Participatory Journalism: Stakeholder Perspectives on Enhancing Online Discussion through Data Talk

Cole Biehle
cole.biehle@gmail.com
UC San Diego, CA, USA

Ritvik Irigireddy
ririgireddy1@gmail.com
UC Los Angeles, CA, USA

Lu Sun
l5sun@ucsd.edu
UC San Diego, CA, USA

Steven P. Dow
spdow@ucsd.edu
UC San Diego, CA, USA

ABSTRACT

Discussing data related to news is one form of participatory journalism. *Data talk* can promote data literacy, foster community around grounded understanding, and encourage civic engagement. What are the challenges and potential of scaffolding data talk on news websites? To understand current practices and perceptions around future visions for audience engagement in news, we interviewed 12 diverse stakeholders including journalists, data scientists, readers, and moderators. To provoke future thinking, we asked participants to react to our interactive design probe with three scenarios: questioning data on a chart, adding data views, and telling personal data stories. We identified several challenges stakeholders face with data talk, including the fast news cycle hindering thorough data discussions and the difficulty of creating accessible insights and visuals. Despite significant data wrangling, discussions around the data remain rare. Participants reacted positively to the interactive scaffolding features in the three design scenarios, noting that these features can make data an effective entry point for discussions, scaffold audience participation in the data pipeline, and lower the barrier for engagement.

KEYWORDS

data talk, data visualization, journalism, online discussion

1 INTRODUCTION

With the rise of Web 2.0, researchers have exalted the potential of participatory journalism to bring audience members into the process of collecting, reporting, analyzing, and sharing news [5, 14]. On many online news sites, the commenting sections can potentially facilitate discussion and deliberation about issues with large audiences to further engage them in the news production process [3]. However, at the same time, many news platforms have scaled back or removed their discussion sections [15]. This is due to toxicity, misinformation [2], the presence of bad actors [8], and a shortage of resources for effective moderation [17]. While fruitful online discussion faces many barriers, the presentation of news has become more visual, engaging, and accessible.

In recent years, journalists have increasingly integrated data through visual chart and interactive visualizations to provide empirical backing to their stories [13]. This rise can be attributed to the proliferation of open datasets and tools supporting the data science pipeline, including data cleaning, wrangling, and visualization [20]. Could supporting discussions around and through *data* be the key to empowering public participation in news?

Research suggests that integrating data visualizations encourages readers to engage more deeply in "data talk", although such discussions are still uncommon [16, 22]. Data talk involves referencing data directly, allowing readers to anchor their discussions in factual information and share their personal stories about data [16].

Prior research has explored strategies to support online discussion around data and data visualization. The SenseUs system supported collaboration and discussion through sharing, commenting, and annotating data in an interactive visualization [9]. The CommentSpace is a collaborative visual analysis system that encourages community members to add tags and links between comments and visualizations. This collaborative system incorporates audience wisdom into data exploration and aids the data analysis process [23]. While these visions for data-centered discussions are compelling, they have not come to fruition for online news platforms.

We aim to investigate the challenges and opportunities for supporting data talk on news websites. We sought to understand the diverse perspectives of stakeholders implicated in data talk, namely journalists, data scientists, readers, and moderators, to gain insight into designing infrastructure to support richer data discussions in news contexts. We seek to answer two research questions:

- **RQ1: What are the challenges for supporting data talk on current news websites?**
- **RQ2: What are the perceived benefits and barriers for an interactive news platform that scaffolds data talk?**

To answer these questions, we conducted an interview study with 12 participants across diverse stakeholder roles. To encourage our interviewees to imagine possible interactive features for news discussions around data, we developed an interactive prototype that

served as a "probe" for eliciting opinions from our key stakeholders [7]. The key features included questioning data, adding data, and telling stories related to data. These semi-structured interviews lasted about 45 minutes and covered their current practices in data journalism, their engagement with online news discussions, and their reactions to our design probe.

Interviews revealed that, despite the abundance of data and visualizations, several barriers hinder effective data talk in online journalism. The fast-paced nature of news prevents journalists from having in-depth engagement with individual articles, limiting communication with diverse audiences. Additionally, the significant time and effort required to transform raw data into effective visualizations result in insights being less accessible to readers. These challenges highlight the need for careful design to encourage meaningful audience participation in data discussions.

Participants expressed positive reactions to the mocked up news website with enhanced features, including scaffolding features and side-presented comments. They agreed that enhanced features on top of the visualization can foster better, more organized conversations. Stakeholders appreciated that data visualizations could serve as entry points for discussion and that the audience could contribute to the data visualization pipeline. They also valued how these features lower the bar for participation by allowing users to question data, add data, and share personal stories about data.

2 RELATED WORK

Data Journalism and Data Talk on News Websites. Data has seen a growing presence in online discussions due to its ability to increase collaboration and allow audiences to receive information more clearly [4]. Many systems have started to enable participatory journalism where journalists collaborate with broader audiences to collect data, edit data, and communicate information about data [5, 10, 14]. Despite the increased presence of data in journalism, readers are generally not discussing the data. One study found that out of thousands of comments on a news website, only a small percent reference the data or visualizations in the article [16].

Incorporating data into written discussions can cause data to be manipulated to fit a specific narrative [11, 19]. This creates the problem that the different forms of data can be inconsistent and therefore confusing to readers [18]. Research also points out that data visualizations, together with narratives, may have framing effects that can affect end-user interpretation of the issues [11]. These challenges underscore the need for understanding the perspectives of all potential stakeholders, from readers to moderators, and carefully designing features to encourage effective data talk.

System Designs for Scaffolding Discussion around Data. Existing systems have explored features to incorporate data into online discussions and enhance data talk. The CommentSpace platform offered features like tagging, linking, and sharing views to scaffold crowds to make sense of the data visualization [9, 23]. SenseUs helped audiences collaboratively analyze the data through view bookmarking, double-linked discussions, and graphical annotation [9]. To facilitate storytelling and aid in explaining complex data, Contextualifer developed features to create data visualizations on top of news articles and to provide customized annotations with

references [12]. Existing research also emphasizes that storytelling based on personal experience with the data can facilitate a more collaborative discussion environment [6]. Researchers highlighted that interactive features that use data as an entry point can potentially scaffold deeper sensemaking on the data narratives [11]. In our study, we designed a probe with three enhanced features that allow audiences to question the data, add data views, and tell personal stories. We further introduce the scenarios for each feature.

3 A DESIGN PROBE TO EXPLORE SCENARIOS FOR DATA TALK

To facilitate a discussion about the potential for data talk around online news, we developed a design probe to incorporate three distinct scenarios where readers engage in discussion around data and data visualizations through the help of interactive scaffolding.

Scenario 1: Questions about the data. The first scenario integrates discussion within the context of the data visualization, encouraging deeper engagement through accessible visual data. When the reader, Ifrah, explores the data visualization section (as shown on the left in Figure 1A), Ifrah notices that COVID-19 cases began rising sharply in early July. Prompted by this observation, Ifrah questions, "What about the government response to this increase in cases?" The data talk Q&A feature contextualizes Ifrah's question within the visualization, marking it with a question mark icon. By clicking on this icon, Ifrah can view responses from other community members. For instance, Brenda responds with a link indicating that the government posted a recommendation to wear masks regardless of vaccination status. This **data Q&A** feature uses data as an anchor to focus attention on the visualization and to provide a concrete way for users to engage in discussion threads for their questions and answers.

Scenario 2: Adding data views. Providing diverse perspectives on the data and adding data views could enable newsrooms to open new forms of participatory journalism. As shown in Figure 1B), readers can switch to different data views in a carousel, each with user-submitted captions. For example, Kevin switches to a different tab in the carousel to view data submitted by another reader to understand how increasing cases affected hospitalizations. This **data views** feature imagines how readers could enhance the discussion by providing additional context and data, enriching the overall understanding of the situation through diverse perspectives.

Scenario 3: Telling stories related to the data. A visualization may also provoke readers to think about how the data impacts them personally or how the data represents them. To support personal storytelling, readers can annotate the visualization to tell their story, and this becomes a hook for other readers to reflect and share, as shown in Figure 1C). Faced with the increasing trend of COVID-19 cases, Ritvik shared the story of when his family caught COVID. Following this personal account, Kristina and Jonathan responded with their own experiences. This **data stories** feature leverages the power of personal narratives to deepen engagement and foster a sense of community among readers.

A

B

C

Figure 1: Three scenarios that participants interacted with during the interview. Scenario A - questioning data, the reader raises a question for clarification. Scenario B - adding data views, the reader adds more data to display different perspectives. Scenario C - telling personal data stories, the reader shares their own experience related to the data.

4 METHOD

Qualitative interviews were used to explore the challenges for supporting data talk. We conducted 12 interviews with diverse stakeholders through online communication tools. Each interview lasted approximately 45 minutes. The first half focused on current practices and the second half explored reactions to the design probe.

Participants. Participants included representatives from four key stakeholders that would be implicated in data-centered participatory journalism: journalists (J1-J3), data scientists (D1-D3), readers (R1-R3), and news website moderators(M1-M3). Demographics are shown in Table 1.

4.1 Interview Procedure

Phase I: Current Practice. In the initial phase, participants discussed their perspectives on news, data-driven discussions, and comment sections. After obtaining consent, we collected background information, including occupation, daily activities, field experience, and engagement with news. The interview focused on their work with news, perceptions of comments, and the challenges around reading and managing comments on news websites.

Table 1: List of Interview Participants

ID	Type	Experience
J1	Journalist	25-year career, covers mainstream news.
J2	Journalist	2 years at a local paper, covers local news.
J3	Journalist	20-year reporter, specializes in public health.
D1	Data Scientist	Uses data for engagement and education.
D2	Data Scientist	Data science educator, uses data in teaching.
D3	Data Scientist	Data science educator, uses data in teaching.
R1	Reader	U-M grad, works in finance.
R2	Reader	Global studies undergrad at ASU, also a tutor.
R3	Reader	Law student, reads cases extensively.
M1	Moderator	Content moderator at a major news network.
M2	Moderator	Focus group panel moderator.
M3	Moderator	Content moderator at a major news network.

Phase II: Design Probe Interaction. In this phase, we introduced a Figma prototype and asked participants to interact with three different versions of the news platform designed around a specific article from the San Diego Union-Tribune [21], as shown in Figure 1. Participants first experienced the *original version* which included only the title and body of the article with one data visualization.

Just as one would experience on a typical news website, there is a comment section at the very bottom of the page. The second version was like the original, except with the comment section beside the article's text and visualization to see if placement alone makes discussion engagement more enticing. The third version is our *enhanced* experience where the discussion can be accessed through annotations on the data visualization, as described in section 3. The enhanced version featured a carousel of visualization thumbnails, allowing users to observe additional data views or add new data. Users could also suggest new caption text for each visualization.

Data and Analysis. All twelve interviews were recorded and transcribed. Two researchers conducted iterative open coding on the transcripts using Google Doc following the thematic analysis [1]. They open-coded the data by identifying topics mentioned by each group of participants in each interview phase. After iteratively discussing the code themes, researchers derived the final themes around stakeholders' challenges for supporting data talk and perceived benefits and barriers for the enhanced features.

5 RESULTS

5.1 RQ1: Challenges for supporting data talk on current news websites

The fast news cycle rushes journalism, hindering potential data talk. Journalists highlighted challenges with the relentless 24-hour news cycle. J3 pointed out that much information is released at night, causing journalists to rush: "They're scrambling to get something in the paper the next day.[J3]" J2 explained that basing articles on data is often infeasible because "data analysts can be a little slow to get [data and data visualizations] to us[J2]" despite their value to stories.

Both journalists and moderators noted that the fast news cycle restricts their engagement with reader comments. J2 admitted, "I don't even always look at the comments, to be honest. You always have to move to the next story pretty quickly.[J2]" Similarly, a moderator mentioned, "We cannot fact-check; there's just too many comments.[M3]" The speed of news hinders moderation, fact-checking, and community-building with readers.

Creating accessible insights and visuals requires significant data wrangling. Effective data visualizations require extensive data wrangling. Journalists shape data to tell a story, which can introduce biases. J1 described, "You start with the raw data and then you're discussing how it should be visualized, and you know there's manipulation that happens too.[J1]" Data scientists worry about collecting data to fit a story. D2 noted, "Collecting data [on the story's subject] can be inconsistent and not very reliable.[D2]" This highlights the challenge of ensuring data reliability.

Creating visualizations needs expertise from dedicated data science teams. Data scientists described their process, from data collection to visualization. Despite concerns about framing effects, they acknowledged that manipulating data helps the community understand it. D2 emphasized, "Understanding data visualizations and data in general can be a powerful tool for bringing change to communities and explaining certain issues very clearly.[D2]" These insights underscore the complexity of data wrangling to create

accessible and interpretable visualizations, balancing data integrity with effective communication.

Discussion around data is rare and can be distracting. Journalists expressed significant concerns about comment sections, describing them as "a venue for attacks, and misinformation.[J3]" They noted that readers often repeat incorrect information, even after it has been debunked. As J3 stated, "I'll just see people repeating things that we've written previous stories about and shown that they're [false information].[J3]" This repetition undermines constructive dialogue and turns comment sections into a battleground.

The quality of comments is also a concern. J2 remarked, "It's pretty rare to find like something constructive or specific or [something that] feels worth paying a whole lot of attention to.[J2]" Many readers react impulsively to headlines without engaging with the article. J1 noted, "It seems like readers are reacting to the headline and not really reading.[J1]" The overwhelming volume and poor quality of comments dissuade meaningful participation. As M3 put it, "There's a lot of junk out there... it does discourage you from participating.[M3]" This sentiment highlights the challenge of fostering valuable discussion in an environment dominated by misinformation and superficial engagement.

5.2 RQ2: Perceived benefits and barriers for an interactive news platform for data talk

Data can serve as an effective entry point for discussion. Participants preferred the news website platform with enhanced features, including three scaffolding features and side-presented comments. They agreed that displaying comments on the side improves discussion quality by fostering better organized conversations.

These features contextualized discussions within specific areas of data visualizations and provided effective entry points for data talk. Readers noted that this structured interaction helps them "get better connections by [participating in discussions] because you can have more specific conversations, and you can also have more organization and focus on what you're talking about.[R2]" Data scientists recognized the potential benefit of "explaining certain issues very clearly at an individual level where people can react.[D2]" Moderators appreciated the mechanisms integrating community contributions into discussion. As M2 noted, "I love how you have the community contributions there. The way you delineated it between the Q&A generated by the community, the data stories, and the data sources.[M2]"

Scaffolding audiences to work as a part of data pipeline. Journalists recognize the value of audience input in shaping stories through comment sections. The data Q&A feature encourages active engagement, aiding in the identification of emerging topics. J2 emphasized that possibility of readers "raising good questions and concerns that could make for the next story,[J2]" highlighting the direct impact of reader inquiries.

Listening to the community without assumptions is key to incorporating audience input. D2 explained "To just listen, not ask [the community] a ton of questions, and understand their world. Not make a lot of assumptions. That was a very first step in starting to collect data.[D2]" As M1 mentioned "here is a mutual benefit for both the journalist and readers, as journalists can be updated on

the status of the community via those who truly want to discuss their content. They can use these discussions to find leads.[M1]" Moderators appreciate how it engages audiences and contributes to data curation. This involvement makes readers integral to the data pipeline, improving the quality of discussions.

Lower the bar for participation through personal storytelling. Personal storytelling in data visualizations boosts audience engagement by humanizing complex information. R1 explained, "I don't start too many posts, but I'm like upvoting and down-voting, and commenting and replying to other people's comments.[R1]" This passive engagement can become active participation by including personal narratives. M1 emphasized that "people just are more drawn in by [personal stories] and can better relate to subject matter when it's told through a personal story,[M1]" illustrating how personal anecdotes can make data more relatable. Additionally, M3 explained, "If someone looks at a chart and it's a bunch of numbers, it can be very impersonal...[if] you can click on it and then have someone talking about it, maybe it would resonate more with people.[M3]" Adding personal anecdotes to impersonal charts increases resonance and fosters a more interactive reader experience.

6 DISCUSSION

Our interviews revealed that participants preferred interactive news websites with enhanced features to scaffold data talk. The data view feature allows audiences to contribute diverse perspectives on data. The data Q&A feature facilitates participatory journalism, lowering the bar for collaboration between audiences, journalists, and moderators. The data stories feature engages audiences by enabling them to create data narratives. While the implications of these scaffolding features are promising, concerns about deploying and maintaining them need to be addressed.

Operating and maintaining a system that incorporates community-generated content poses challenges for moderators. Moderators highlighted the difficulty in determining the accuracy of crowd-sourced visualizations and stories. Journalists expressed concerns about the time investment required to verify the accuracy of changes made by commenters. Despite these challenges, moderators see value in expanding interactive programs like the reporter reply program, as it fosters a more informed and connected community.

Limitations. The prototype in the study was created in Figma, limiting the amount of interaction and immersion. Also, the focus was on individual reactions, neglecting the social aspect of these discussion spaces. Only 12 interviews were conducted, more interviews would be necessary.

7 CONCLUSION

Data talk has potential benefits for all stakeholders within the journalism ecosystem but it is infrequently seen in current practice. A fast news cycle, intensive data wrangling, and insignificant contributions are notable barriers. Our interviewees expressed that an improved discussion system could make data the entry point, include the audience in the data pipeline, and lower the bar for participation. Considerations of managing and not overwhelming users would need to be addressed, but stakeholders were hopeful about an increased involvement within these online news spaces.

8 ACKNOWLEDGEMENT

REFERENCES

[1] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.
[2] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial behavior in online discussion communities. In *Proceedings of the international aaai conference on web and social media*, Vol. 9. 61–70.
[3] Kevin Coe, Kate Kenski, and Stephen A Rains. 2014. Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of communication* 64, 4 (2014), 658–679.
[4] Catalina M Danis, Fernanda B Viegas, Martin Wattenberg, and Jesse Kriss. 2008. Your place or mine? Visualization as a community component. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 275–284.
[5] Katherine M Engelke. 2019. Online participatory journalism: A systematic literature review. *Media and Communication* 7, 4 (2019), 31–44.
[6] Sheena Erete, Emily Ryou, Geoff Smith, Khristina Marie Fassett, and Sarah Duda. 2016. Storytelling with data: Examining the use of data by non-profit organizations. In *Proceedings of the 19th ACM conference on Computer-Supported cooperative work & social computing*. 1273–1283.
[7] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: Cultural probes. *Interactions* 6, 1 (jan 1999), 21–29. https://doi.org/10.1145/291224.291235
[8] Daniel Halpern and Jennifer Gibbs. 2013. Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in human behavior* 29, 3 (2013), 1159–1168.
[9] Jeffrey Heer, Fernanda B Viégas, and Martin Wattenberg. 2009. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM* 52, 1 (2009), 87–97.
[10] Bahareh R Heravi and Mirko Lorenz. 2020. Data journalism practices globally: Skills, education, opportunities, and values. *Journalism and media* 1, 1 (2020), 26–40.
[11] Jessica Hullman and Nick Diakopoulos. 2011. Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2231–2240.
[12] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. 2013. Contextifier: automatic generation of annotated stock visualizations. In *Proceedings of the SIGCHI Conference on human factors in computing systems*. 2707–2716.
[13] Jessica Hullman, Nicholas Diakopoulos, Elaheh Momeni, and Eytan Adar. 2015. Content, context, and critique: Commenting on a data visualization blog. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1170–1175.
[14] Regina G Lawrence, Damian Radcliffe, and Thomas R Schmidt. 2018. Practicing engagement: Participatory journalism in the Web 2.0 era. *Journalism practice* 12, 10 (2018), 1220–1240.
[15] Jiawei Liu and Douglas M McLeod. 2021. Pathways to news commenting and the removal of the comment system on news websites. *Journalism* 22, 4 (2021), 867–881.
[16] Brian James McInnis, Lu Sun, Jungwon Shin, and Steven P Dow. 2020. Rare, but valuable: Understanding data-centered talk in news website comment sections. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–27.
[17] Maria N. Nelson, Thomas B. Ksiazek, and Nina Springer. 2021. Killing the Comments: Why Do News Organizations Remove User Commentary Functions? *Journalism and Media* 2, 4 (2021), 572–583. https://doi.org/10.3390/journalmedia2040034
[18] Zening Qu and Jessica Hullman. 2017. Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 468–477.
[19] Cindy Royal and Dale Blasingame. 2015. Data journalism: An explication. In *International Symposium in Online Journalism*, Vol. 5. 24–46.
[20] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1139–1148.
[21] Paul Sisson and Jonathan Wosen. 2021. With a wave of new coronavirus cases, will it soon be masks up in San Diego? *The San Diego Union-Tribune* (2021).
[22] John Stasko. 2014. Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. 46–53.
[23] Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. 2011. CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 3131–3140.