

Discipline and promote: Building infrastructure and managing algorithms
in a ‘structured journalism’ project by professional fact-checking groups

Introduction

In May of 2016 the Duke Reporters Lab, an academic and professional hub for ‘structured journalism’ projects, unveiled an unusual journalistic device: the Share the Facts widget, a tool designed to help fact-checking organizations around the world increase the spread and impact of their work. The term ‘widget’ reflected the hard-to-describe nature of this new object, which is at once a visible badge attached to individual fact checks that makes it easy to share them online, and a bit of invisible computer code implementing a new data standard designed to harmonize the work of different fact-checking outlets. That standard, called ClaimReview, was the basis for a major achievement announced in 2017, when Google began extracting fact-checkers’ verdicts as ‘snippets’ featured prominently in search results. In 2018, Facebook embraced the same standard in an effort to boost the effectiveness of its own much-criticized campaign to stem online misinformation.

This paper develops the Share the Facts (STF) project as a case study in what has become known as the ‘structured journalism’ movement, a series of attempts to re-conceptualize aspects of reporting for the digital age — and specifically to make news stories legible as data for analysis or transformation by both humans and machines. The STF project is interesting and important because it shows fact-checkers themselves driving a campaign to increase the algorithmic relevance of their new form of public-affairs reporting; like other structured

1
2
3
4 journalism projects, it demonstrates that journalists (and presumably other media workers) in
5
6 some cases draw on their own professional agency to intervene in and shape algorithmic
7
8 regimes. We chronicle the development of the STF widget and the underlying ClaimReview
9
10 standard, show how the widget ties together different communities of practitioners and
11
12 sociotechnical systems, and highlight tensions that have emerged as this infrastructural
13
14 technology promotes particular institutional arrangements in the fact-checking world.
15
16

17
18 To understand the how the STF widget mediates between the developing world of
19
20 professional fact-checkers and the wider digital media environment we highlight two facets of
21
22 the technology: a *promotional/distributional* aspect and a *regulative/disciplining* one. From the
23
24 perspective of the fact-checkers promoting it, the widget succeeds by giving their work greater
25
26 purchase in algorithmically governed media environments. But it accomplishes this partly by
27
28 constituting that work in restrictive ways — by cementing specific views of how fact checks
29
30 are made, what they should look like, and who can legitimately produce them. More than that,
31
32 its adoption reinforces particular lines of authority and status among fact-checkers, helping to
33
34 establish what in hindsight may look like a professionalized, norm-governed institutional order
35
36 within this emerging international subfield of professional journalism.
37
38
39
40

41 In analyzing the STF project as an emblematic case of contemporary newswork, this
42
43 paper explores a deceptively simple question: How can journalists shape the trajectory of the
44
45 facts they produce in digital environments where gatekeeping authority has been ceded to
46
47 powerful new intermediaries? How do these efforts in turn shape or constrain journalistic
48
49 practice and epistemology? These questions matter not only to professional fact-checkers but to
50
51 journalists more broadly as they embrace the data-centric and informational understanding of
52
53 their own work reflected in structured journalism.
54
55
56
57
58
59
60

Newsroom technology and journalistic practice

Like other newsroom technologies, the STF widget embodies a set of understandings — overlapping but not always in perfect alignment — about the proper mission, conduct, and audience of journalistic work. For fact-checkers, the immediate purpose of the widget is to promote their work by creating a standardized, easy-to-understand summary of the fact-check which invites social sharing and, crucially, makes the underlying structure of fact-checks friendlier to search engines and other algorithmic mediators. (See Fig. 1.) In doing this, though, the technology helps to reinforce a particular view of the relationships between various actors involved in making, distributing, and using news, including the public. How news organizations have institutionalized technologies from the telegraph to the content management system (CMS) is the result of complex and contingent negotiations among those actors but may also shift lines of authority between them, invite changes in practice, and, in hindsight, seem to ratify particular visions of journalism (Anderson and Kreiss, 2013; Boczkowski, 2005).

[Insert Figure 1 about here]

These relationships emerge unusually clearly in the case of the STF widget. This is not a long-embedded newsroom artifact with layers of understanding and practice hardened around it, but an experimental tool designed to promote a new genre of journalism. It was created in a deliberate campaign to enroll different actors necessary to establish a data standard for fact checks, premised on the notion that the technology will be seen differently by key constituencies — notably fact-checkers, search engines, and online audiences. One way to understand the widget is as an engineered version of a ‘boundary object’ linking actors with different views in a

common project (Star and Griesemer, 1989). Its deliberate, designed flexibility differs somewhat from the ‘organic infrastructures’ (Star, 2010: 602) usually associated with that concept; but, precisely to the extent it becomes entrenched as a layer of digital infrastructure, the STF widget meets the core criterion of enabling cooperation without consensus.

In this way, this paper offers a view of the early stages of ‘infrastructural work’ (Bowker and Star, 1999). A turn toward Science and Technology Studies (STS) in communications research has drawn attention to the role of material infrastructures of media production in constituting journalistic autonomy, authority, and epistemology (e.g. Anderson, 2018; Annany, 2018; Braun, 2015; Carlson, 2017). The STF project offers a good example of what Annany (2018: 4) calls the ‘networked press’, including ‘journalists, software engineers, algorithms, relational databases, social media platforms, and quantified audiences’. As he notes, it is in specific, patterned relationships among these varied human and non-human actors — in a ‘system of separations and dependencies’ — that taken-for-granted formations such as the institutional press and the democratic public actually take shape and have meaning. The STF widget is precisely an attempt by subfield of journalists to devise a sociotechnical lever to increase the audience for and authority of their own work, while also preserving a degree of control over that work. (In this sense, the STF widget and ClaimReview aim to strengthen fact checks as ‘immutable mobiles’; Latour, 1987, 2005.)

The case we analyze also may add something new to debates over the power of algorithms. As Kitchin (2016) observes, the role of algorithms in mediating ever-wider regions of social and economic life demands critical attention but also raises difficult empirical and conceptual questions about how to study a kind of influence that is pervasive but opaque, and difficult to isolate from the existing systems and institutions algorithms are embedded in. By

now broad alarms about algorithmic power (e.g. Pariser, 2011; Pasquale, 2015) have been supplemented with nuanced organizational or meso-level accounts of how algorithms are deployed in particular occupational spheres, including journalism (Christin, 2017; Gillespie, 2018; Petre, 2015, 2018; Zamith, 2018). These accounts unpack the human work implicated in algorithmic power, show how professional practices and discourse adapt to it, and sometimes highlight strategies of resistance by individuals and organizations.

The case study developed here illustrates something slightly different: journalists and technologists taking advantage of algorithmic systems as active agents in order to advance their own organizational and institutional goals (see also Bucher 2018). This is not to say the STF initiative subverts the power of either algorithms or platform companies; as discussed below, Google and Facebook have clear strategic reasons to partner with fact-checkers amid rising public and regulatory concern about online misinformation. More broadly, this project, and structured journalism in general, can fairly be described as helping to make journalism ‘algorithm ready’ (Gillespie, 2014).¹ At the same time, the case shows journalists actively shaping algorithmic logic. And crucially, as developed below, the disciplining influence of this new technology on fact-checkers reflects independent professional-institutional agendas as much as the demands of algorithmic gatekeepers like search engines and social networks.

The next section of the paper discusses the methods and data used to develop this case study. The third section turns to an overview of the ‘structured journalism’ movement, which is intertwined with the early history of fact-checking and deeply informs the STF project. The fourth section describes genesis and development of the STF widget and the underlying ClaimReview standard, highlighting the role and understandings of key actors as well as tensions

that surfaced as that standard became increasingly established. The final section analyzes STF and ClaimReview as institution-building technologies with distributive and disciplining facets.

Methods

The research here combines ethnographic fieldwork by two authors on two related strands in contemporary journalism, the fact-checking movement and structured journalism. In conversations in 2015, the authors identified efforts to develop a new data standard for fact-checks — what would become the Share the Facts project — as a case that spanned those two professional discourses and could offer a revealing view of how journalistic practice is changing in a digital environment. The authors tried to develop the case study in an iterative and reflexive way as the STF initiative unfolded in 2016 and 2017, and successive versions of the analysis have shifted substantially; what began as a focused investigation of a newsroom technology for producing public facts became an account of institution-building and professional gatekeeping in an emerging subfield of journalism.

The argument developed here draws on three primary sources. First, interviews with key figures involved in STF were used to reconstruct the origins and history of the project, to explore how different constituencies understood it, and to monitor ongoing developments as the STF widget and the ClaimReview standard became widely adopted. In all, ten focused interviews with six journalists and technologists closely involved in the project were conducted by one or both authors between late 2015 and the end of 2017. In addition, both authors observed a live webinar in August, 2017, designed to promote adoption of the new standard.²

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The second source of data is fieldwork by the first author on the global fact-checking movement. This includes observation at four annual conferences of fact-checkers from around the world, between 2014 and 2018, as well discussions of the initiative on the global mailing list of the International Fact-Checking Network (IFCN). The four two-day fact-checking conferences took place in London in June 2014 and July 2015, Buenos Aires in June 2016, and Rome in June 2018. These forums offered the chance to see how promoters framed the STF project internally within the fact-checking community at different stages, beginning with an off-the-record session at the 2015 conference a year before STF went live and concluding with Facebook’s embrace of ClaimReview at the 2018 meeting. They also highlighted perhaps inevitable tensions as this new standard has become increasingly important in mediating relationships within the fact-checking community and with outside actors like platform companies.

The third important source of data is fieldwork by the second author on the structured journalism movement. This centered on three months in the summer of 2015 researching Structured Stories, an experimental structured journalism project in New York City run by the Duke Reporters Lab at Duke University, which also houses the Share the Facts program. This research included full ethnographic access to the small team running the experiment: observing the three days of Structured Stories training in early June, hanging out in the Structured Stories ‘newsroom’ to watch the production process at work, and attending daily editorial meetings in which student participants and editors discussed philosophical issues that arise in structured journalism projects. This wider view of the structured journalism mindset, with a history that dates to 2006, helped to ground the emphasis developed below on journalistic agency in algorithmically mediated environments.

Structured Journalism

The Share the Facts widget exemplifies a form of newswork known as *structured journalism*. The simplest way to think about structured journalism is that it not only uses data to generate news stories, but also seeks to turn events in the world, and the stories we tell about those events, into structured data — that is, data organized in a fashion to be machine-readable — which can then be fed back into other stories. This idea is only a decade old and remains an outlier in professional practice, not nearly as established as data journalism, for example. Only a handful of structured journalism projects dot the media landscape, in hubs like the Duke Reporters Lab and the BBC Research and Development lab, and some initiatives have already closed down for lack of funding. At the same time, structured journalism arguably represents the vanguard of computational practice in the newsroom, proposing to fundamentally redefine newswork in a way that marries journalistic and computational thinking.

One of the earliest discussions of structured journalism can be found in a seminal piece by Adrian Holovaty, the developer of an influential 2005 project known as the Chicago Crime Map. Developed as an alternative to lurid, episodic newspaper narratives about individual incidents of crime, the Chicago Crime Map gave a more comprehensive picture by mapping incidents across the city using databases of structured crime data. Holovaty's (2006) piece, 'A Fundamental Way Newspaper Sites Need to Change', didn't just argue that journalists should make better use of data in news stories, as advocates of Computer Assisted Reporting and data journalism had previously (Anderson 2018, Coddington 2015). Rather, it critiqued the very idea of the narrative story as the primary journalistic form. 'Newspapers need to stop the story-centric worldview,' he wrote. 'The problem here is that, for many types of news and information,

newspaper stories don't cut it anymore.' At this point, Holovaty used the term structured information, a term commonly deployed in computer science, but which had not been used in conversations about journalism before this time:

So much of what local journalists collect day-to-day is structured information: the type of information that can be sliced-and-diced, in an automated fashion, by computers. Yet the information gets distilled into a big blob of text — a newspaper story — that has no chance of being repurposed. ... Repurposing and aggregating information is a different story [from simply reformatting it], and it requires the information to be stored atomically — and in machine-readable format.

For example, say a newspaper has written a story about a local fire. Being able to read that story on a cell phone is fine and dandy. Hooray, technology! But what I *really* want to be able to do is explore the raw facts of that story, one by one, with layers of attribution, and an infrastructure for comparing the details of the fire — date, time, place, victims, fire station number, distance from fire department, names and years experience of firemen on the scene, time it took for firemen to arrive — with the details of previous fires. And subsequent fires, whenever they happen.

These intellectual principles were observed in practice during the summer of 2015 when the second author conducted fieldwork at Structured Stories NYC, an experimental initiative of the Duke Reporter's Lab. Based in New York City, the project employed three volunteer journalists to report breaking news on city housing issues and crime in a novel way, focused on the semantic elements — people, organizations, events — that make up the news. In practice the Structured Stories journalists spent most of their time gathering basic units of information, particularly the *nouns* and *verbs* that made up a news story, trying to relate these recurring elements in a systematic way. (This often meant going through their story notebooks with a

range of colored highlighter pens to classify information for the database.) The dominant daily intellectual exercise, in other words, was to choose the standards by which new information could be sorted, classified, and combined into a consistent database of journalistic happenings in New York City (author's fieldnotes, July 9, 2015).

As we will see, the STF widget also relies on structured journalism principles, applied on a wider scale and with somewhat higher stakes. To devise standards to define and organize fact-checking by a wide variety of news organizations, working in different countries and in some cases with different missions and methods, while also meeting the needs of platform companies, becomes even more obviously a kind of political work. At the same time, the entire genre of political fact-checking as it developed since the early 2000s aligns with the ethos of structured journalism; both discourses are meaningfully native to journalism, and show journalists adapting to digital media in ways that embody core professional concerns, even as they also make reporting work more 'algorithm-ready' (Gillespie, 2014) by turning stories into data.

Fact-Checking as Structured Database Work

Unlike traditional, internal fact-checking by journalists, what is sometimes called political or external fact-checking focuses on debunking false statements by politicians and other public figures. External fact-checking emerged as a distinct genre in the US in the early 2000s (though it has earlier roots) and grew into a professional reform movement led by veteran political journalists advancing a forceful critique of conventional 'he said, she said' reporting (author cite). As the movement spread internationally to more than 50 countries over the last

decade, it has widened to include not just news organizations but academic and civil society groups of many stripes (author cite).

Fact-checking invites a structured approach to reporting work. In contrast to conventional news narratives, fact checks often build to a single data point, the verdict. They also feature a consistent set of elements: a claim, an analysis, a verdict, a list of sources, and so on. This recurring structure was evident even in the earliest online fact-checking efforts, such as Snopes.com and FactCheck.org. The first to make it explicit was PolitiFact, launched in 2007: The site turned each recurring element into a distinct field in a database that archives all of the statements checked by its trademarked ‘Truth-O-Meter.’ Reducing reporting work to uniform classes of data in this way allows it to be aggregated as the basis for ‘higher-level’ analysis — for instance, comparing the records of different candidates or parties or revealing patterns in political discourse (e.g. Drobnic-Holan 2016). The developer of the original PolitiFact site, linking directly to Holovaty’s manifesto, described the project this way shortly after launch:

The site is a simple, old newspaper concept that’s been fundamentally redesigned for the web. We’ve taken the political ‘truth squad’ story, where a reporter takes a campaign commercial or a stump speech, fact checks it and writes a story. We’ve taken that concept, blown it apart into it’s [sic] fundamental pieces, and reassembled it into a data-driven website covering the 2008 presidential election. (Waite, 2007)

As noted, rather than database-driven journalism, this is journalism as a database. Every PolitiFact article exists as both a reported narrative and a defined set of data points, highlighted in a rectangular badge at the top of every story: the name and photograph of the speaker, the statement being checked, the date, and the verdict. Like the STF widget which they are a model for, PolitiFact's badges can be inserted into a new article as a visual reference to previous work. Organizing PolitiFact around a database has shaped the group's news values, editorial strategy,

and business model (see author cite). For instance, writing for a database means timeliness and exclusivity are deemphasized in favor of an ethic of record-keeping. Similarly, this structured approach became the basis for PolitiFact's formal methodology and for its affiliate program, which has allowed PolitiFact to expand through media partnerships.

Origins and Development of the STF Widget

The Share the Facts widget is computer code designed to give fact-checks greater purchase in online discourse by making them easier to find, easier to understand, and easier to share. It accomplishes this in two ways. First, the code creates a visible capsule (like the Truth-O-Meter badge described above) which identifies the claim being checked, the person who made it, and the verdict. This capsule can be shared on social media and embedded in news articles; the project aims to make the widget a familiar discursive object, like the embedded tweet. Second, and more important, the widget injects data tags into the fact check which allow it to be parsed by search engines and other algorithmic mediators. That tagging scheme, called ClaimReview, is an open standard developed mainly by Google engineers in consultation with fact-checkers.

A useful starting point is to review the key actors involved in establishing the STF widget and ClaimReview. Even a brief list has to include several organizational actors, each in turn comprising or standing in for any number of elements at least tacitly enrolled in the project:

- The Duke Reporters Lab led by Bill Adair (and supported by a wider network of human and technical resources at the University) as the organizational home of

the STF program, and as a representative of the fact-checking community in conversations with Google about the ClaimReview standard;

- The Washington Post's Fact Checker, PolitiFact, and FactCheck.org (and their respective parent organizations) as leading US fact-checkers committed in advance to implementing the widget;
- The International Fact-Checking Network (IFCN), led by Alexios Mantzarlis and based at the Poynter Institute, as the primary institutional hub for discussing the initiative (via its reporting, conferences, and mailing list) and as the base for parallel efforts to professionalize fact-checking, discussed below;
- Google, as an important source of financial and engineering support for the project (both directly and via the standards body Schema.org), and crucially for the promise to take advantage of the new data layer by integrating fact checks into search;
- Other platform companies, like Facebook and Bing, whose adoption has ratified ClaimReview as an accepted standard.

Of course, the fate of the widget is tied to other constituencies — most vitally to fact-checkers and their audiences. The widget can't function as a traffic-driving engine if audiences don't play their part by sharing and embedding and clicking on it, and by flowing to the standards-compliant outlets whose work is featured on Google. And the coherence of the widget depends on the degree to which the underlying data standard is adopted by a diverse global network of fact-checkers. In the most ambitious view, adoption by practitioners and by various platforms — search engines, social networks, consumer-electronics devices — will be mutually

reinforcing and ClaimReview will by degrees become a basic infrastructural standard in the data environment, allowing fact checks to circulate in entirely new ways and contexts.

The STF widget was conceived as a vehicle to encourage adoption of ClaimReview. In web development, the term ‘widget’ has come to mean a small application embedded in a web page (typically via a CMS) to perform a specific function, such as displaying recent tweets.³ Many such widgets have dual facets, both displaying a new visual element and sharing the code that allows them to be copied to other web sites. However, to journalists in this case the term also captured the fact that the technology is illegible, a black box whose workings they don't understand — a point made repeatedly in interviews and in discussions at fact-checking conferences. ‘I don’t understand enough about how all this works’, the head of FactCheck.org clarified at the outset of an interview as the site was implementing the widget:

[But] the point is to get our material in front of more people. That’s what we’re trying to do. So the way it’s been presented to me is by doing this, it has two advantages: one, and most important really, is that it would allow people who go to Google and search for information on a particular candidate or statement or issue, that our material would be pushed up, you know, and get a better — would come up higher in the results. That, to me, was very attractive. And also, what’s very attractive too is this idea of the widget, that we can insert it not only in our stories but throughout social media. (E Kiely 2016, personal communication, 5 February)

A brief history of the development of the STF widget will help to illustrate the multiple facets of this distribution technology, setting the stage for discussion of the different, sometimes competing agendas it embodies.⁴ The initial impetus for the project came from a prominent US fact-checker, the Washington Post's Glenn Kessler. Attending an overseas conference on democracy and technology in early 2015, Kessler ran into a Google executive whom he knew

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

from an earlier period in both of their careers, when Kessler covered foreign policy. He took the opportunity to suggest that professionally produced fact checks should be highlighted authoritatively in search results as a way to combat misinformation, which would be especially useful in countries ‘where access to accurate information is particularly difficult’:

I had had this idea in my head that, wouldn’t it be great if Google were able to, when it’s searching the web for news and information, if it would actually say — you know, given the fact that there’s so much bad information out there — that it would actually elevate fact-checks in the search results. If people are trying to find out something about something that a politician said, you know, Google would say, ‘Ah, this is actually a legitimate, vetted organization that has a good track record of identifying true facts, and so therefore we’re going to put this thing higher up on the page.’ (G Kessler 2016, personal communication, 8 April)

The conversation led to an initial meeting in Washington, DC, after which Kessler enlisted another prominent fact-checker, PolitiFact founder Bill Adair, to carry the project forward. Adair was in a unique position to develop the idea: in 2013 he had left PolitiFact to join the faculty of Duke University and become head of the Duke Reporters' Lab, which he repositioned as a hub for fact-checking and structured journalism, launching experiments such as Structured Stories NYC. In that position Adair also helped to organize the first global meetings of fact-checkers and to establish the International Fact-Checking Network, now based at the Poynter Institute, a journalism education and training nonprofit. Over the course of 2015 and 2016 Adair continued conversations with Jigsaw, a ‘technology incubator’ owned by Google's parent company, and with Schema.org, a public repository for structured data standards founded by Google, Microsoft, and Yahoo. As Adair explained,

1
2
3 Schema is a partnership ... that comes up with the standards for things on the web, the
4 schema in which things are displayed. And the whole idea of it is to have consistency in
5 how things are displayed, so that you have structure. ... They are all very much believers
6 in structured journalism stuff. (B Adair 2016, personal communication, 14 January)
7
8
9

10
11
12 The effort to establish a new data standard for fact-checking underscored how structured
13 journalism, a discourse that bridges editorial and programming communities, provided a
14 language for journalists to address organizational and technical concerns that are common in the
15 world of engineering standards. The ideal approach to deploying the standard would offer clear
16 benefits to drive adoption; it would be lightweight enough to implement easily on different
17 publishing platforms; and, Adair stressed, it would not demand ongoing supervision from a
18 central body like the Reporters Lab. Perhaps most important, the solution had to strike a balance
19 between highlighting fact-checkers' work in an authoritative way while still creating an incentive
20 to click through to their sites, a theme that came up in interviews as well as discussion at fact-
21 checking conferences (author's fieldwork, July 2015 and June 2016). 'I want to make sure I
22 don't create something that ends up screwing the fact-checkers,' Adair noted during
23 development (B Adair 2016, personal communication, 14 January).
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 The idea for the widget itself came from Justin Kosslyn, a Jigsaw engineer who had
41 worked on Google News and was active in developing 'structured data' standards. (Previously he
42 helped to develop a standard for human-rights NGOs to mark up their cases in machine-readable
43 ways.) Modeled on the embedded tweet, the visual facet of the widget adds what Kosslyn calls
44 'network-independent value' — a reason for publishers to add the metadata to their articles even
45 before the new data standard benefited from network effects. As he explained, 'the goal was for
46 the machine-readable aspect to be useful, but I knew that would take a while. I knew it would
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

have to have some intrinsic value to publishers’ (Kosslyn 2017, personal communication, 18 May). Publishers can choose to adopt the underlying ClaimReview schema on their own, but the widget provides an easier path while offering additional benefits. He also suggested an easy way to generate widgets: a customizable form maintained by the Duke Reporters Lab, called the ‘widget generator’, which asks for the key details of a fact check and automatically produces a block of HTML code to paste into the article before publishing it.

This modular approach to creating widgets requires an additional step from the reporter but resolves some of the tensions, both technical and journalistic, which tighter integration might produce. First, it avoids the difficult work of customizing each publisher’s CMS to support the new tagging scheme, by allowing the scripts to act as a kind of mediator. Each participating organization uses its own version of the script; new outlets can be added (or changes to existing ones accommodated) by tweaking an individual script rather than altering a central code base. It also meant ClaimReview could develop without requiring updates to the CMS of each participant — a vital point, since the primary goal of the new standard from the outset was adoption by Google and similar platforms. Crucially, by hosting the widget-generator on a password-protected site, the Duke Reporters’ Lab maintains control over the project without a need for day-to-day management.

A rough mockup of the STF widget was introduced to fact-checkers in July 2015, almost a year before launch, during an off-the-record session at the second global fact-checking conference (author’s fieldnotes, July 2015; permission was later obtained to use this material). The presentation drew on the language of structured journalism, modeled visually and described this way: ‘Structured journalism reimagines the news story and breaks it into component parts, giving more flexibility to journalist and reader’ (see Fig. 2). The reward for embracing this

1
2
3 approach appeared on another slide: ‘Search engines love structure.’ The discussion also
4
5 acknowledged that thorny questions about who counts as a legitimate fact-checker — ‘something
6
7 that we’re not going to resolve here today’ — would inevitably become more acute as fact-
8
9 checking was integrated more tightly into search. This dovetailed with two major, linked themes
10
11 of the conference that year: encouraging fact-checkers to embrace new technologies to boost
12
13 their impact, but also to adopt professional standards to improve quality and defend against their
14
15 many critics (author’s fieldnotes, July 2015).
16
17

18
19 [Insert Figure 2 about here]
20
21

22 Development continued through the fall, and from early 2016 the widget was in testing at
23
24 three pilot sites, PolitiFact, FactCheck.org, and the Washington Post’s Fact Checker. The STF
25
26 project and web site launched publicly in May 2016. The official announcement described the
27
28 widget as a tool that ‘provides a new way for readers to share fact-check articles and spread them
29
30 virally across the Internet’. It did not mention search engine integration, which had not been
31
32 announced, but hinted at future possibilities: ‘Share the Facts boxes are fully machine readable,
33
34 enabling new ways of assembling automated collections of fact-check findings from across the
35
36 Internet’ (Adair, 2016). A presentation the next month at the 2016 global fact-checking
37
38 conference, in Buenos Aires, highlighted the promise of greater traffic from search engines but
39
40 also the hope that the widget would make embedded fact-checks a common element of news
41
42 articles, like embedded tweets (author’s fieldnotes, June 2016).
43
44
45
46

47 Meanwhile, work on the underlying ClaimReview standard at Schema.org continued
48
49 through 2016, shepherded by a structured data specialist at Google. In October the search engine
50
51 made its first public announcement relating to the project: Articles that used ClaimReview and
52
53 followed ‘commonly accepted criteria for fact checks’ would be tagged with a new ‘Fact Check’
54
55
56
57
58
59
60

label on Google News, in order to ‘help readers find fact checking in large news stories’ (Gingras, 2016). Google expanded the program incrementally until in April 2017 it took the step that fulfilled fact-checkers’ initial vision: The search engine began to preview the same key details displayed by the widget — the claim, the speaker, and the verdict — in an authoritative ‘snippet’ featured prominently in the results. (See Fig. 3.) The announcement explained that to participate, publishers must either use the widget or apply ClaimReview markup directly; in addition, only publishers ‘algorithmically determined to be an authoritative source of information will qualify’ (Kosslyn & Cong, 2017). The search engine Bing, owned by Microsoft, followed suit several months later (Schwartz, 2017).

[Insert Figure 3 about here]

New applications for ClaimReview and the STF widget have continued to emerge, in keeping with the rhetoric of progress attached to the initiative from the outset. For instance, the Duke Reporters Lab has developed a voice-enabled fact-checking app for the Amazon Echo and Google Home devices, also called Share the Facts, that automatically includes outlets using the widget (Ryan, 2017). In mid-2018, Facebook embraced ClaimReview as one of several steps to strengthen its efforts to fight online misinformation, explaining that this will save time and effort for its fact-checking partners (Funke, 2018). More broadly, the ClaimReview schema has become a basic enabling standard for a series of collaborations with artificial intelligence researchers that aim to automate parts of the fact-checking process by aggregating the work of multiple fact-checking outlets (author cite).

Discussion: Distribution and Discipline

1
2
3 The history sketched above shows a group of journalists engaging in a coordinated,
4 multi-year campaign to increase the algorithmic relevance of their own work; it also highlights
5 some of the tensions involved in that effort, which will be further developed here. That
6 campaign, drawing on the conceptual vocabulary of structured journalism and based mainly in
7 ancillary professional organizations and forums, has been remarkable successful in many ways.
8 At the 2018 fact-checking conference, in Rome, a session on ‘the future of ClaimReview’
9 reported that the initiative was generating ‘huge traffic’ for participating outlets⁵ and highlighted
10 the potential for automation: ‘the mind boggles at how we can use schema to hold people and
11 entities accountable for passing along bad information’ (author’s field notes, June 2018).
12
13
14
15
16
17
18
19
20
21
22
23

24 However, the same forum raised difficult questions that have become more consequential
25 as the technology increasingly mediates between fact-checking outlets, professional
26 organizations, and platform companies like Google and Facebook. Fact-checkers from the
27 developing world complained that the ClaimReview standard, as implemented by Google,
28 doesn’t recognize their language, or that they had adopted the schema but were still being
29 outperformed in search by state-backed sources of misinformation. Others asked why Google
30 does not privilege organizations that have signed the IFCN Code of Principles, which Facebook
31 requires from its fact-checking partners. One of the engineers who led development of
32 ClaimReview countered that such decisions should be left to practitioners: ‘It’s not for us at
33 Schema.org, it’s not for us at the tech companies to say, “This is what fact-checking looks like”’
34 (author’s field notes, June 2018).
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 The introduction to this paper identified two facets of the ‘Share the Facts’ project: a
50 *promotional/distributional* aspect and *regulative/disciplining* one. It should now be clearer how
51 the STF/ClaimReview complex plays both these roles at once. The promotional facet is more
52
53
54
55
56
57
58
59
60

obvious, highlighted constantly in explicit rhetoric around the project; at least three distinct senses stand out. First, in keeping with the public-service mission of these organizations, the widget promises to make their work more visible and impactful. Fact checks shared on social media, embedded into news reports or highlighted on Google will be seen by more people and may be more likely to have an effect on political discourse. Second, more narrowly, the widget promises to drive traffic to fact-checking sites. Higher traffic increases exposure to their brands, boosts direct revenue in the form of advertising revenue or individual donations, and constitutes vital evidence to secure funding from major charitable foundations. (As noted, however, curation by an intermediary like Google may increase visibility without yielding more traffic; hence the importance of evidence of ‘huge traffic’ gains presented at the 2018 conference.) And finally, as discussed, adopting the ClaimReview schema eases compatibility with emerging forms of distribution, along the lines of a syndication standard like RSS. This can include new hardware devices, such as the Amazon Echo or so-called ‘smart TVs’, but also social media platforms, automated fact-checking projects, and other new applications.

At the same time, however, the technology also has what might be called a regulative or even disciplinary dimension, defining the thing being promoted — fact-checking — in particular ways. First and most obviously, the STF/ClaimReview complex enforces a particular vision of fact-checking as a structured format. The affordances of the tagging scheme favor fact checks built around a single, discrete claim, as opposed to a series of related claims or a broad rhetorical front; from a specific claimant, as opposed to a general rumor; and yielding a pithy, decisive verdict. Each of these elements must be quite brief: While ClaimReview itself does not specify a character limit, particular platforms recognize different field lengths, which has resulted in fact checks being cut off to produce nonsensical ‘snippets’ on Google. The search engine’s guidelines

1
2
3 tell ClaimReview users to front-load their analysis ‘in case the sentence is truncated to fit the
4 display’; in a session at the 2018 fact-checking conference, a Google engineer urged attendees
5
6 adopt other editorial practices that would work better with the standard, such as never
7
8 paraphrasing claims and, more controversially, harmonizing their rating systems to a single
9
10 ordinal scale (author’s field notes, June 2018).
11
12
13

14
15 In this way the success of STF/ClaimReview moots what had been a long-running debate
16
17 among fact-checkers over the use of ratings systems (author cite). Several high-profile outlets,
18
19 including FactCheck.org and the U.K. site Full Fact, have vocally objected to rating schemes
20
21 such as PolitiFact’s ‘Truth-O-Meter’ as pseudoscientific and reductive. To accommodate this
22
23 diversity, ClaimReview allows organizations without fixed ratings to substitute any brief phrase
24
25 as a verdict; in practice, however, these terse evaluations are often indistinguishable from
26
27 standardized ratings. (For instance, via the STF widget, FactCheck.org now routinely stamps
28
29 labels such as ‘False’ and ‘Not the whole story’ on its fact checks.) Conforming to the standard
30
31 thus promotes choosing subjects that lead to pithy judgments, and it concedes a larger argument
32
33 about how fact checks should circulate in the world and what the public should expect from
34
35 them. The chief technologist at Full Fact, which uses ClaimReview, nevertheless argues that the
36
37 standard effaces time, language, and geography in ways that entail a dangerous lack of context
38
39 for some fact checks (M Babakar 2018, personal communication, 11 January). She echoed this
40
41 concern at the 2018 conference: ‘We’re atomizing our own content... What does that mean for
42
43 us when we don’t have control over the design of our fact-checks anymore, we don’t control how
44
45 people see it?’ (author’s field notes, June 2018).
46
47
48
49
50

51
52 Less obvious but arguably more important are the ways the new technology helps
53
54 professional fact-checkers to police the borders of their subfield, by creating another anchor for
55
56
57
58
59
60

1
2
3 boundary work (Gieryn, 1983; Carlson and Lewis, 2015). Access to the trademarked STF
4
5 widget, controlled by the Duke Reporters Lab, offers a new token of legitimacy in the field, and
6
7 thus another way (in addition to IFCN membership, conference participation, etc.) to exclude
8
9 organizations deemed illegitimate. The STF initiative has been closely aligned with internal
10
11 efforts to professionalize the field. Participating in STF does not yet require formally adopting of
12
13 the IFCN Code of Principles, introduced in 2016 and signed by 59 outlets by late 2018. But core
14
15 IFCN principles of transparency (about sources and methods) and ‘fairness and nonpartisanship’
16
17 are used to evaluate STF applicants. Adair gave the example of a newspaper columnist who only
18
19 debunks claims by Donald Trump, and was rejected; ‘When they came to us and said they
20
21 wanted to use the widget, I wrote back and said, ‘So glad you’re interested, for the widget we
22
23 want organizations that check all sides’ (B Adair 2017, personal communication, 31 August).
24
25 Adair has been a leading advocate of the IFCN code, and the three US fact-checkers involved in
26
27 developing the widget are original signatories.
28
29
30
31
32

33 Unlike the widget, the underlying ClaimReview schema is an open standard which any
34
35 site in theory can use to tag its articles. However, various platform companies only recognize
36
37 ClaimReview code from outlets that meet additional criteria. As noted, Facebook requires its
38
39 fact-checking partners to be approved signatories of the IFCN Code of Principles. (This has led
40
41 to suggestions that Facebook, under political pressure to include conservative outlets as fact-
42
43 checkers in the US, has in turn pushed the IFCN to approve their applications; e.g. Ingram 2018.)
44
45 Meanwhile, Google applies a series of tests: First, the organization producing fact checks must
46
47 qualify for inclusion in Google News, itself an opaque and controversial process (e.g. Christian,
48
49 2017). In addition, sites should ‘follow the commonly accepted criteria for fact checks’ (Gingras
50
51 2016). Developed in consultation with fact-checkers, these require fact checks to be identified as
52
53
54
55
56
57
58
59
60

such, to focus on a ‘discrete, addressable claims’, and to be ‘transparent about sources and methods’; violations can lead to delisting from Google News. Finally, Google has indicated that publishers must be ‘algorithmically determined to be an authoritative source of information’ (Kosslyn and Yu 2017). At the 2018 fact-checking conference, a Google engineer explained that the site screens algorithmically based on a wide range of signals to exclude sites — including ‘porn sites, car dealerships, dentists’ — trying to game its rankings by adopting ClaimReview. Asked why Google does not formally require the IFCN Code of Principles, the engineer replied, ‘The slippery slope is right here. ... The problem is that IFCN is an external entity, and so we [would] give a lot of power to an external entity’ (author field notes, July 2018).

That response points to a final regulatory dimension of the STF/ClaimReview complex: Further cementing the authority of particular institutional gatekeepers in the new professional field, including a handful of leading fact-checkers as well as ancillary organizations like the IFCN and the Duke Reporters Lab. Despite not formally adopting the IFCN code, Google’s involvement in the initiative already *does* give ‘a lot of power’ to the IFCN and other gatekeepers who manage the relationship with major platforms and funders, lead the way in developing technologies, and have been most active in promulgating professional standards. While the criteria for fact-checking advanced by the IFCN, the Duke Reporters Lab, and Google do not coincide perfectly, they reinforce one another, and emerged from overlapping conversations in a tightly defined milieu. As noted, Google was instrumental in developing ClaimReview and has been a major funder of the IFCN events where initiatives like Share the Facts and the Code of Principles are discussed and debated.

Through these linked distributive and regulative facets, STF and ClaimReview bind a bid for algorithmic relevance by fact-checkers to their ongoing professional project. The data-

1
2
3 producing practices normalized by this technology can obviously been seen as adaptations to a
4 digital media environment dominated by platform companies — even as vehicles for what
5
6 Caplan and boyd (2018) call ‘isomorphism through algorithms’. At the same time, they also
7
8 strongly embody an independent, norm-driven professional discourse, and enroll Google and
9
10 other platforms in building an institutional-professional order among fact-checkers like the ones
11
12 found in the wider journalistic field in many countries — an order anchored by a handful of
13
14 nonprofit professional organizations, by formal and informal ethical codes, and by a more-or-less
15
16 agreed-upon hierarchy, led by professional standard-bearers and embodied in various markers of
17
18 status, such as awards, certifications, and so on.
19
20
21
22

23
24 Finally, it is vital to note Google, Facebook, and other platform companies have a stake
25
26 in this institution-building project: These algorithmic gatekeepers increasingly benefit from to
27
28 being able to point to recognized, independent institutional authorities to relieve public and
29
30 regulatory scrutiny of their singular role in mediating global information flows (while preserving
31
32 the discretion to prioritize commercial goals in directing those flows). As platforms assume a
33
34 more central role in determining the manner by which an informed public comes into being, and
35
36 as scholarly attention increasingly turns to these sites of activity, it is important to keep the
37
38 multi-directional nature of algorithmic power in mind. The case developed here underscores the
39
40 continuing relevance, and potential influence, of the long history of professional movements —
41
42 including fact-checking and structured journalism — that have actively sought to reform or
43
44 improve journalism with the embrace of new technologies and methods.
45
46
47
48
49
50

51 ¹ Thanks to Liliana Bounegru for helping to develop this point.
52 ² The Poynter Institute’s News University hosted the webinar, called ‘Sharing the Facts: How a Hidden Code Can
53 Boost Your Fact-Checks’, on 23 August 2018; see <http://www.newsu.org/courses/fact-check-markup-code>.
54 ³ The word ‘widget’ has been in use for nearly as century as an ‘indefinite name for a gadget or mechanical
55 contrivance’ (OED, 2017). This sense is reflected in its usage in computing, since the 1980s, to refer either to a
56
57
58
59
60

small application with a dedicated function (such as displaying time or weather), or to an element of a graphical user interface (like a menu or dialogue box) and the code that enables it.

⁴ The account of the development of the widget and the ClaimReview standard draws on interviews with Bill Adair on 2 December 2015, 14 January 2016, 18 May 2017, and 31 August 2017; Mevan Babakar on 11 January 2018; Glenn Kessler on 8 April 2016; Eugene Kiely on 5 February 2016; Justin Kosslyn on 18 May 2017; and Alexios Mantzarlis on 19 February 2016.

⁵ For instance, PolitiFact reportedly saw its share of traffic from Google rise from about one-third in early 2016 to 60 percent in 2016, though some of that may be attributable to changes in Facebook's algorithm.

Works Cited

- Adair B (2015) At the Global Fact-Checking Summit, a call to look ahead. In: *Duke Reporters' Lab*, 23 July. Available at: <https://reporterslab.org/at-the-global-fact-checking-summit-a-call-to-look-ahead/> (accessed 7 August 2018).
- Adair B (2016) New Share the Facts widget helps facts – rather than falsehoods – go viral. In: *Duke Reporters' Lab*, 12 May. Available at: <https://reporterslab.org/new-share-facts-widget-helps-facts-rather-falsehoods-go-viral/> (accessed 7 August 2018).
- Ananny M (2018) *Networked Press Freedom: Creating Infrastructures for a Public Right to Hear*. Cambridge, MA: The MIT Press. Available at: <https://mitpress.mit.edu/books/networked-press-freedom> (accessed 4 April 2018).
- Anderson CW (2018) *Apostles of Certainty: Data Journalism and the Politics of Doubt*. New York, NY: Oxford University Press.
- Anderson CW and Kreiss D (2013) Black Boxes as Capacities for and Constraints on Action: Electoral Politics, Journalism, and Devices of Representation. *Qualitative Sociology* 36(4): 365–383. DOI: [10.1007/s11133-013-9258-4](https://doi.org/10.1007/s11133-013-9258-4).
- Boczkowski P (2005) *Digitizing the News: Innovation in Online Newspapers*. Cambridge, MA: MIT Press.
- Bowker GC and Star SL (1999) *Sorting things out: classification and its consequences*. Inside technology. Cambridge, Mass: MIT Press.
- Bucher T (2018) *If...Then: Algorithmic Power and Politics*. Oxford Studies in Digital Politics. Oxford, New York: Oxford University Press.
- Caplan R and boyd danah (2018) Isomorphism through algorithms: Institutional dependencies in the case of Facebook. *Big Data & Society* 5(1): 2053951718757253. DOI: [10.1177/2053951718757253](https://doi.org/10.1177/2053951718757253).
- Carlson M and Lewis SC (eds) (2015) *Boundaries of Journalism: Professionalism, Practices and Participation*. New York, NY: Routledge.
- Christian J (2017) We still don't know how Google News works. In: *The Outline*, 22 November. Available at: <https://theoutline.com/post/2512/we-still-don-t-know-how-google-news-works> (accessed 5 April 2018).
- Christin A (2017) Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society* 4(2): 2053951717718855. DOI: [10.1177/2053951717718855](https://doi.org/10.1177/2053951717718855).

Coddington M (2015) Clarifying Journalism’s Quantitative Turn. *Digital Journalism* 3(3): 331–348. DOI: [10.1080/21670811.2014.976400](https://doi.org/10.1080/21670811.2014.976400).

Funke D (2018) In Rome, Facebook announces new strategies to combat misinformation. In: *Poynter*, 21 June. Available at: <https://www.poynter.org/news/rome-facebook-announces-new-strategies-combat-misinformation> (accessed 7 August 2018).

Gieryn TF (1983) Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review* 48(6): 781–795.

Gillespie T (2018) *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ, and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. 1 edition. Cambridge, MA: The MIT Press.

Gingras R (2016) Labeling fact-check articles in Google News. In: *The Keyword*, 13 October. Available at: <https://blog.google/topics/journalism-news/labeling-fact-check-articles-google-news/> (accessed 5 April 2018).

Holovaty A (2006) A fundamental way newspaper sites need to change. Available at: <http://www.holovaty.com/writing/fundamental-change/> (accessed 4 April 2018).

Ingram M (2018) The Weekly Standard and the flaws in Facebook’s fact-checking program. In: *Columbia Journalism Review*, 18 September. Available at: https://www.cjr.org/the_new_gatekeepers/the-weekly-standard-facebook.php (accessed 24 December 2018).

Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20(1): 14–29. DOI: [10.1080/1369118X.2016.1154087](https://doi.org/10.1080/1369118X.2016.1154087).

Kosslyn J and Cong Y (2017) Fact Check now available in Google Search and News around the world. In: *The Keyword*, 7 April. Available at: <https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/> (accessed 5 April 2018).

Latour B (1987) *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.

Latour B (2005) *Reassembling the social: an introduction to actor-network-theory*. Clarendon lectures in management studies. Oxford ; New York: Oxford University Press.

OED Online (2017) widget, n. Oxford University Press. Available at: <http://www.oed.com/view/Entry/228908> (accessed 9 August 2018).

Pasquale F (2016) *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University Press.


Petre C (2015) *The Traffic Factories: Metrics at Chartbeat, Gawker Media, and The New York Times*. New York: Tow Center for Digital Journalism at Columbia University. Available at: <https://doi.org/10.7916/D80293W1> (accessed 24 August 2018).

Petre C (2018) Engineering Consent. *Digital Journalism* 6(4): 509–527. DOI: [10.1080/21670811.2018.1444998](https://doi.org/10.1080/21670811.2018.1444998).

- Ryan E (2017) Fact-checking moves into the Google Home. In: *Duke Reporters' Lab*, 8 July. Available at: <https://reporterslab.org/fact-checking-moves-google-home/> (accessed 10 August 2018).
- Star SL (2010) This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology & Human Values* 35(5): 601–617. DOI: [10.1177/0162243910377624](https://doi.org/10.1177/0162243910377624).
- Schwartz B (2017) Bing now officially supports Fact Check label with ClaimReview markup. In: *Search Engine Land*, 15 September. Available at: <https://searchengineland.com/bing-now-officially-supports-fact-check-label-claimreview-markup-282636> (accessed 10 August 2018).
- Waite M (2007) Announcing PolitiFact. In: *mattwaite.com*, 22 August. Available at: <http://www.mattwaite.com/posts/2007/aug/22/announcing-politifact/> (accessed 7 August 2018).
- Zamith R (2018) Quantified Audiences in News Production. *Digital Journalism* 6(4): 418–435. DOI: [10.1080/21670811.2018.1444999](https://doi.org/10.1080/21670811.2018.1444999).


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Share The Facts




Donald Trump

President of the United States



POLITIFACT





"This Russia thing with Trump and Russia is a made-up story. It's an excuse by the Democrats for having lost an election that they should've won."

In an interview with NBC News – Thursday, May 11, 2017

SHARE

READ MORE

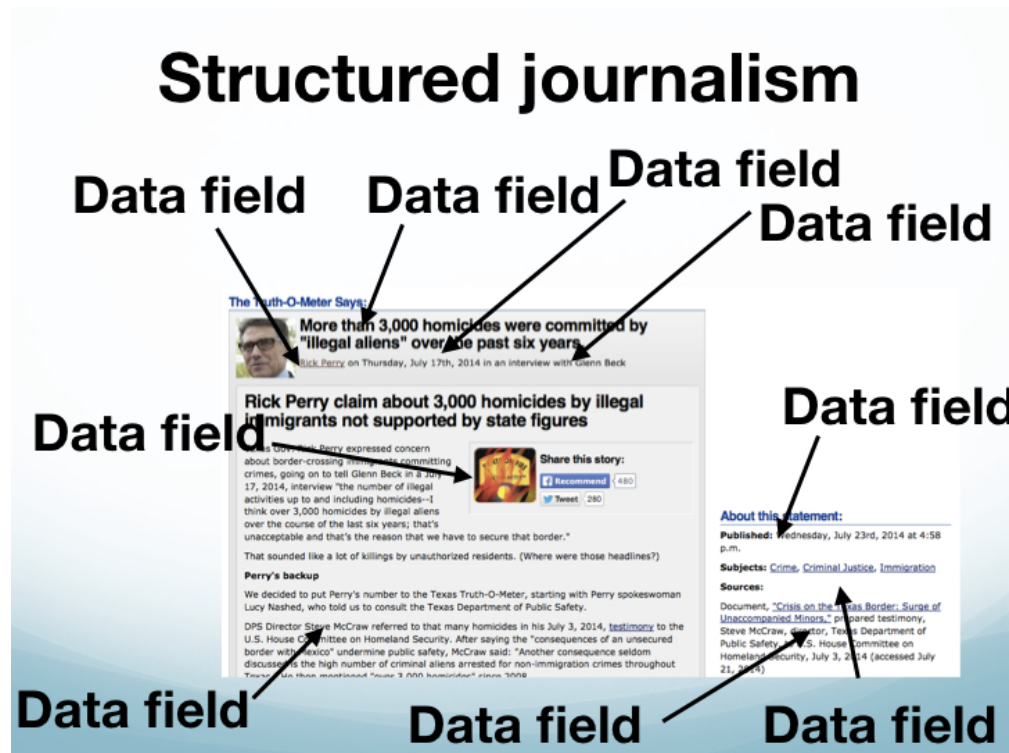




Embed

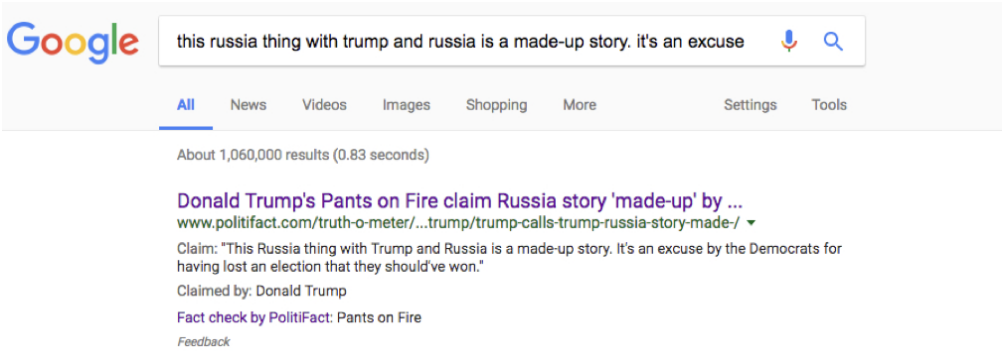
The Share the Facts 'widget' summarizes a fact check and invites sharing on social media.

305x167mm (72 x 72 DPI)



The presentation that unveiled the STF widget to fact-checkers drew on structured journalism.

254x190mm (72 x 72 DPI)



A 'snippet' in Google previews the verdict of a fact-check prominently in search results.

354x129mm (72 x 72 DPI)