

# **An Introduction to Schemas for Journalists**

**Understanding and applying markup to  
online local news stories**

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

**An initiative of**

**CREDIBILITY  
COALITION**

**Nieman**

# Contents

Introduction.....	1
What Is a Schema? .....	2
Why Do Schemas Matter to Journalism? .....	6
What Is Schema Markup? .....	8
Schema Markup in the News .....	11
A Markup for Fact-Checking.....	12
What Are Signals? .....	15
New Signals and Schemas for News .....	16
What Are Labels?.....	17
Questions? .....	18
Glossary.....	19
Additional Resources.....	20

# An Introduction to Schemas for Journalists

Understanding and applying markup to online local news stories

*Produced in partnership by the [Credibility Coalition](#) and the [Nieman Foundation for Journalism](#) at Harvard and written by Samantha Sunne.*

## Introduction

The following curriculum is meant to empower journalists to better understand — and help platforms to better execute — the discovery and ranking of quality local journalism on news delivery platforms.

This document was developed as part of the Schema-Online Local News discussion series that was convened over several months in 2021, by the [Credibility Coalition](#) and the [Nieman Foundation for Journalism](#) at Harvard, and with financial support from the Google News Initiative. The purpose of this discussion series was to establish a common vocabulary for understanding between local news outlets and online platforms by convening regular, structured conversations between journalists and representatives of technology platforms.

Journalists from local news outlets and representatives from technology platforms participated in one introductory training module and three subsequent discussion sessions. Participants worked together in small groups to brainstorm ideas, processes or mechanisms they think platforms should adopt to better surface local news. At the end of the discussion series, these ideas were refined and submitted to platforms for consideration.

In order to facilitate these conversations, both journalists and platform representatives needed to be able to understand and use common vocabulary around news algorithms. This vocabulary, and in particular the way it represents how online information is organized and processed, can be known as a “schema.” Schemas are unfamiliar to many journalists, but nonetheless have a huge impact on their work and how it is distributed.

## What This Document Covers

This document provides a hands-on introduction to practical applications of schemas for online news, including:

- Structured data
- Rich snippets
- Schemas for online news articles
- Applying schema markup and Schema.org markup
- New initiatives to markup news content
- Signals, indicators, and ranking factors
- Labels
- Additional resources for further exploring how local news is distributed online

A short glossary is provided at the end of this document that defines potentially unfamiliar concepts and vocabulary.

## What Is a Schema?

A schema is one way to wrangle data — which can be messy, uncategorized, and enormous in scope — into a structure that is easier for a machine to understand. Since the goal of this guide is to better understand how news is distributed online, we will explore the schemas — or “structured data” — that help an algorithm (such as Google News) understand a piece of content (like a news article).<sup>1</sup>

Here’s one example: every day, CNN publishes dozens, if not hundreds, of stories on its website. Search engines like Bing want to know if these articles would be good search results. For example, if someone searches for “Jeff Bezos” on Bing, should it return an article about him on CNN.com? To know the answer to that, Bing and other search engines need to process these articles — not just on CNN.com but at every news outlet in the world.

That’s way too much for a human to sort through, and it’s even too much for many computers. Yet news algorithms, such as the ones behind Facebook and Google News, aspire to understand the content of every article to determine what they think are valuable to readers, such as:

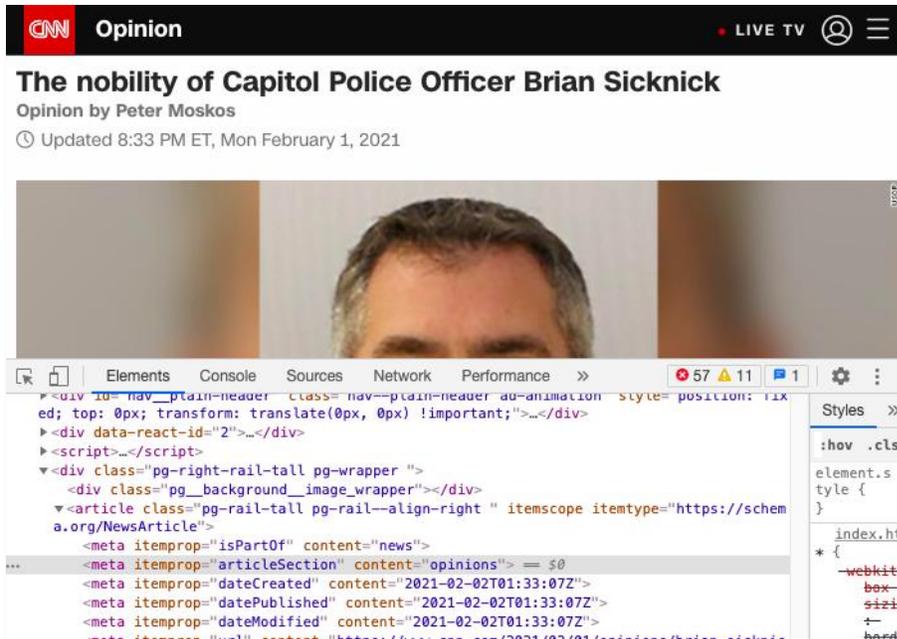
- Relevance, for instance, topics such as sports
- Recency, or when the article was published

---

<sup>1</sup> “Schema.org Markup,” MOZ.com, accessed June 3, 2021, <https://web.archive.org/web/20210603041718/https://moz.com/learn/seo/schema-structured-data>.

- Length
- Accuracy, such as if it is fact-checked

To help search engines find and categorize information, the CNN website applies tags to news stories, such as `opinion` and `datePublished`. These tags are an example of a “news schema” in action.



The highlighted code in this screenshot can signal to a computer that this article is classified under an outlet’s Opinions section. This helps a search engine categorize it, which could prevent it from landing in “harder” news sections of online news apps.

For example, the screenshot above shows that CNN considers the article to be “opinion,” which can help a search engine distinguish CNN’s opinion content from hard news. The tags are embedded into HTML, the code that presents web pages published on CNN. There are other ways to add these tags as well.

Technical know-how is not necessarily required to create or design schemas for online news. [ClaimReview](#), one of the examples mentioned in this guide, got its start in a conversation between a Google employee and a fact-checker at The Washington Post.<sup>2</sup>

<sup>2</sup> Shanha Lim, “A Better ClaimReview to Grow a Global Fact-check Database,” Reporters Lab, April 18, 2019, <https://web.archive.org/web/20210123075453/https://reporterslab.org/a-better-claimreview-to-grow-a-global-fact-check-database/>.

Of course, as an individual, your ability to influence the universe of online news by tagging stories is limited. However, algorithms are the powerful engines behind sorting, prioritizing, and distributing news online. Understanding schemas and structured data can help journalists tackle some of the challenges these algorithms present.

Let's get started!

### **Try It Yourself: Structured Data**

Before we dive in, let's get a sense of the difference that structured data makes when organizing information online, and why this organization matters.

#### **Previewing “Rich Snippets” With Structured Data Linter**

When a webpage contains structured data, a search engine can present a preview of the page with additional information — called a “rich snippet” — in a search result. To understand better, open a tool called the [Structured Data Linter](#).<sup>3</sup> This tool is called a “linter” as a reference to a machine that separated fibers from seeds as part of the cotton ginning process. In this case, Structured Data Linter is going to show you how a web page is separated into bits of structured data.

The tool is run by the [Structured Data Initiative](#), a group that advocates for, as you may have guessed, structured data.<sup>4</sup> As we move through this exercise, keep in mind that structured data is data that has been organized neatly into parts, making it easier for web browsers and search engines to read.

---

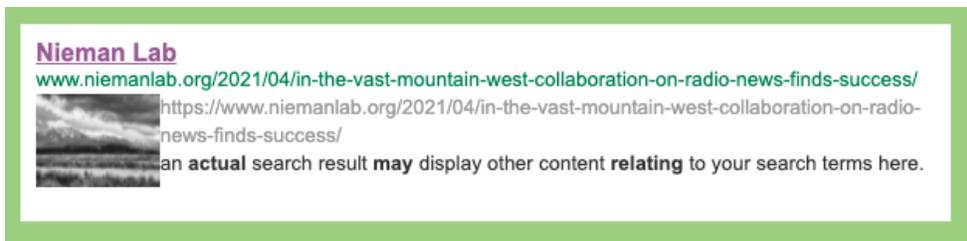
<sup>3</sup> Structured Data Linter, accessed May 12, 2021, <https://web.archive.org/web/20210512000504/http://linter.structured-data.org/>.

<sup>4</sup> “Structured Data on the Web,” Structured Data Linter, accessed February 26, 2021, <https://web.archive.org/web/20210226205237/http://structured-data.org/>.

To get started:

1. Paste the link to this [Nieman Lab article](#) in the [Structured Data Linter](#)
2. Click “Submit”

As you’ll see in the screenshot below, the Structured Data Linter tool will generate a preview of the structured data that makes up the Nieman Lab article. The tool will also generate a preview of what the article might look like as a Google Search result.



*This “rich snippet” shows that the computer, in this case using Google Search, recognizes different bits of data in this Nieman Lab article, like the image, site name and URL.*

The preview above, which includes an image, a publisher, a URL, and other elements, is called a “[rich snippet](#).” Schemas and structured data can be used to generate rich snippets, which not only have higher click-through rates, but also help audiences quickly understand information they encounter online.

Below the preview, the Linter tool will also generate a table of the various bits of structured data the linter separated out: the site name (Nieman Lab), the title (the article headline), and so on.

Raw structured data extracted from the page:

@id	https://www.niemanlab.org/2021/04/in-the-vast-mountain-west-collaboration-on-radio-news-finds-success/(0)
rdfs:type	rdfs:Resource http://types.ogp.me/ns#article
rdfs:label	In the vast Mountain West, collaboration on radio news finds success
article:author	
fb:pages	29057838653
og:description	"During the pandemic, we made all of our content free to smaller, community radio stations that don't have a news budget."
og:image	https://www.niemanlab.org/images/sawtell-peak-and-the-centennial-range-idaho-cc-700x394.jpg
og:image:height	394
og:image:width	700
og:site_name	Nieman Lab
og:title	In the vast Mountain West, collaboration on radio news finds success
og:type	article
og:url	https://www.niemanlab.org/2021/04/in-the-vast-mountain-west-collaboration-on-radio-news-finds-success/
rdfs:comment	"During the pandemic, we made all of our content free to smaller, community radio stations that don't have a news budget."
https://api.w.org/	https://www.niemanlab.org/wp-json/
@id	http://types.ogp.me/ns#article(1)
rdfs:type	rdfs:Class

Table detailing some of the structured data included on a web page. Screenshot from [Structured Data Linter](#).

### Compare with Non-Structured Data

In contrast, try using the Structured Data Linter to examine [this plain HTML web page](#).

As you will see, since this plain HTML page has no structured data, the tool will not return a nicely designed preview, let alone bits of information separated out for a search engine to understand.

### Why Do Schemas Matter to Journalism?

Relevance, along with accuracy and truth, is listed as a core element of reporting in "[The Elements of Journalism](#)," a widely read introductory text.<sup>5</sup> Journalists have a professional obligation to ensure stories are relevant, and a well-applied schema can help ensure this responsibility is fulfilled. Journalists can't directly control how an algorithm distributes their work,

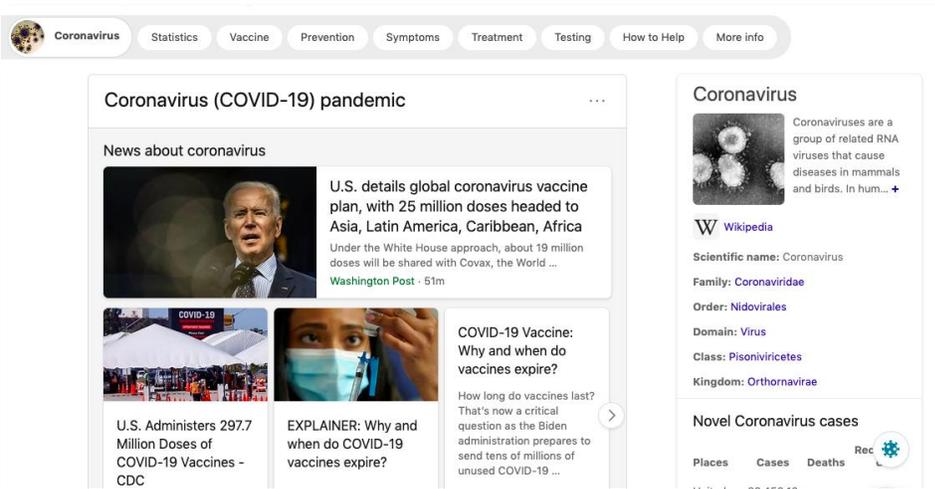
<sup>5</sup> "The Elements of Journalism," American Press Institute, accessed March 28, 2018, <https://web.archive.org/web/20190328172041/https://www.americanpressinstitute.org/journalism-essentials/what-is-journalism/elements-journalism/>.

but schemas offer an opportunity to indirectly help their audience find accurate, relevant information, by way of helping search engines.

Besides relevance, many journalists also want to help their audience find recent, accurate, informative local stories about whatever topic they're searching for.

Technology platforms such as Google, Facebook, and Microsoft state similar goals with their news products. For example, Bing uses personalized results in their interpretation of "relevance" to encourage individual users to consistently return to use their search engine.

For news and media organizations that publish content online, news algorithms are a powerful force for distribution, traffic, and revenue. In 2020, Facebook alone was responsible for 16% of traffic to news sites. A few years earlier, a mistake in Facebook's traffic analysis resulted in many journalism organizations "pivoting to video" and [losing many readers and staff members](#) as a result.<sup>6</sup> Google Search sometimes will generate even more traffic than that for news publishers.<sup>7</sup>



Some information on this Bing page, including headlines, images, section headers, and URLs, is powered by structured data.

<sup>6</sup> "How Facebook's Chaotic Push Into Video Cost Hundreds of Journalists Their Jobs," The Atlantic, accessed June 18, 2021, [https://web.archive.org/web/20210618113151if\\_/https://www.theatlantic.com/technology/archive/2018/10/facebook-driven-video-push-may-have-cost-483-journalists-their-jobs/573403/](https://web.archive.org/web/20210618113151if_/https://www.theatlantic.com/technology/archive/2018/10/facebook-driven-video-push-may-have-cost-483-journalists-their-jobs/573403/).

<sup>7</sup> Echobox Team, "4 Insights into Social Media Traffic from our Social Media Index," Echobox, accessed May 19, 2021, <https://web.archive.org/web/20210519170805/https://www.echobox.com/resources/blog/social-media-is-driving-traffic-to-news-websites-and-one-source-dwarfs-all-others/>.

While some news publishers and news sites have direct relationships with online platforms, journalists cannot control how an algorithm chooses to distribute their stories. Traditionally, editors and publishers would place an article in a specific spot on a page or website for readers, allowing them much more control over how it is received within the outlet's own space.

However, articles in news feeds are not organized according to this original formatting context. This is why journalists and journalism organizations should consider learning to curate things like metadata — information about their content — to influence algorithms in some way.<sup>8</sup>

Besides influencing how their own work is organized and displayed online, understanding schemas and structured data provides journalists across the industry access to a powerful vocabulary. By understanding schema definitions, journalists can more easily communicate with and encourage technologists to consider the potential impacts of their algorithms on the long-term survival of our vital local news ecosystems. By potentially informing or creating them, journalists may also be able to help contribute to the improvement of the news ecosystem as a whole.

## What Is Schema Markup?

A schema “markup” is the language used by a search engine to identify data structures within web pages. In this section, we will learn a few types of schema markups to serve as a starting point for how schemas work and how to use them.

In general, markup is a system of tags and annotations that enables computers to process text in a meaningful way.

There are many different markups and ways of encoding schema-related tags — as you read through additional resources, you may come across terms like RDFa and Microdata, which can be found in the glossary at the end of this tutorial.

As a journalist, you may be somewhat familiar with another web markup language, called HTML. Have you ever had to type `<p>` into your article to wrestle it into paragraphs, rather than a big block of text? Then you have written HTML!

---

<sup>8</sup> Nicholas Diakopoulos, “Automating the News,” (Harvard University Press, 2019).

HTML tells a computer program how a piece of web content should be displayed — should content be displayed as italics (<i>), in bold format (<b>), or arranged as a bulleted list (<ul>)?

In contrast, schema markup instructs a computer or a machine how the content should be classified. You can have more than one kind of markup on the same page.

For example, an <h1> tag can instruct a browser such as Google Chrome to present CNN’s headline as large and bold. An `opinion` tag can instruct an algorithm, such as Facebook News, to process the CNN story as a piece of “Opinion.”

One language, HTML, decides what content looks like when presented on a web page; the other markup makes it easier for tools, such as search engines, to categorize the content.

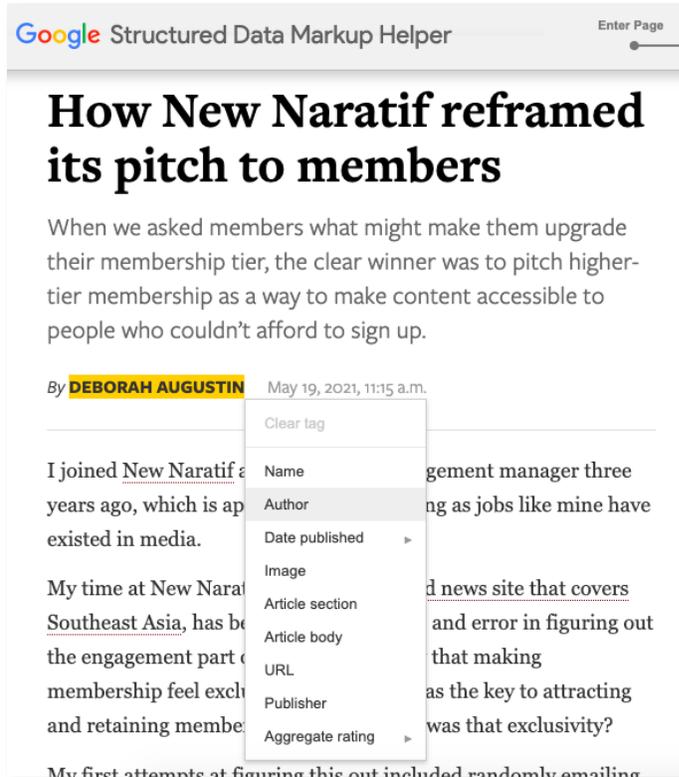
### **Do it Yourself: Schema Markup**

To get a sense of how a schema markup works, let’s write our own using Google’s [Structured Data Markup Helper](#).<sup>9</sup>

Note: You will need to create or sign in to a Google account to use this tool.

---

<sup>9</sup> Google Support, “Structured Data Markup Helper,” accessed May 5, 2021, <https://web.archive.org/web/20210505123714/https://support.google.com/webmasters/answer/3069489>.



*With the Structured Data Markup Helper, anyone can highlight sections of a web page and assign tags to them. In this example, “Deborah Augustin” should be tagged “Author.”*

To get started, navigate to Google’s [Structured Data Markup Helper](#). This Google tool allows you to highlight a section of an article and mark it up — that is, assign it tags like “author,” “date published,” and “image.”

Typically, these tags would be assigned by someone or some program writing the code of the publication’s website. But the Structured Data Markup Helper allows you to point and click on pieces of content through a friendly interface, letting you see how the markup works visually.

Next:

1. Select the radio button for “Articles”
2. Paste in a link to a news article, such as [this article](#) from Nieman Lab
3. Click “Start Tagging”

For now, let's assign these tags to the article:

- Name (headline)
- Author
- Image
- Date published

Once you have created those four tags, click “Create HTML” at the top right to pop up a bit of code.

This is markup!

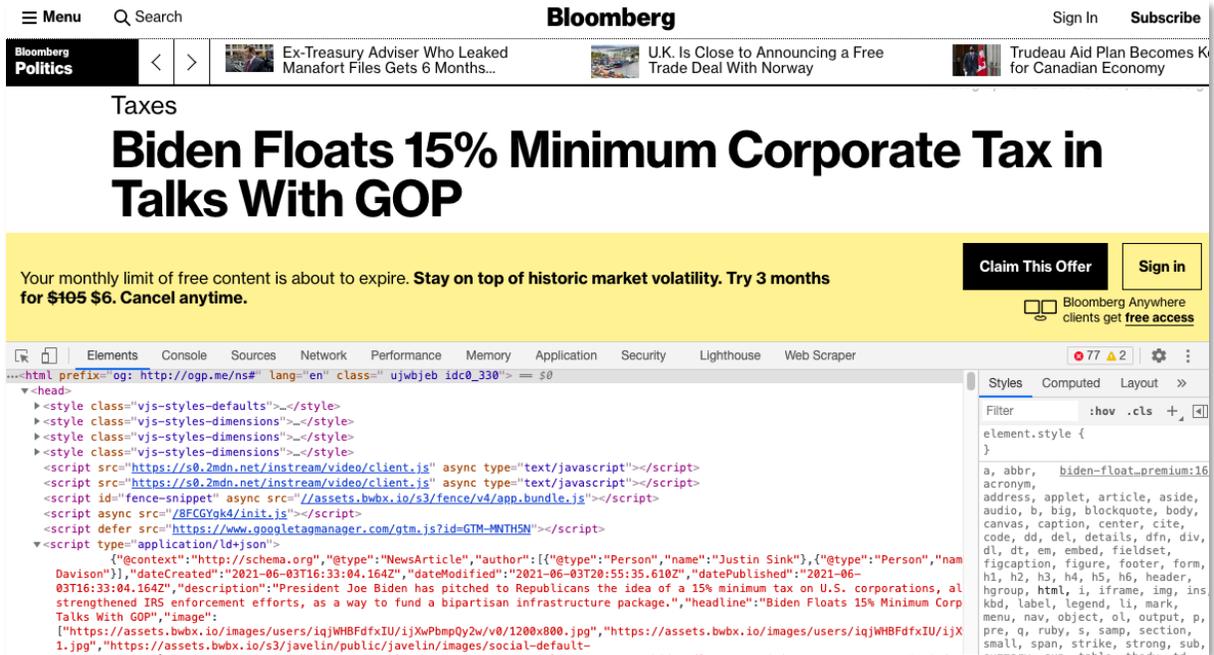
## Schema Markup in the News

Many news companies already use schema markup as part of their story distribution workflow. For example, The Associated Press and Reuters both use a metadata system called “[NewsCodes](#)” to send their stories out over the wire.<sup>10</sup> NewsCodes is organized by the International Press Telecommunications Council (IPTC), an organization that creates standards for online news media.

Another set of structures is defined at [Schema.org](#), which many news publishers, including CNN, Bloomberg, and The Dallas Morning News, use on their websites.

---

<sup>10</sup> IPTC NewsCodes Working Group, “What Are the IPTC NewsCodes?” IPTC.org, accessed May 6, 2021, <https://web.archive.org/web/20210506133004/https://iptc.org/std/NewsCodes/guidelines/>.



The red text in the code of this Bloomberg article indicates that the site uses Schema.org. By processing the red text, algorithms can interpret the information according to the tags defined at Schema.org.

Schema.org was created by Google, Yahoo, and a few other search engines in 2011.<sup>11</sup> These companies decided to put their heads together and come up with a common data structure that would help their search engines understand the vast oceans of information that flood the web.

If you want to help your newsroom’s website talk more directly to the algorithms behind Yahoo and Google News, Schema.org markup is [available and ready for use](#) on your newsroom’s website.<sup>12</sup> For even more detail about Schema.org markup for news articles, read through Schema.org’s [NewsArticle markup](#) — the tags that it uses on news article content — and their [recommendations](#) for marking up news.

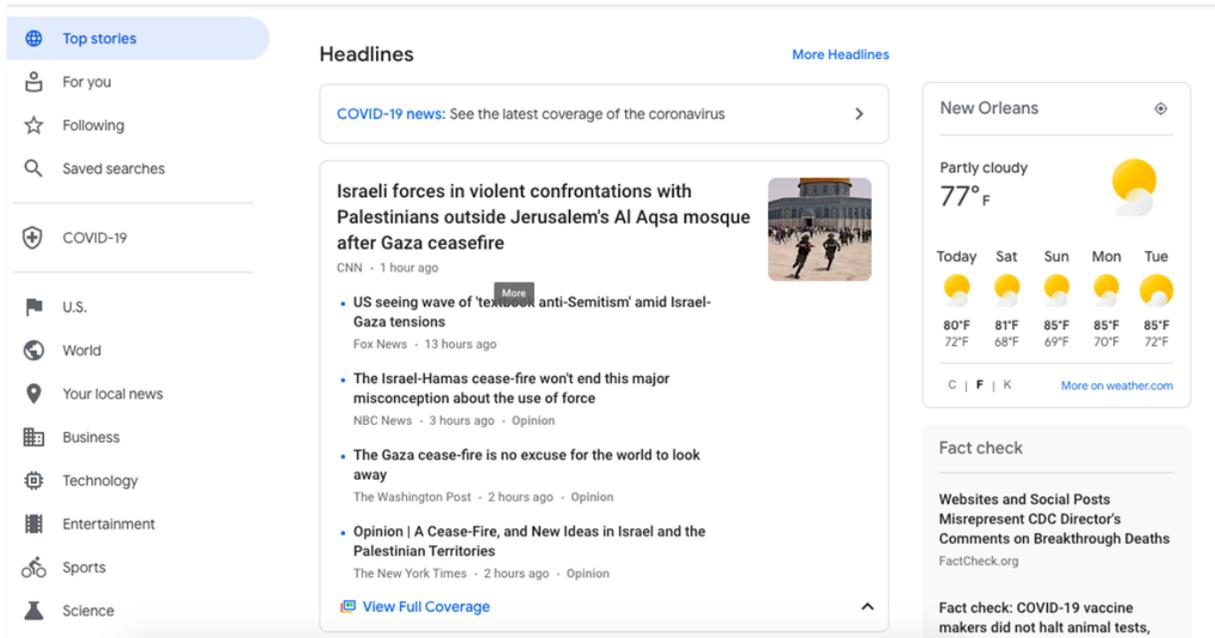
## A Markup for Fact-Checking

Some news-related organizations are also experimenting with new methodologies for helping algorithms understand and classify news

<sup>11</sup> “Schema.org,” Wikipedia, accessed on May 31, 2021, <https://web.archive.org/web/20210531u65022/https://en.wikipedia.org/wiki/Schema.org>.

<sup>12</sup> “Getting Started,” Schema.org, accessed June 16, 2021, <https://web.archive.org/web/20210616042605/https://schema.org/docs/gs.html>.

content on the web. These projects show possible paths forward in helping algorithms and computers understand complex topics, such as credibility and misinformation, that have yet to be solved in a way that works for tech companies, news outlets, and the public.



The Google News homepage includes a “Fact check” box in the lower right. This fact check box is populated directly by data that is formatted according to ClaimReview’s definitions.

[ClaimReview](#) is a schema markup designed specifically for fact-checking. Created as a collaboration by coders at Google and journalists at the Duke University Reporters’ Lab, ClaimReview is intended to act as a tagging system that publishers use to flag fact-checked articles for search engines, apps, and social media platforms, and now directly influences platforms like Google, Facebook, YouTube, and Bing.<sup>13</sup>

ClaimReview markup can be applied to an article either manually, through tools like Google’s [Fact Check Markup Tool](#), a WordPress plugin developed by [Full Fact](#), or automatically [through a newsroom’s CMS](#). If you want to look at examples of fact-checked data, you can see it on an open-source website called [Data Commons](#).<sup>14</sup>

<sup>13</sup> Lim, “A Better ClaimReview to Grow a Global Fact-check Database,” Reporters Lab, accessed April 18, 2019, <https://web.archive.org/web/20210123075453/https://reporterslab.org/a-better-claimreview-to-grow-a-global-fact-check-database/>.

<sup>14</sup> “Fact Check Tools,” Google Fact Check Markup Tool service, accessed June 21, 2021, <https://web.archive.org/web/20210501080707if /https://toolbox.google.com/factcheck/about#fcmt>.

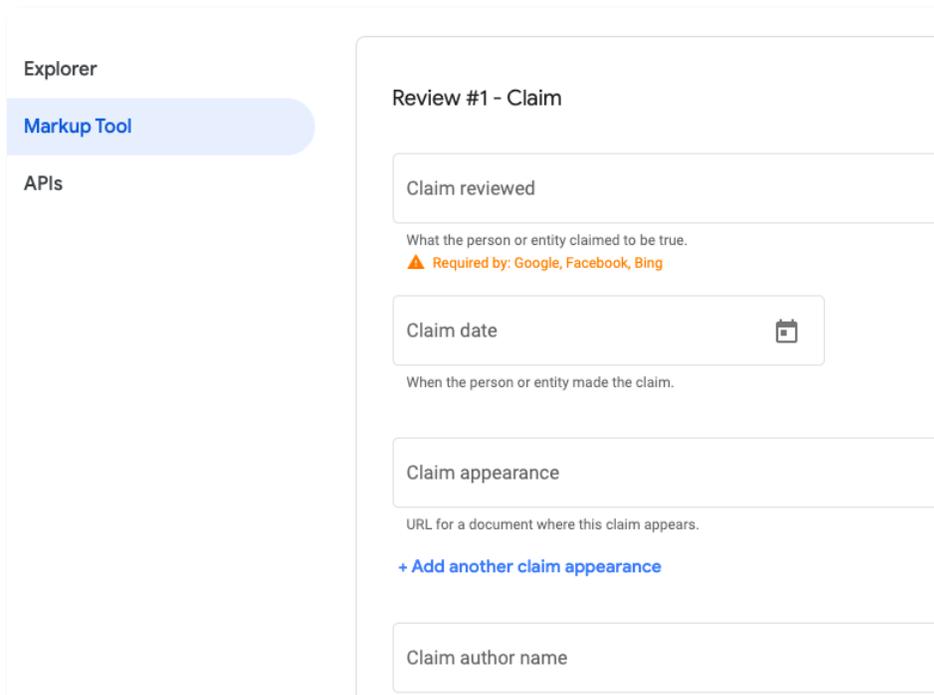
Fact-checkers around the world recently began testing a new schema, an extension of ClaimReview called MediaReview, which these organizations will use to consistently tag misinformation that's presented as images, video or audio.

ClaimReview was also being used to [develop a tool](#) to fact-check claims in real time. The tool was experimental, but it, along with the other projects mentioned in this paper, may turn out to have a huge impact on online news distribution.

### Do it Yourself: ClaimReview

ClaimReview, like Schema.org, is free and open to the public. Any publisher with a [Google Search Console](#) account can enter a fact-check manually using Google's [Fact Check Markup Tool](#). This is separate from the Structured Data Markup Tool we used earlier.

The fact-checker enters items like a summary of the claim, the author of the claim, and a verification rating.



The screenshot shows the 'Fact Check Markup Tool' interface. On the left is a sidebar with 'Explorer' (containing 'Markup Tool' and 'APIs') and 'Review #1 - Claim' in the main area. The form fields are: 'Claim reviewed' (with a note 'What the person or entity claimed to be true.' and a warning 'Required by: Google, Facebook, Bing'), 'Claim date' (with a calendar icon and a note 'When the person or entity made the claim.'), 'Claim appearance' (with a note 'URL for a document where this claim appears.' and a '+ Add another claim appearance' link), and 'Claim author name'.

*The Fact Check Markup Tool allows publishers to manually enter claims — a specific type of web content— and their organization's fact-check of it.*

If you have a Search Console account, you can open the [Markup Tool](#) and try it now. If your submission meets Google's structured data guidelines

and [fact-check eligibility guidelines](#), your fact-check may even show up on news products like Google Search and Bing.<sup>15</sup>

## What Are Signals?

Sometimes called “indicators,” “signals” are bits of information that help a computer understand, evaluate, and categorize a piece of content. The signals journalists are interested in can be thought of as the factors that algorithms use to decide about how to surface or rank a news story.

The platforms behind search engines and other related services are heavily invested in keeping these proprietary signals and processes a secret, so it can be frustrating to figure out how to optimize content for them.

The News Quality Initiative (NewsQ), a group that studies online news rankings, [compiled a summary](#) of some of the values that platforms seem to use to rank news content.<sup>16</sup>

Some signals they might be using to measure these values might include:

A factor to measure	A computational signal that could measure it
Relevance	Does the content contain the same the text as the search?
Location	Was the user’s geographic location recently tracked via apps or services?
Prominence	Does the text of the content contain terms that are currently trending?
Authoritativeness	Is this website linked to by many other websites?

It’s important to understand signals so that journalists understand how algorithms — and their designers — make their decisions. The key here is to understand “computational signals” — how does a computer understand something like relevance?

<sup>15</sup> “Fact Check documentation,” Google Developers, accessed June 16, 2021, <https://web.archive.org/web/20210616063029/https://developers.google.com/search/docs/data-types/factcheck>.

<sup>16</sup> “Panel Orientation: Platform News Products,” News Quality Initiative (NewsQ), accessed January 22, 2021, <https://web.archive.org/web/20210122192617/https://newsq.net/panel-orientation-platform-news-products/>.

We know Google Search measures relevance based partly on whether the content contains the search terms the user has entered. For example: a reader opens up their phone and types “Atlanta Braves game.” If a piece of news content contains the words “Atlanta,” “Braves,” and “scores,” it is likely to get points for relevancy.

There are many other factors at play — the day the article was published, how many other sites link to it, and so on. But the key for journalists is to turn journalistic principles into things that computers can understand. A computer, after all, doesn’t understand sports. But it can tell whether a piece of text contains the same words as a search entry. This is an example of a “signal.”

As Nick Diakopoulos writes in [Automating the News](#), “In a way, people need to develop some form of ‘empathy’ for algorithms so they can see a problem through computational eyes.”<sup>17</sup>

## New Signals and Schemas for News

As with schema markup, there are a variety of emerging projects to develop better computational signals for news. One such initiative is JournalList’s effort to create [trust.txt files](#) for newsrooms — basically, simple web pages that would list specific signals of association for platforms, advertisers, and others to interpret. Primarily, the trust.txt files list membership in associations, along with ownership of brand-specific social channels. With wide adoption, there will be a network effect of machine-readable indicators of trust.<sup>18</sup>

A community group at the W3C, a global organization that tries to standardize the web, posted an [early draft](#) of what could be used as signals to measure credibility.<sup>19</sup>

One other new project is The Journalism Trust Initiative (founded by Reporters Without Borders), which [published](#) a set of potential [standards](#) through the European Committee for Standardization. The standards now have a robust online tool that allows publishers to see how well they

---

<sup>17</sup> Diakopoulos, “Automating the News,” page 249.

<sup>18</sup> “Specification for Trust.txt File and Underlying System,” JournalList.net, accessed January 22, 2021, <https://web.archive.org/web/20210122053352/https://journalist.net/wp-content/uploads/2020/12/trust.txt-specification-v-1-1.pdf>.

<sup>19</sup> Sandro Hawke, editor, “Technological Approaches to Improving Credibility Assessment on the Web,” Credible Web Community Group, accessed March 8, 2021, <https://web.archive.org/web/20210308083048/https://credweb.org/report/20181011>.

measure up to the standard, and then publish the results in a uniform way. As well, if publishers elect to do so, they can receive a third-party certification of those results. Those uniform results and the certification may also serve as signals to the platforms and others.

All of these projects are in the early stages, but may eventually have an enormous impact on how news is discovered online.

If you aren't ready to implement code-intensive or site-wide changes, you can try to adapt your content to trigger certain signals, or, as we will see later in this guide, labels. To get a better understanding of which signals we know platforms use, and how best to use that information, check out the Online News-Schema Series' [compilation of existing research](#).<sup>20</sup>

For a compilation of some of the existing potential computational signals for news, see a [working directory](#) maintained by NewsQ.<sup>21</sup>

## What Are Labels?

If signals still seem too daunting, you can also consider implementing consistent labels. Labels are similar to computational signals, except they tend to live in the body of an article, rather than in the code. Ideally, there should be metadata to accompany the text in the article.

One example would be including an "Opinion" label in the headline of every opinion article your news organization publishes, which then may appear in the article's URL. This doesn't always speak to algorithms in the same manner that a schema markup does, but it has the added benefit of helping your readers parse out the information, too.

Trusting News, a project by the Reynolds Journalism Institute and the American Press Institute, published a short, helpful [guide](#) to help local newsrooms and on-the-ground reporters label their content.<sup>22</sup>

The Trust Project, a consortium of news organizations, has also developed [Type of Work Labels](#) that platforms and news outlets can use to identify

---

<sup>20</sup> "Training Agenda," Schema-Online Local News discussion series, accessed June 21, 2021, [https://web.archive.org/web/20210617195324if\\_/https://docs.google.com/document/d/18WEhW5EBrs4P\\_PaFkg4yW3evnKgUDo65poVuSuWzMAo/edit#heading=h.izyruo2ag236](https://web.archive.org/web/20210617195324if_/https://docs.google.com/document/d/18WEhW5EBrs4P_PaFkg4yW3evnKgUDo65poVuSuWzMAo/edit#heading=h.izyruo2ag236).

<sup>21</sup> "NewsQ Signal Database," NewsQ Initiative, accessed June 21, 2021, <https://web.archive.org/web/20210122181035/https://newsq.net/newsq-signals/>.

<sup>22</sup> "Strategy: Label Your Stories," TrustingNews.org, accessed June 17, 2021, <https://archive.is/5N3we>.

their content.<sup>23</sup> Some of these labels have been incorporated into Schema.org’s NewsArticle structure. The “Local Reporting” label, for example, signals that a story has “local roots, origin or expertise.”<sup>24</sup>

Labels are something that require little technical expertise, if any, although they would be most effective if implemented at the site-wide level, such as making it a formal newsroom policy.

## Questions?

As we’ve learned, all sorts of web services including Facebook, Twitter, and Google News use algorithms to help find and rank their content. The more journalists and tech platforms learn about how their counterparts approach online news, the closer both fields will get to their goals of sharing accurate, relevant information with the public.

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

[schema@credibilitycoalition.org](mailto:schema@credibilitycoalition.org)

## Acknowledgements

We would like to acknowledge the assistance and insights of Mark Stencel, Joel Luther, Beck Levy, Scott Yates, and Momen Bhuiyan.

---

<sup>23</sup> Type of Work Labels Working Group, “Indicator: Type of Work Labels,” The Trust Project, accessed June 21, 2021, [https://web.archive.org/web/20210506025105if\\_/https://docs.google.com/document/d/1pLUM-JvGD15NMUNMi2Hl74MnsoBGw\\_B3cYIAQ8NyO\\_M/edit](https://web.archive.org/web/20210506025105if_/https://docs.google.com/document/d/1pLUM-JvGD15NMUNMi2Hl74MnsoBGw_B3cYIAQ8NyO_M/edit).

<sup>24</sup> Type of Work Labels Working Group, “Indicator: Local Reporting,” accessed June 21, 2021, [https://web.archive.org/web/20190718082517if\\_/https://docs.google.com/document/d/1elu7Q94Qf9amlhVDDmVQgGf96iauSGSfXINA1JeW5H4/edit](https://web.archive.org/web/20190718082517if_/https://docs.google.com/document/d/1elu7Q94Qf9amlhVDDmVQgGf96iauSGSfXINA1JeW5H4/edit).

## Glossary

Structured Data	Data that has been put into a structure, such as a schema, in order for computers to understand it more easily.
Schema	A data structure that organizes information in a specific way.
Schema.org	A nonprofit dedicated to overseeing and maintaining specific schemas, along with suggested markup created by several major search engines.
Schema Markup	A language of tags, related to a data structure, to attach to text and content. Different companies and groups will often use the same markup.
Signal	A bit of information used by an algorithm to make a decision. For example, the publication date of an article can indicate its recency.
Indicator	Generally, another word for “signal.” For example, whether a publication has won a Pulitzer Prize could be used as a “credibility indicator.”
Rich Snippet	A box that highlights structured data in a search result, such as an article’s headline and image.
Enhanced Search Result	A presentation of a search result that pulls from structured data within that result. A “rich snippet” is an example of an enhanced search result.
Metadata	Data that describes other data. For example, a news article is a piece of data; its publication date is metadata.
Microdata	A set of tags introduced in HTML5, the most recent version of HTML.
RDFa	A markup recommended as a standard by the W3C coalition for embedding metadata.
Ontology	In this case, another colloquialism for a set of vocabularies or a schema.
Vocabulary	An element of a schema with a clearly defined meaning.

Also see Northwestern University’s [News Algorithms Glossary](#), developed by Nicholas Diakopoulos.

## Additional Resources

For more resources, check out the series' compilations of existing tools and research:

- Online News Series [Training Agenda](#)
- Online News Series [Training Slides](#)

The following sources were compiled to provide 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 led the compilation of 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) 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 for Digital Journalism 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 geo-location**:

“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)