia, United States. Four days after the incident, the top results when searching for “Atlanta shooting” from an Atlanta-based IP address were from The New York Times, The Washington Post, and CNN. [See full screenshot of search results.](#)

As news increasingly moves online, these search rankings, online traffic metrics and website referrals matter a great deal to the bottom lines of many local news operations. It is not only important that these local outlets get credit for their reporting; ranking higher in search results can help justify the resources they invest in newsgathering and reinforce their image as an authoritative, trusted source for readers.

Failing to surface their content has become another setback for local news outlets already struggling with cutbacks, shrinking subscription bases, and increasing pressure to produce more content with ever-dwindling resources. A vibrant local news landscape plays a vital role in keeping the public informed on issues that directly impact their lives. If local outlets can't reach readers online, the whole news ecosystem suffers.

Defining Local

Although the Schema Discussion Group only examined one small aspect of the multifaceted crisis facing local news, it focused on an issue that many local journalists say needs urgent attention: the struggle to surface more local news results online. One of the first challenges the group discussed was the lack of a definition of local news that is universally understood by both platform representatives and journalists. This has emerged as a sticking point in trying to develop computer algorithms that recognize something as local using the same criteria that a news editor might. It has proven challenging to devise a common definition of the term that is also acceptable to both journalists and computer programmers. Developing uniform definitions and vocabulary around what is and isn't local news is an important first step in creating systems to share it.

The first definition of "local" that the group discussed is common to many newsrooms: a story produced *by* an outlet in a particular area, *about* events in that area, *for* readers in that area. It is easy, however, to think of cases in which one or more of those elements may not be true. A New York Times reporter based in Seattle, for example, could break a story about a local issue there, even though the Times is a national newspaper. Or a local Seattle publication could cover a national story. Put another way, "local" could apply to a journalist, an audience, or an outlet. And "local relevance" can vary between stories published in the same outlet: depending on where it is from, where it is sourced, and where it is produced.

These definitions become more nuanced, however, when drilling down further into what the concept of local means in different contexts. As the Schema Group participants discussed, "localness" might be defined as a series of concentric circles expanding outward from a neighborhood or zip

code to a municipality, county, state or metropolitan statistical area (MSA). In addition, readers might self-identify — and publications might target — specific local groups within a larger area, such as African Americans in the South, or the LGBTQ community in California. For some readers, the definition of local may have nothing to do with their location, but is instead defined by where they work, where they *previously* lived, where their family is from or by their particular identity or affinity groups. A reader's concept of local may also change depending on a variety of circumstances. For example, a reader who lives in Arlington County, Virginia, and works in Washington, D.C., might be interested in local news from both areas at different times, depending on relevance.

For platforms delivering local content, it is also important to be able to determine user intent and interpret signals that point to the kind of news a reader may be seeking at a particular time, in order to surface the sources that might best serve them. In addition to previously mentioned criteria such as location, user intent can also change depending on whether readers are seeking a local, regional, or national take on a particular issue or from a specific outlet. Search results may also differ depending on the platform being used: for example, a reader might look for a certain type of story through search engines such as Google or Bing, and another type from Facebook or Twitter. Despite these differences, there is an argument to be made for including at least some local outlets in top search results, regardless of platform.

These varying definitions of “local” present challenges for platforms in developing algorithms that simultaneously present useful stories to readers and give local publications credit for their work. Human judgment can help to obviate some of these challenges, but it is difficult and expensive to scale. Human curation may also lack the full context of each reader's news environment in order to make informed decisions.

So, how should computer algorithms understand all of this? How can programmers incorporate these nuances to be able to surface the most reliable or relevant version of a story and better prioritize local news in search results?

Signaling Localness and Originality

Although the discussion participants representing platforms said they were not authorized to divulge the inner workings of their proprietary algorithms, the journalists in the group identified many ways they feel the algorithms devalue local news sources.

Several journalists in the discussion group shared ways they feel the signals that platforms use to identify concepts like “originality” may be insufficient. One platform’s algorithm may determine that an outlet had a story first by seeing that it contains the first instance of a particular quotation, for example. However, other platforms may define “first” as the first time any version of that story is posted on its *own platform*. In such cases, the algorithm would give priority ranking to an article’s timestamp over the originality of its content. In that case, if a news outlet is late to upload a story to a particular platform — even if it broke the story — it might be easily overshadowed in search rankings by competitors.

In addition, sometimes local publications fail to provide the proper signals that would identify their stories as relevant to local audiences, such as using datelines in stories to indicate their location. Television and radio stations, for example, may use station call letters on stories, such as WCVB, which mean little to an algorithm that rewards explicit location-based signals, such as ABC Boston. Also, discussion participants pointed out that headlines are often written with a core local audience in mind, rather than in a way that might convey local expertise to a wider national audience.

Another potentially confusing issue the group identified is trying to determine whether a platform algorithm is prioritizing recency over originality in news story rankings. For example, while a local outlet may have the advantage of providing authoritative reporting early on for a story breaking within its coverage area, if an algorithm prioritizes recency, that local outlet can quickly be overshadowed by larger national publications with more resources to frequently update its version, thus pushing it higher in search results. Platforms may struggle with determining how long to keep an initial local story at the top of search results, when another publication may have advanced the story or consolidated up-to-date information from multiple sources. Some participating platform reps admitted it is sometimes a challenge to determine how important a story development must be to warrant replacing a newer version of a larger publication’s story for a local one that may be closer to the scene.

This issue is particularly challenging when the lines between local and national are blurred in a story. For example, in March 2021 in Atlanta, Georgia, there was a mass shooting whose victims were predominantly Asian women working in massage parlors. The Atlanta Journal-Constitution was one of the first outlets on the ground reporting the story, and the paper developed many of the storylines that other outlets would follow, including: the role that religion may have played in the alleged shooter’s motivation, a GoFundMe campaign set up to support the victim’s families, and other developments that other outlets followed. National

publications quickly picked up the story and started advancing the reporting with wider themes of national significance, such as a nationwide spike in hate crimes perpetrated against Asian Americans. Although these national publications were clearly being guided by developments from the Journal-Constitution, the paper's stories were supplanted by competitors and driven lower in search result rankings.

Breaking News: Speed Versus Quality

A similar situation happened when searching for news results about a Black motorist who had been killed by a white police officer in Brooklyn Center, Minnesota, around the time that white former Minneapolis police officer Derek Chauvin was on trial for the murder of George Floyd, a Black man whose death sparked global protests. When the Schema Discussion Group participants searched for articles about the Minnesota incident in real time, news from Minnesota-based outlets was not returned in the results.

Instead, the results were seemingly based more on recency and outlet size and included stories from national outlets such as USA Today, ABC News, and The View, as well as reports from television stations in Vermont and Tennessee. One discussion participant, based in Canada, was initially shown an article by CBC News — an outlet local to the reader, but not to the story itself.

In part, the search prioritization of large and/or national news outlets may have been a result of the *time* the user was undertaking the search. Search experiments conducted during the discussion sessions seemed to indicate that platforms may prioritize “live” results, which offer the most current updates on an unfolding story, over more detailed summaries later in the news cycle. Several discussion participants said such a structure can create pressure on under-resourced local outlets to emphasize speed and constant incremental story updates over quality, in-depth reporting, which is less algorithmically rewarded. Publications dependent on web traffic and clicks say they are disadvantaged by these current search algorithms, and that their cumulative effect contributes to the declining quality of local coverage overall.

Struggling to Surface Local News Online

The group's examination of recent news articles revealed that the largely automated methods developed by technology platforms seem insufficient to properly surface local news in search results or news feed products. Often the signals are skewed toward larger outlets and seem to obscure the original sources of stories generated by local outlets.

Within these constraints, online platforms currently have different ways of identifying and surfacing local news. For example, platforms use natural language processing (NLP) to look for signals within stories that may indicate local or original content. By attempting to automatically identify locations within stories, platform products can deliver these stories to users based on geolocation data, so that all readers within a certain media market will receive the same location-based content. However, such results depend on what a specific platform's product is designed for. For example, search engines — unlike news aggregators — are designed to deliver relevant results for a user's particular search.

In contrast, news aggregator algorithms are designed more to guess at a user's intent and then recommend/surface relevant articles for them. However, as mentioned in the discussion about the challenge of defining localness, these processes sometimes result in a mismatch between a reader's conception of local and their physical location when accessing platform technologies. They also often seem to prioritize larger outlets' coverage of local issues, or stripped-down versions of local stories that run on The Associated Press wire, for example.

Also mentioned in the discussions was the idea, which some platforms have experimented with, of enabling publications to tag their content as "breaking news" as a signifier that a story is new or original *to that outlet*. While some publications have used this effectively, it emerged in the schema group discussions that others have abused the breaking news tag to generate additional clicks on their stories while contributing little original content.

In addition, it was pointed out that large, national publications have the institutional capacity to "flood the zone" with content, which has the potential to manipulate algorithms to push their content to the top of search results when news stories are prioritized based on their number of clicks. Some news organizations, such as television news stations, can also garner clicks and push their search results higher by doing condensed summaries of local stories from other outlets, rewriting them, and posting them online, where they might appear higher in search results than the outlet that originated them. The same is often true with a wire service version of a local story. Once a local story is picked up and disseminated nationally (and sometimes worldwide) by the AP, for example, even though the AP credits the original outlet in its version, the AP version — or another outlet that is posting the AP's version on its website — often appears higher in search results than the original story. Several local journalists who participated in the schema discussion group spoke of their frustration at seeing a story their newsroom broke or originated being

returned to them in search results as an AP version posted to a newspaper site that might be several states away.

The Search for Solutions

Despite the formidable challenges in surfacing local news, the journalists and technologists of the Schema Team worked to brainstorm possible solutions. These included both practices that might be implemented by organizations to optimize their content within existing algorithmic frameworks, and new tools and techniques that platforms might implement to better feature local stories.

On the news production side, journalists discussed additional ways that their articles might more accurately signal local content. For example, publications might expand the practice — which some platforms already use to identify origin — of labeling original quotes with terminology such as “as told to The Seattle Times.”

Apart from these types of internal signifiers, however, algorithms or programmers can only make judgments on what they are able to “see” or read, which in most cases does not include information on which outlet originated a story or what went into the reporting process. Participants also discussed the idea of publishers providing additional information in a Content Management System (CMS) plugin, which outlets could use to mark stories with information such as the length of time spent reporting, the number of sources interviewed, or content of local interest, which could give platform algorithms more indicators of originality.

Another participant suggested that in addition to calibrating headlines and subheads to appeal to national audiences, publications might create two versions of a story: a broader national audience version and a more locally focused one. For example, if a hurricane is heading for Florida, a local audience may be more concerned with practical information such as where to seek shelter, whereas a national audience might be drawn to coverage of extreme weather.

The discussion group’s ideas for platforms and programmers included a suggestion that platforms or third parties develop databases of trusted local news organizations with track records of original reporting to regularly feature in search results. There are existing databases of pre-vetted newsrooms, a participant pointed out, that they might use.

Building on that idea, a smaller discussion group focused on ways that platforms might add information for readers about the outlets returned in search results. One suggestion was to include an information button that

readers might click to learn about news outlets they might not be familiar with. Another suggestion involved offering readers more detailed information about the curation process for search results — a kind of “Why am I seeing this?” feature that would appear alongside results. In addition to developing a database of news outlets, some platforms have shown interest in developing a database of reporters, since journalists often change jobs, taking their knowledge and expertise with them.

Since “local” can mean so many different things to readers, the group also discussed how platforms might give users more control in setting their own preferences for local news. Currently, some platforms allow readers to select local news settings, for example: location preferences within Facebook, or the “News Near Me” setting on Twitter’s “trending” page. In most cases, however, readers are not given the choice of which outlets are shown to them in search results. Instead, most platforms automatically customize news based on what kind of stories a reader usually engages with — a process that can encourage filter bubbles or vicious cycles of prioritizing larger national publications over local ones.

Giving Control to Local Audiences

Some Schema discussion participants proposed that users be given an option to activate checkboxes for local publications of interest, which would give those news sources priority in search results. They could include publications which readers already subscribe to, so they could be accessed without a paywall. Giving select publications more weight in search results, the group argued, would better serve readers with more relevant local content, and help publications see a direct benefit between user engagement and digital subscriber retention.

Other participants suggested that additional user-centered options be incorporated into the design of search result pages. For example, one group suggested that a widget could be added to search results to elevate local news, or readers could select an “alternative view” based on a prompt to see local news sources. Another idea was to cluster results for similar stories from different media markets, and let users decide which sources to click on based on their own publication preferences.

Toward a New Model

One of the smaller breakout groups within the larger discussion decided to workshop an idea for a more comprehensive search system that an online platform might use to provide readers with local content. The group proposed that platforms surface at least one story by a local news source in

the first page of all search results. The breakout group defined a local news source as either an outlet covering the area where the news occurred, or one located near the reader. The group used as an example either a local event of national interest or a national event with local implications.

In the first case, the group imagined fictional news consumer (a cartoon character from the TV show “The Simpsons”), Marge Simpson. In the example, Marge is contrasted with her daughter Lisa, a more discerning news consumer who seeks out a variety of news sources, and Marge’s husband Homer, someone who pays little attention to the news. In contrast, Marge understands the need to be informed but does not actively seek out particular sources of information.

In the example outlined by the group, Marge learns of an accident at the nuclear power plant in her hometown of Springfield. While such a story would be of national interest, Marge’s immediate needs as a resident of Springfield would include needing specific information on how she and her neighbors might be immediately affected. In the group’s example, if Marge typed “Springfield nuclear power plant” into a search engine, at least one (preferably two), of the top three news stories returned would be from local outlets and tagged with labels indicating local relevance.

Meanwhile, people living beyond the location of immediate impact might be more interested in a broader, more nuanced or “big picture” perspective of the incident. For this example, the group came up with a fictional Seattle resident named Dour Starbuckian, whose online search results regarding the Springfield accident would return some larger, national publications but should also include at least one local publication in Springfield, especially during the early stages of the developing story when the local outlet would be closer to the scene.

For a different type of news story with local and national implications, however, the group argued that the reader profiles of Ms. Simpson and Mr. Starbuckian might change. For example, if they each searched separately for coverage of the passage of the CARES Act (which provided financial assistance during the Covid-19 pandemic), “local” might be defined as close to where they live — Springfield versus Seattle. The group suggested that in this case, the top three search results should yield at least two state or local stories that could help Marge and Dour understand the direct implications of the law for their specific household or community.

The purpose of fixing local news is to serve people who have local needs. Suggesting (and clearly tagging) local stories would allow readers to self-select coverage of most interest to them, rather than trying to algorithmically guess on their behalf. This could help news organizations

provide needed information to readers while rewarding their original content with clicks and potential subscribers. More importantly, the group argued, surfacing these local results could give readers more consistent access to reliable journalism produced in their own communities and help them stay better informed about the issues affecting their lives.

Conclusion

The issues this group of journalists and technologists explored during this discussion series raised some important questions about the ways in which local news is surfaced and prioritized online. Among the questions the group explored:

- How do platforms determine reliable markers of a local story when news is breaking and a story is evolving? And how can platforms best serve readers with unique local content in those situations?
- How can algorithms determine when user intent changes, in order to offer them the most relevant sources, whether local or national in scope?
- How can platforms credit the original work of local publications while simultaneously deprioritizing and disincentivizing outlets from repackaging the work of others (including summary versions of stories on the AP wire) rather than reporting the story themselves?

These were just a few of the issues that emerged during the discussions, which participants hope will continue between journalists and technologists. Several participants from local newsrooms said that platforms/news search engines should take more responsibility for the negative impact their algorithms are having on the local news ecosystem.

“We must pivot this question, immediately. These are not challenges for local news. These are challenges for democracy, challenges for technologists and challenges for platforms,” one participant wrote. “Local journalists cannot fix these issues, because they do not control the mechanisms that prioritize and disseminate news and information to the majority of the U.S. public. The longer we continue to foist responsibility for addressing these problems onto local journalists, the less likely we are to succeed in curbing the damage.”

Resources

The following sources were compiled to provide the Schema-Online News discussion series participants with a common starting point for considering the landscape of local journalism, and how to define it.

Thanks to Kate Harloe of the News Quality Initiative (NewsQ), who compiled these resources.

Definitions of “Local” and “Local News”

- [Is it Time for a New Definition of Local News?](#) (Kristen Hare for Poynter, 2021) “...We need to think long and hard about what it means to be local in the digital age and the communication technologies that provide local media.”
- [What is Local News?](#) (Dan Kennedy for Editor & Publisher, 2019)
- [What are the Local News Dynamics in Your City?](#) (Pew survey, 2019)
- [Most People Think Local Journalism is Financially Healthy. Here’s the Troubling Reality.](#) (Margaret Sullivan for The Washington Post, 2019)
- [For BIPOC Communities, Local News Crisis Extends Beyond Major Cities.](#) (Columbia Journalism Review, 2020)
- [Digital News Report.](#) (The Reuters Institute and the University of Oxford, 2020) A broad analysis of how local news is perceived across the globe.
- [The Future of Local News in New York City.](#) (Sam Ford for The Tow Center at Columbia, 2018) This report seeks a definition of local news: “We need to expand our thinking as to what constitutes local news and who practices local journalism. We need to include ethnic media, community media, user-generated media in our discussions.”
- [The Sociology of News](#), by Michael Schudson, 2011. Information on historical definitions.

Computational Definitions

- [Defining Local News: A Computational Approach.](#) (Nick Hagar, Jack Bandy, Daniel Trielli, Yixue Wang, Nicholas Diakopoulos, 2020) References in this article also point to additional definitions from a computational perspective.

A definition of **audience-oriented geolocation**:

“Our overall approach involves analyzing geographic patterns in a news outlet’s Twitter audience.

Specifically, it consists of (1) matching news outlets to Twitter accounts, (2) collecting IDs of users who follow a news account on Twitter, (3) collecting the location field from a statistically significant sample of those users’ profiles, (4) geocoding the location fields into geographic coordinates, and (5) analyzing geographic patterns in the geocoded locations.”

- [Towards a measure of Community Journalism](#). (Mass Communication and Society, 2008)

Impact of Platforms on the Local News Ecosystem

- [“The Local News Landscape is Broken: NewsQ Panel Review of Platform Products.”](#) (Gabriel Kahn with Meredith Clark, Al Cross, Claudia Irizarry Aponte, Mandy Jenkins, David Kroman, 2020) “Local news means news produced by or in collaboration with reporters who are part of the community they serve.”
- [Google and Facebook Have a News Labeling Problem](#). (Emily Bell and Sara Sheridan for Columbia Journalism Review, 2020)
- [How Facebook Stole the News Business](#). (Josh Constantin for TechCrunch, 2018)
- [Don’t Stop the Presses](#). (Henri Gendreau for Wired, 2017) “[...] Today nearly all new digital ad revenue goes to Google and Facebook, leaving only crumbs for the rest of the publishing industry.”

Projects Seeking to Support or Reinvent Local News

- [Center for Innovation and Sustainability in Local Media](#). (University of North Carolina)
- [Institute for Nonprofit News](#)
- [LION Publishers](#)
- [A Playbook for Launching a Local, Nonprofit News Outlet](#). (The Shorenstein Center on Media, Politics and Public Policy and the News Revenue Hub)
- [Local News Lab](#). (Columbia University’s Brown Institute)
- [Local News Digital Adaptation Index](#). (Tow Center for Digital Journalism)

Questions?

If you have questions or would like to learn more, contact the Schema-Online Local News discussion series team:

schema@credibilitycoalition.org