

## 16<sup>th</sup> Annual International Symposium on Online Journalism

Day 2, April 18, 2015: Research Panel – 11:30-12:45p.m.

### ***The Changing Tide in Journalism Structures, Systems and Processes***

**Chair & Discussant: Amy Schmitz Weiss**, Associate Professor, **San Diego State University**

#### **Panelists:**

- **Nicholas Diakopoulos**, University of Maryland: *Picking the NYT Picks: Editorial Criteria and Automation in the Curation of Online News Comments*
- **Soraia Souza**, Estácio de Sá University, Brazil: *The Age of Digital Collaboration: A Case Study of Arquitetura da Gentrificação, a Brazilian Collaborative Platform for Citizen Journalism*
- **Cindy Royal** and **Dale Blasingame**, Texas State University: *Data Journalism: An Explication*
- **Joshua Scacco**, Purdue University, Alexander L. Curry and Natalie J. Stroud, University of Texas at Austin: *Digital Divisions: Organizational Gatekeeping Practices in the Context of Online News*

---

**Nicholas Diakopoulos:** So thanks, Amy. So, I'm Nick Diakopoulos from the University of Maryland, College Park. And I'm going to talk to you today a little bit about my research looking at New York Times picks, comments, and what makes a New York Times pick, and thinking about the editorial criteria that play into those decisions, as well as how we might think about automating some of those decisions.

So, I'm going to take you back a couple of years. Some of you might remember good old Vlad Putin published this Op-Ed in *The New York Times* in September of 2013. Among other things, he questioned American exceptionalism, which if there is one thing I know you shouldn't do in America, it's that. [laughter] So, he was provoking the American public a little bit. And it worked. He got 6,367 comments submitted from readers against his Op-Ed, of which 4,447 were eventually published by *The New York Times*.

So, how could we possibly think about organizing 4,000-some-odd comments, right? I mean, and still getting to some sort of interesting and insightful ones. Well, like other commenting systems, *The New York Times* allows you to vote on comments. You can recommend them. You can sort in different ways. You can sort by recommendation score, by oldest or newest first, and so on. But they also have this feature of The Times Pick. So, these

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

are comments that are selected to sort of represent the most interesting and thoughtful comments that have been made on the article. And what's perhaps most impressive about this system is that each of these 4,400 comments was actually read by a human moderator before it was published. So, they have a system of pre-moderation that they use.

This man, Bassegy Etim, is the Community Manager at *The New York Times*. And together with his team of 13 other moderators, they read almost every comment that's published to *The New York Times*. Part of that job is deciding whether or not a comment gets published to begin with. And part of that job is also, you know, selecting these high quality New York Times Picks comments. He told me that New York Times Picks is the most popular comment queue on the site, and that they really spend a lot of time tweaking that queue and trying to get it right.

Now, there are some potential benefits to selecting high-quality comments on news sites. The idea is that if you can signal to your audience what is a high-quality comment, what you expect from a comment, the idea is that this can create a beneficial feedback loop for your audience, where you're signaling and telling them, "Hey, you should do more of this." So, that is the theory at least.

The question that I want to pose in this research though is, you know, what are really the criteria that are used for selecting these comments? What are the editorial criteria? And if we were going to help *The New York Times* go from the 22 stories a day, where they allow comments now, to hundreds of articles perhaps having comments open on them, how could we scale up this process? Could we use automation to potentially augment moderator's capabilities to recognize and consider and detect higher quality comments, so that they can highlight it?

So, as a good researcher, I started in the literature, reading into various studies that have looked at letters to the editor, online comments that have been remediated for print, as well as on-air radio comments, and thinking about, what are the criteria that journalists have used in other contexts to select comments for publication? And there's various criteria that come up.

There's the sort of negative and exclusionary criteria, personal attacks, profanity, [and] abusive behavior. Some of these are supported in software like Keepcon that tries to help publishers deal with scale and excluding these kinds of comments.

That is not the focus of my study. I'm actually much more interested in the positive, high-quality end of comments and the sort of inclusionary criteria that are used by moderators to select high-quality stuff. So, this would be things like argument quality, entertainment value, readability, mentions of personal experiences, thoughtfulness, relevance, fairness, and novelty. So, there are 12 of these criteria that I identified.

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

And certainly, there's a lot of work that needs to be done in the future for thinking about how to operationalize these in interesting and useful ways. But as a sort of first stab at this, I approached this as a crowdsourcing project and was thinking, you know, do *New York Times* Picks actually reflect some of these editorial criteria? And if so, which ones and to what extent?

So, for the eight, I guess, nine criteria that are highlighted there, I did a crowdsourcing experiment and asked human readers on Amazon Mechanical Turk crowdsourcing platform to rate these comments on a scale from 1 to 5 along each of these dimensions. You can get more details in the paper. [I] had 500 comments: 250 that were Times Picks comments and 250 that were not *New York Times* Picks comments. And then collected these ratings on a scale from 1 to 5 for each of these criteria. So, based on the interrater reliability, it was sort of slight to moderate, and so, just sort of keep that in mind as we move forward.

So, I was also interested in looking at automation. Which of these criteria could we potentially automate? And certainly many of these would be very challenging for computer scientists to even think about programming in code. So, [I] sort of started with the low-hanging fruit and picked off three of them that were easy enough to develop algorithms for scoring these things. So, readability, personal experience, and brevity/length were automatically computed.

So, something like readability, I used a reading grade level score that was actually developed in the late sixties that just measures sort of the complexity of word use in the comments. I also developed a more novel score to measure personal experience, where I look at the proportion of words in a constructed dictionary that's called LIWC. It comes out of Austin. And this dictionary has words that reflect personal experiences or family and friends kinds of relationships. And basically, [you] look at counting up the proportion of words in a comment that come from those dictionaries and that becomes the score.

The length is trivial to operationalize computationally. You just tokenize based on the white space between words and then you have a count of the number of words in the comment.

So, here are the results in a nutshell with the exception of two of the criteria. I found a statistically significant difference in the criteria as rated for *New York Times* Picks comments compared to non-*New York Times* Picks comments. so, *New York Times* Picks in the darker blue is consistently greater along these six dimensions than non-*New York Times* Picks comments. And again, you can sort of see the details in the paper.

There was no physical difference for entertainment value. Again, humor is a very difficult thing to measure. The crowdsourcing apparatus that I had

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

wasn't able to sort of key in on that. And then emotionality was also something that there was a little bit of a trend difference, but it was not substantially significantly different.

In terms of the automated scores, again, we see the exact same trend. So, for *New York Times* Picks comments, they had a higher brevity score, they were more readable, according to this index of complexity of language use, and then they also had a higher personal experience score. And these were all statistically significant.

So, where does that leave us? Well, I think the development of automated scores sort of uncovers a number of opportunities for us to think about moderating comments and news in new ways, as well as developing new user experiences.

For instance, the personal experience score that we developed could have some really nice applications in terms of amplifying the value of comments for moderators. We know many moderators are interested in finding personal perspectives and so on, and these could be selected for using that filter.

And in this little scatter plot that I'm showing, we're working on some follow-up work, an interface called Comment IQ, for moderators to be able to see visually how different scores are sort of splayed out along these axes. And you can sort of lasso and select comments that, for instance, are more relevant and reflect more personal experience. And we think that this will enable moderators to be able to find these comments more efficiently.

We can also imagine new end-user experiences. So, if we can automatically score comments in these different dimensions, we can put that into the front-end user experience. Someone could sort of dial in, you know, whether or not they want to see comments that are more readable or with higher personal experience and so on, and let them sort of adapt and personalize their experience of the comments for themselves.

I think automation also raises some interesting questions. In particular, we know that different communities, different types of topics—sports, for instance—they do require different treatment from moderators and some understanding that different contexts demand different types of criteria that are applied.

So, I think we need to be careful about over-generalizing algorithmic solutions. I don't want us to head down the road of sort of a 'one size fits all,' "Yeah, let's just take the algorithmic hammer and all of a sudden we have this magical high-quality comment detector." I think we need humans in the loop [and] journalists in the loop to sort of be able to adapt the algorithms to the particular context that they are in and know what fits when and where.

## 16<sup>th</sup> Annual International Symposium on Online Journalism

And so to that end, we're sort of, again, in this Comment IQ interface, we're playing around with this idea that you can reweight the criteria interactively and sort of specify. The journalists can specify, "Do I want more of this and less of that?" and so on. And that would let them adapt to different kinds of moderation context.

So, that's really it for me. I would be thrilled to take questions during the panel. And for more information, you can also see a paper that was recently published in the Computer Supported Cooperative Work comments, where I was also looking at *New York Times* Picks comments.

And a special thanks to the Knight Foundation for funding this work as part of a Knight prototype grant. Thank you.

[Applause.]

**Soraia Souza:** Hello, everybody. Good morning. My name is Soraia Souza. And I'm talking with you about *Arquitetura da Gentrificação*, a Brazilian journalist platform for citizen journalism.

First of all, I want to talk with you about Journalism in Brazil. So, we have ten families that control mainstream media in Brazil. That means that few people choose what Brazilian citizens must know, must see, must read; hence, a little bit concern. And second issue is that gathering information, and most important, gathering digital information in Brazil is really tough. It's nearly impossible.

We have a specific act called Information Access Act that assures all citizens to access information, but even so, we face a hard time to get it. The third one is that Sao Paulo faces a real problem with gentrification. So, wealthy people have access to transportation [and] wealth care, and dispossessed people don't. All this scenario and the project itself led me to these three questions.

The first one: Is it possible to make investigative articles without the active support of a large media mainstream—company, sorry. Is crowdfunding a valid initiative to finance journalistic projects in Brazil? And the last one: Is the digital platform the best suited to the publication of broad and extensive research as presented by *Arquitetura da Gentrificação*?

In order to answer these questions and to understand the project, I read some of these authors. So, I have [in] public space, Habermas. We have self-communication, Costells. I also have some Brazilian authors. And I also did an interview with Sabrina Duran. Sabrina Duran is a Brazilian journalist, and she's the leader of the project.

So, we have *Arquitetura da Gentrificação*, the project. The first phase of *Arquitetura da Gentrificação* was completed in December three years ago. It

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

has a website. It also has social media accounts and a blog. It's all data-driven journalism based, and it's also [an] open data environment access. And during the research Sabrina Duran and her team had a tip from a source. And this tip led to the Phase 2, privatization of the street. It's about an area in downtown Sao Paulo. And this second phase has two projects: a timeline that's already available and a video documentary that's going to be completed next semester.

In order to understand a little bit better about this project, I brought to you an article called Bancada Empreiteira. It's about loving a downtown area of Sao Paulo. So, you have the article itself. You also have the infographic that's a very interactive infographic, so you can see all the projects. You can see the big companies, the big constructors. You can see the political parties involved in that. You can see also the city representatives. So, you can see all the path of the money, all the donations.

If you don't believe in the article, if you don't believe in this interactive infographic, you can also access all the data that they used to do it. So, it's available online and you can see the -- also, you can see the original data. So, it's available for everyone who wants to know about it.

All this data led me to some conclusions. The first one is that crowdsourcing and crowdfunding in Brazilian journalism is possible, but crowdfunding isn't enough to provide all the financial support that Sabrina Duran and her team needed to do the project. So, they had to look for other forms of income. It's not the best thing to do in this kind of project.

Also, independent investigative journalism as well [as] access to information in Brazil is really, really time consuming. Because if you don't belong to this mainstream media that I already told you [about], we face a real hard time to do it. So, not only because we don't have the financial support, but also because access to public information is time consuming. So, you set an interview, but nobody knows you, and nobody answers it. Then you have to ask them again. They don't answer you. You have to call a lawyer.

So, I have this Information Access Act. Can I access this public data? They still say, "OK, let's do it." It took almost 30 days to get the answers. And they are like, "Yes. No. Maybe." So, you have to call your lawyer again and ask all the questions. So, it's really time consuming.

And the last one is that, yeah, the online platform was best suited, so you can access the data, you can access the video, the photos, [and] everything that they used to do the articles [and] to do the timeline. So, it was the best suited platform.

I'd like to thank Estácio for supporting me, financially supporting me, and I also must thank all of you for this presentation. Thank you.

## 16<sup>th</sup> Annual International Symposium on Online Journalism

[Applause.]

**Cindy Royal:** Thank you very much, Amy. Thank you, everyone. Put my water back here. And I just want to say before I get started how much I enjoy coming to this conference every year. So thank you, Amy and Rosenthal. I have to tell you all there have been some very monumental things that have happened to me as a result of this conference, from meeting people who I've been on panels with at other places, like Trei Brundrett—he was actually in my class this week helping me teach my class—[to] students getting jobs from coming to events like this, and my meeting someone who led to my Knight Fellowship at Stanford last year. So, look to the left. Look to the right. If you don't know the people next to you, make a friend, because you have no idea what a meaningful impact that they can have on your life.

I always love coming here and bringing Texas State students. We have a big contingent from Texas State this year. Y'all raise the roof, everybody. Excellent. So, I teach at Texas State. I teach digital media skills, concepts. My colleague, Dale Blasingame, right here in the front, worked with me on this project. He also teaches in this area. Dale came to us from a background as an Emmy award winning news producer. So, I recruited him to work with me on this project to do the video elements of it. We're so excited to be part of the multimedia aspects of the new journal.

So, let's get started here. you know, one of the things -- one of the people that I met here in 2008 was Aron Pilhofer, who was at *The New York Times* at the time. And very quickly, I invited myself to spend a week with them at the Interactive News Team at *The New York Times*. And that next year, I presented that research here at ISOJ about just sort of understanding who these data journalists are. So, I did the research in 2009. So, we're six years on from that. God, that just seems like yesterday.

So, this research is sort of the next step for that. And actually, it's a step back. It's like, well, what do we actually mean when we talk about data journalism now? What is it? Is it a process? Is it a product? Is it roles or skills that we're talking about? Who does it affect? Does it even matter? You know, is it really anything new at all? Is it just journalism? Or, is it all of the above? Is it all these things and we need to be maybe taking a more nuanced and systematic look at how we describe it and how we categorize it?

There have not been a ton of research projects done in this area. There was the one that I did and there was one on, I think, it was in new media and society that some French scholars did on *The Chicago Tribune*. And just recently Seth Lewis has edited a digital journalism issue on big data, so a lot of new research has come with that. So, I think there's plenty of opportunities for us to study this field once we kind of know what it encompasses.

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

So, the method that I used was grounded theory. I wanted to generate as many assertions as possible of people defining or describing, what is data journalism? And I had to identify what those sources were. And because there were not a lot of academic papers written on it, I looked first in the communication source EBSCO Host, but then I also looked at other sources that I knew had been doing work in this area; some other academic sources like Neiman Lab and the Tow Center and some professional sources that I identified, like PBS MediaShift and Poynter and things like that. I did some Google searches to make sure I wasn't missing some of the key people, some of the key sources who were looking at it.

So, I ended up with 63 assertions and from 23 sources. More than half came from those other academic sources not represented in sort of the typical journal articles, but I think we'll probably see that changing. So, I identified the sources. I identified the assertions. I did a word frequency analysis. And you'll see the results of that in a second. I used a python script to just count words and all the assertions, so you can see what words were used. Coding of dimensions. And then I developed a conceptual definition that's hopefully a jumping off point for some of you.

And then we went to Online News Association in Chicago, and I was like, "We're going to have all these great data journalists there. Why don't we ask them, what is data journalism?" So, at the very end of this, I'll show you a quick trailer, but we have several videos embedded into the research paper that we've sort of coded against the dimensions that I identified.

So, the assertions by year. You can just see that, you know, the interest in data journalism has increased over the years. And this was the frequency analysis I did of the terms. [It] may be hard for you to see, so you can look at it in the journal, but words like *storytelling* and *reporting* and *sources* came up. Sounds like journalism, right? The things that are new are when you talk about *personalization*, *customization*, *databases*, *interactive*. So, there's some new elements and there's some existing elements that really -- I mean, we are talking about journalism and storytelling here.

And then the dimensions that I identified through the grounded theory process. There were seven of them. They are: process, product, convergence of field, traditional, outside influence, skills, and then there was like a hybrid area.

So, the biggest one was the process dimension. And I won't read all of these, because there's several examples in the paper, but process dimensions -- process assertions were anything that talked about the activity. It's a reporting process. It's about aggregating, filtering, or visualizing when they talked about the process.

Product dimensions, where anytime they mentioned, say, an infographic, interactives, charts, tables, databases.



## 16<sup>th</sup> Annual International Symposium on Online Journalism

Convergence of fields had to do when they were saying, "Well, it merges this field with that field. It's journalism and computer science. It's statistics and investigative reporting." So, anytime it was multiple fields converging together, we coded it that way.

And then there were traditional aspects of it where we say that, you know, maybe this is nothing new. It's the same as it's always been. Data journalism hasn't changed very much. It requires some of the same skillsets.

And then there were those that reflected outside influence, and that's, who does it matter to? Things that mentioned the individual, their ability to use these products, and democracy.

Then skills we talked about. We found places that talked about the skills that people needed to have to be able to practice this.

And then there were a lot of hybrid definitions. That, you know, kind of cut across two and sometimes three or four of the dimensions.

So, I know that's a very quick sort of trip, but I want to get to the video because, again, it's a trailer. Some of you may see yourself in this if you were at ONA. The comprehensive definition I came up with is: *Data journalism is a process by which analysis and presentation of data are employed to better inform and engage the public. Its roots are in the fields of computer-assisted and investigative reporting, but data journalism products may add engagement through customization and user contribution made possible by web development and programming techniques.*

That's a big conceptual definition. I'm hoping that these dimensions will allow you to create operational definitions that will focus your research and provide input into this very important growing and dynamically changing area.

So, I'm going to bring up the video.

[Video begins.]

*Man: What is data journalism?*

*Man: This has to be one long definition, huh?*

*Man: Uh...data journalism...*

*Man: I have lots of definitions for data journalism.*

*Woman: See this mess? This is the answer. That is data journalism.*

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

*Man: Data journalism is all about quantifiable information. It's about information that can be organized and information that can be sorted through.*

*Woman: I believe data journalism is looking at structured or organized information to find meaning in our world.*

*Man: Data journalism treats data as a source like any other. You—you go out and you find it. You identify it as having information about something you need and you ask it questions.*

*Woman: Data journalism is digging into information and making sense of it for people.*

*Man: Data journalism, my opinion is just using data for journalistic purposes.*

*Man: Data journalism, to me, is telling stories and reporting with data, with numbers, having a fluency with the stuff so that you can make it accessible and interesting to regular people.*

*Man: Data journalism is journalism that uses information from spreadsheets and information sources from the web, tabular data. I'm trying to not use the word 'data' in the definition, but it's a way to use data to directly do the work of journalism. So, to tell stories and to reveal the facts and to find the bad guys and show people what they've done.*

*Man: Data journalism is journalism that is done where one of the sources that you're interviewing is some kind of data, some kind of number set.*

*Woman: Right now, we just trend the data up as saying data journalists because we use big data, big numbers. Instead of doing interviews one-on-one, we have computer-assisted journalism that help us get lots of data, but it's lots of interviews.*

*Man: It's gathering a mass amount of information and trying to find stories within it to help among our communities.*

*Woman: For us, doing data journalism is a big effort, because we don't have the resources.*

*Woman: Maybe there's two broad strokes. The first one is the nitty-gritty data analysis. Getting stories out of the noise that comes with a bunch of data and finding that clarity to figure out, like, what is my story in this data? And then the other area is*

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

*sort of like data presentation, and that includes data visualization and everything that you have to do to get that story that was so awesome that you found online, so people can sort of -- you just show people the awesome.*

*Man: I actually think we should have a broader definition of data journalism and think about how products that journalists create can be data in order to then reformat, reuse, and rework that into different—different types of stories entirely.*

*Man: I'm thinking about it lately in terms of experiencing information, not so much kind of poring through it all and studying it, but being able to as much as possible step into it, which in the 3D VR space, which I've been working a lot in, in the last year, I think there's a lot of potential there.*

*Man: It's about either creating new data or finding data that someone else has created and finding stories within it and using them as evidence to tell your story.*

And that's it. Thank you very much. And thank these rock stars who helped us with this project.

[Applause.]

**Joshua Scacco:** So, good afternoon, everyone. My name is Josh Scacco. I'm actually an assistant professor at Purdue University. And this research was part of the Engaging New Project that is housed here at the University of Texas. We had a unique opportunity here last year, which was we got to sit down and listen to some of the top minds in digital journalism talk about what they do. And out of that, out of those discussions and those talks, we sat and we listened, came some of this research about digital divisions and understanding organizational gatekeeping in the context of online news.

So, a little bit of background information about the Engaging News Project. The Engaging News Project provides research-based techniques for engaging audiences in a democratically beneficial and also commercially viable manner. So, we see ourselves as the bridge between what many consider to be two divides in the newsroom. And we see ourselves as an intermediary to understand, how can we not only come up with practices, but also tools for newsrooms to use to help bring these two aspects together?

This research is generously supported by the Democracy Fund, Hewlett, Google, Rita Allen, as well as the Moody College of Communication here at UT. As I previewed to you last year, in both Austin and also in Princeton, New Jersey, we brought together groups of online news personnel at some of the largest news organizations and also some of the key regional news organizations in the United States for a series of weekend workshops.

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

And we asked pretty basic questions. We asked things such as, are you concerned about political polarization? What do you consider to be good metrics? What kind of metrics would you come up with in an ideal setting to understand how your audiences are engaging with your news product?

And we looked at the data and realized that there was a lot of richness here in understanding not only gatekeeping processes, in terms of what bits of information become the news, but also in terms of how the news organization itself, particularly related to online news, is not only an opportunity generator for digital news, but also in a lot of ways still a challenge in understanding the modern news environment.

So, our particular lens for this was the hierarchical influences model and understanding how things like the individual journalist routines, social institutions, as well as the organizational level of the news organization, ultimately has an influence on what becomes the news. Particularly in the academic literature and especially related to online news, there's not a lot of information about how that organizational level is ultimately leading to particular online news outcomes.

And when we talk about the organizational level, we're meaning things like the structure of the newsroom and the organization of staff, how resources are allocated, [and] how learning occurs among journalists. So, we looked at those particular things in relation to this research. And we were inspired a lot by the late political scientist, Tim Cook, who not only envisioned the news media as an institution and talked about it, but then shortly before his death, [he] talked about the importance of researchers not being too tied to this notion of newsroom homogeneity—that really we should be testing assumptions in terms of, where are the limits of the institutionalization of news?

And so we look at this, and this is also important for us as researchers, because we are working with news organizations. And if we are going to bridge those commercial and democratic games, we have to meet news organizations where they are. And so, that's important for understanding where the limits are of academic theory, as well as where our assumptions can come into the play in terms of research going forward.

So, we brought together, as I said, representatives from newsrooms across the United States for a series of two-day workshops last year. And [there were] a couple of big themes that came out of this. We noticed that there was a lot of talk about resources. No big surprise. But interestingly enough, when we looked at how resources were being talked about, resources were being talked about not only in relation to within the newsroom, what's going on within the newsroom, but also resources that exist outside the newsroom. And I'll get into that in a second. We also saw a second big dimension, which was how news organizations are engaging in a socialization process;

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

particularly, digital news within the newsroom, but also outside of the borders of the newsroom as well.

So thinking about these things, when we looked at how resource allocation is occurring inside the newsroom, a couple of the key areas that we noticed were that resources still matter. Despite the fact that these tools for collecting data, these tools for engaging in digital journalism are quite cheap—and we've heard it here at this conference that these tools are quite cheap—that there are other resource considerations that have to be taken into play.

For example, once you have the data, who's going to look at it? How are they going to look at it? How is the newsroom going to be structured in a way to look at it? And these are resource intensive practices. This also gets into what I will talk about in a little bit with socialization as well. But importantly enough, things like understanding data, AB testing of headlines, which some of our news organization representatives talked about, were met in a lot of ways by other participants with surprise that news organizations had resources to, for example, AB test 50% of the headlines that were going up on their website. And we noticed here that there was a lot of divergence in terms of how resources were being allocated in the newsroom for these particular things.

Another area that we looked at was resources related to external, outside of the newsroom. And here, we looked at two particular things. First is that surprisingly, and this shouldn't be surprising really in terms of the competitive environment, that news organization's digital news is looking to their competitors to see who is the leader, who's the industry leader on digital news practices? And they're learning from their competitors in terms of their best practices, in terms of their mistakes, and implementing them in their own newsroom.

Also interestingly enough, the audience itself also becomes a key resource in terms of understanding, for example, how do you moderate comments? So, some of the newsroom representatives talked about the importance of bringing in citizen commenters to help with moderation of comments.

So, this was the first big dimension that we looked at. And I'm happy to talk about it more in Q&A in terms of some of the details that came out of the discussion. But the second big area that I think often gets lost in the process is, how are people learning this process? And at the organizational level of news gatekeeping, socialization is important to understanding how best practices are carried forward [and] how outdated practices are stopped. And particularly, when we think about internal socialization, we're looking here at a couple of different perspectives that emerged in terms of the structure of newsrooms.

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

So, some of the newsroom representatives talked about the fact that they incorporated data analytics teams, where they either had one data analysis team analyzing the data or they had several different teams analyzing the data. Interestingly enough, you could say, "This isn't a bad idea. It's quite efficient." One of the downsides potentially to this could also be a compartmentalization of knowledge— that ultimately newsroom teams might not be sharing with other teams the benefits that they're gleaning, the metrics that they're gleaning from what they're seeing in terms of audience data, compared to some of the other flatter designed organizations.

And we had several digital representatives as well, who engaged in a flatter organizational design, where smaller newsrooms could meet with each of their individual contributors and their journalists to talk about their data analytics and figuring out how they could also increase the breadth and the width of the benefit that they were getting from their product.

The second area that we looked at related to socialization was socialization externally to the newsroom. Interestingly enough, we actually found something here that I kind of consider to be norm setting. So if we think of agenda setting as issue setting, norm setting here, in terms of, how do journalists model the behavior that they want outside of the newsroom when it comes to engagement?

So, things such as comment sections. How do news organizations reach out to the public to engage in the comment section in a particular type of manner? And we're seeing that community engagement is becoming very important in terms of this new dimension of newsroom engagement with the public.

And this is very much related to some of the other research that the Engaging News Project has done related to comment moderation. Finding that, indeed, selling at norms that are inside a newsroom and inside a comment section can be replicated when commenters, when contributors see particular behaviors being enacted in a comment section.

So, a couple of big conclusions is that, really, the homogeneity hypothesis, this notion that, yes, newsrooms are similar, yes, there's a broader institution, but there's also a lot of divergence that we have to as scholars, as researchers, as practitioners understand and appreciate that the news is fragmented, partly based on these organizational perspectives that organizations are developing different socializing techniques in relation to the resources that are available, and that these are giving variety to the perspectives of the packaging and dissemination of online news.

So, I look forward to your comments. and thanks so much for your time today.

[Applause.]

## 16<sup>th</sup> Annual International Symposium on Online Journalism

### Q&A Session:

**Amy Schmitz Weiss:** Great. Well, I think this was a fantastic panel this morning to look at exactly what is changing, continually changing in the area of journalism now, from looking at the aspects of how data journalism is really transforming how we think about the field and the stories that we tell, to looking at exactly how these new changes in the newsroom are having an impact externally, internally, and how they operate, to looking at crowdfunding, to then looking at how much online commenting is also playing a bigger role in terms of how we understand the dialogue and conversation that's happening in these digital spaces.

So, I had a couple of questions for each of you that I wanted to ask. First, starting off with Nick, with your research in particular. I wanted to ask you about this, because I know this was brought up yesterday with Fiona's piece at the end of the day yesterday. And when we look at what's happening with online commenting, how much do we feel that there is currently an echo chamber that's happening within these spaces? Where do you think online commenting may be going next in terms of looking at how these different systems are created? And are they continually providing that kind of echo chamber or are they going into something else in that regard?

**Nicholas Diakopoulos:** So, I think this is actually one of the things that moderation can really help with, right, is sort of breaking out of that echo chamber. And diversity is one of these criteria that's mentioned by professional journalists as important in selecting for high-quality comments. And I think that the challenge is, diversity amongst what? Right? And it can be difficult to understand exactly what someone's position is or what their background is based on, you know, a single comment. And so, one of the things that we're developing at University of Maryland is a technique for actually data mining a person's history? So, if someone's left 2,000 comments, and there are people who've left thousands and thousands of comments at *The New York Times*, if you can go back and data mine everything that they've ever said, can we start getting a better sense for who they are [and] where they worked? Are they male or female? Maybe we can even characterize their age and these other kinds of demographic factors. And then provide that as a sort of visual input for moderators to be able to select more diverse opinions amongst the different dimensions. And so, you know, we're basically trying to build the tools that enable these types of conversations that we think are valuable.

**Amy Schmitz Weiss:** Soraia, I've got a question for you in particular. With the project that you were looking at in Brazil, specifically I wanted to ask you about the crowdfunding aspect to it. As we heard yesterday from El Español, the startup yesterday morning, and the work that they've been doing and jumping into and looking at a different model for their crowdfunding, which was equity based, I wanted to ask you, kind of, are you noticing in Brazil, in

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

particular, other kinds of models along those lines for crowdfunding in Brazil? And particularly with this news organization, news group project that you were looking at, how are they planning to identify different kinds of crowdfunding moving forward?

**Soraia Souza:** Well, crowdfunding in Brazil is starting. They don't use it for journalism. For example, it's.... I was talking with some researchers from Brazil here. So, there are tools that aren't well known. For example, Storify. And crowdfunding here, we heard [about] some projects, but in Brazil, crowdfunding is not a good way to get money in order to do projects in journalism.

So, this specific project, *Arquitetura da Gentrificação*, we tried to use crowdfunding. But as you can see in the article, I don't know if I was so clear in the presentation, it wasn't enough to provide financial support. So, it's something we have to work it out in Brazil. It's not a good way to provide financial support for this kind of project.

**Amy Schmitz Weiss:** OK. Next question for Cindy and Dale, in particular. Looking at data journalism and how you came to the idea of creating the conceptual definition for it as a process, where do you see the human element in that? I noticed you incorporated the public into the definition, but I was curious to understand a little bit more where you see the human inquiry part, the human at the end of that data point, and where you see that falling into the conceptual definition that could also lend itself into future operationalization of variables in that regard?

**Cindy Royal:** Yeah, sure. I mean, I think you identified the area that was sort of, who does it influence on the outside? The individual? How does it affect democracy? And being able to look at the elements of user interactivity, personalization, I mean, that's all things that provide benefits to the user in the process of using these tools. So, a lot of times we'll be looking at this from the perspective of, well, who's doing this work? I mean, that's what a lot of the work has been so far.

I did some research on some of *The New York Times* data visualization projects; some of the actual products. But we haven't done a lot with how people are engaging with this. And I think that individual area would be like a sort of area to kind of target to be a jumping off point, to be able to do more research about, you know, do people really get something out of this? And are there different measurements and different types of impact that we need to be aware of for somebody engaging with a story in this manner? Do you want to add anything?

**Dale Blasingame:** Yeah. I think the best answer is in the video. [It] came from Rodney at *The Tribune* and Robert Hernandez, who talked about making it as easy as possible to digest these huge data stories. So, it may be for the



## **16<sup>th</sup> Annual International Symposium on Online Journalism**

people at home who are actually consuming it to understand the story and get something from it.

**Amy Schmitz Weiss:** OK. And the last question is for Josh. Sorry, I said you were from UT.

**Joshua Scacco:** That's okay.

**Amy Schmitz Weiss:** I know you're from Purdue. So, my apologies for that. [laughs] So, I was thinking about this as you were presenting. I was inspired by Jay Rosen this morning during the research breakfast, from his question there of looking at when we think about culture and how much by which changes are not happening can actually influence change that does happen. And so, I wanted to kind of get your perspective on that in terms of all of the work that you and your coauthors did in looking at -- noticing that there's these particular aspects of socialization that are occurring that are changing, and some that are not, and the internal and external resource allocation that's going on as well, and what your thoughts are in terms of, in some cases, things that may not be happening are actually influencing things that can be changing, just in a different way or in a different process.

**Joshua Scacco:** I would think that based on what we're seeing, there's still quite a bit of difference across newsrooms, in terms of some newsrooms are taking examples of, you know, experiences that they have and learning from them, and other newsrooms are still very much trying to negotiate the traditional digital divide. And it's really telling, I think, when you get those moments, especially up here during the panel yesterday, where you have some of the biggest news organizations in the United States still talking about the fact that they still don't get it or they still haven't gotten it necessarily within their newsroom. And so, a lot of it is very much cultural. It's very much organizational.

It's easy to say, "Here's a tool. Use it." It's easy to say, "You can buy it." But at the end of the day, as I said, if no one has been trained to look at the data, if no one has the skillset to run an AB test, or even if individuals have, are they sharing the knowledge in a way that is beneficial to the entire newsroom and the entire news operation? And I think in those manners, there's still quite a bit of richness and a lot of room for learning.

**Amy Schmitz Weiss:** Thank you. So, we've got a couple of folks that have questions. So, we're going to start over here with Mark Coddington.

**Mark Coddington:** Yeah. I have a question for Cindy and Dale. I really liked the study. And I'm curious. I think the response that intrigues me the most is the data journalist who said, "There's nothing new here. We've always been doing this. Been doing this for a long time." And I'm curious, I guess, (A) the degree to which you agree with them. Do you think there is anything substantially new here? And I guess I'm wondering, do you think there's

## **16<sup>th</sup> Annual International Symposium on Online Journalism**

anything sort of culturally or professionally going on there with journalists? Like, why would they answer that way? So, I'm just... I guess I'm curious without trying to tip my own hand [chuckles] and make a speech. [laughter] So, I'm just curious as to if you have any idea as to what's behind that and the degree to which -- do you think they're right?

**Cindy Royal:** I mean, I think it's pretty clear that there are elements of data journalism that may be completely the same thing as journalism, as I saw in many of these assertions, where we're telling stories, we want to have impact on people's lives with the ways we tell stories. But there are some key differences to the product that actually gets developed and the features that are within it that not every journalist is able to do. It's a different way of using data than what we've even done before with computer-assisted reporting, because now we're actually making products that the users can interact with and customize and personalize. We talked about personalization earlier today and how important that's going to be to people.

So, I think that there are cultural aspects on both sides of that question. I think the person who made that response in this, they are really wanting data journalism to be kind of embraced as journalism. They want their time in the sun and to get their credit for what they're doing as journalism. It's not this weird techy thing that somebody's doing in a corner.

And then I think that there are people who want to define it as its own competency, so that it also gets the credit it deserves for being different and unique. So, like everything in this paper, it's a jumping off point for a lot more research. I think that this is an area that is just ripe for lots and lots of new study and research, so we can identify, you know, what we need to know about the processes [and] what we need to know about these products. And we have so many opportunities for people to do these jobs to be part of that.

I think another part of the culture is at the end of my paper, I get to the parts where they talk about journalists don't like numbers. We don't like math. So, I think there's that sort of cultural clash with that, and that has to be overcome in certain ways.

**Dale Blasingame:** Right. And I also think the journalists who are willing to try new things, to them, this is just another example of trying something new and still telling stories and still doing the work that they need to do. So, the ones who might view it as not journalism are the ones who are scared by trying new things.

**Amy Schmitz Weiss:** So, we've got a question over here. Trei.

**Trei Brundrett:** Hello. So, this question is for Nick about your research. Well, first of all, all the research was awesome. Thanks for all of that [and] presenting it here. But there's something that's like really interesting in this

## 16<sup>th</sup> Annual International Symposium on Online Journalism

research about comments, and I think it's particular in news organizations, where there's this focus on sort of like kind of in a vacuum about quality of the comments. And I think that's maybe somewhat related to this debate about whether or not news organizations should have comments at the end of their articles. And it's sort of wrapped up in that. There's certainly lots of Twitter chatter happening about that right now behind you.

**Nicholas Diakopoulos:** Whoa! [laughter]

**Trei Brundrett:** My question is actually about what I feel, like, because of that. It's strange that the.... And I'm curious if this comes up in your research, is that it sort of seems to be outside the context of community, right? Like, we talk about comments are just like talking about the article. And how quality are they? All these different factors that you would.... Maybe it's relevance to the article. But in terms of community and reputation and identity, I feel like we've sort of gotten away from what was really promising and exciting about virtual community around content, was that we were going to kind of build these communities that provided context and quality in that way. So, did any of that come up in your research?

**Nicholas Diakopoulos:** I mean, I think we're living in challenging times for building community online in the sense that content is so scattered across social media, and we encounter content in so many different ways and places, and come into publisher sites from so many different angles. You know, how many communities can I really be part of? There are certainly commenting community that exist on news websites. And I think to a large extent, they exist because they are cultivated and because, in many cases, the community is there, and the online version of that just happens to support the communication between people in that community.

I mean, I think those communities can be cared for by moderators. And I think it takes a delicate touch in some cases in making sure that it's a safe place and so on. But I think that this has to be something that news organizations want to invest in. They have to put people—paid people—employees to be there and to cultivate those kinds of relationships.

And I think ultimately, there can be a lot of value there. There's tips coming in. there's loyalty that's built. You know, if you want to build a nice subscriber base, hey, why not create some loyalty and some community around your content?

**Man:** I want to thank all the presenters, too, for staying on time. It really was fantastic. Thank you. So, there was a great back channel going on, coming out of a comment that Nick made about Cindy's presentation. And I wanted to sort of just bring that to the front and see if you guys could talk about it a little bit on stage. So, Nick commented about the place where the interviewing data metaphor might break down on Twitter, and a lot of people have been talking about that for the last ten minutes or so. So, Nick, if you

## 16<sup>th</sup> Annual International Symposium on Online Journalism

want to put your question out there, and then, Cindy, if maybe your respondents actually had something to say about that.

**Nicholas Diakopoulos:** Yeah. So, I think this is sort of a popular metaphor. I've heard it several times. You know, it's like, you have that data journalists interview data. And I wanted to push back on that a little bit, in the sense that I don't actually think that it's -- that the metaphor is complete. We don't actually talk to data in a way that it can talk back. Data is inert. It doesn't respond to you. It doesn't have agency. In fact, the agency is in the asker. I decide the question I want to ask and how I want to ask it, which dictates how the data -- what interpretation I gain from that data. And so, it's not really a dialogue so much as it is a monologue with some kind of object that you are kind of peeling back and trying to understand and interpret. So, that's sort of my slight semantic tweak on the idea of interviewing data.

**Cindy Royal:** I think that probably came up in some of the interviewer's comments, where they talk about interviewing the data, right? Yeah. And I think that that is the way that they understand it, because they want to view data as a source, so they want to kind of treat it like it's any other source. But I agree with your comments that it's a different relationship that we have with data than what we have with a human source.

But as you know, and anybody who's worked with data knows, there's a lot of work that has to be done once you get a dataset to get it in some sort of a fashion that you want to present it. And you have to spend a lot of time with it before you can even understand what some of the questions might be. So, students in my class know this—you don't necessarily grab a dataset and start answering questions. You have to kind of work with it, massage it over time, see what emerges, see what's weird, and then you have to explain that. So, you have to go out and do more research to figure out why this is the way it is. You can't just take a dataset and be like, "Cool. I'm just gonna make a chart."

So, I think that's what they are getting at. But it's not a perfect metaphor, yeah.

**Rosental Calmon Alves:** It's not a perfect metaphor, but it is a great metaphor. And it has been very useful for journalists to explain to journalists what computer assisted reporting was at that time and what data journalism can do now. So nobody—nobody ever used that metaphor believing that they the data would talk back, other than responding to the smart questions that are created for the response [that] comes from the data provoked by this. But the important thing of the metaphor was to say, first, interview the data. Interview people, but first, interview the data. And the importance of this was also to tell that it's not enough to interview the data. It's not enough to do data journalism based on only data. It's part of a more complex process, etc. So, leave this beautiful metaphor alone. [laughs/laughter] I love it.

## 16<sup>th</sup> Annual International Symposium on Online Journalism

**Cindy Royal:** I feel if Rosenthal gets up and makes a comment, then you've really like sparked some provocative item. [laughter]

**Nicholas Diakopoulos:** We can chat more later, Rosenthal.

**Amy Schmitz Weiss:** We have another question over here.

**Frank Russell:** Hi. I'm Frank Russell. I'm from the j-school at University of Missouri. My question is for Joshua, but it's sort of prompted by Nick's presentation and Cindy's presentation. Do you think, did the journalists that you interviewed, did they report that their gatekeeping practices were affected in any way by the technologies they used, or the people who were responsible for providing and maintaining those technologies, whether they're from within their news organization or company, or whether they are external?

**Joshua Scacco:** So, some of the discussion did turn to some of the platforms from where all of this data was coming from. In terms of talk about the other people on the other side of those platforms, there wasn't a lot of that. It was more, you know, once the newsroom is getting the data from, you know, one of seven platforms, one of eight, how is the newsroom going to deal with it? So, that's actually a really interesting point in terms of the maintenance of those particular platforms, because the technology itself will ultimately determine what is coming into the newsroom. And that's an additional screen on the process that we don't necessarily consider in the paper, but it would be a potential gatekeeping source outside of the newsroom as well that would influence it.

And I think this is kind of getting into a lot of the discussion that's emerging up here, which is, the lenses that ultimately we have to still consider for the final news product; that ultimately the data coming in is only as good as the people interpreting it. It's only as good as how it's being interpreted, the lenses that are being brought to it to bear on it, and in addition to that, how the technology itself is set up and how it's maintained.

**Dale Blasingame:** So, he has a point.

**Frank Russell:** Yeah. And not to bring up the metaphor again, but when John said you have to question the data, that means both talk with it and actually question its validity. If you ask horrible questions, then you get bad data. So, I think that gatekeeping role is still there. We're responsible for telling good stories and making sure the data is reliable.

**Amy Schmitz Weiss:** OK. Well, a big round of applause for our panelist for their research.

[Applause.]