13th Annual International Symposium on Online Journalism

Day 2, April 21, 2012: Morning Session - 9:30-11:00 a.m.
The New Narrative: How data is changing the way we tell stories online

Chair: Aron Pilhofer, Interactive News Editor, The New York Times

Panelists:
- **Brian Boyer,** News Applications Editor, Chicago Tribune Media Group
- **Alberto Cairo,** Lecturer in Visual Journalism, University of Miami
- **Alastair Dant,** Lead Interactive Technologist at Guardian News and Media, London, UK
- **Angelica Peralta Ramos,** Multimedia Development Manager, La Nación, Buenos Aires, Argentina
- **Ben Welsh,** Database Producer, Los Angeles Times

Q & A: Aron Pilhofer and the Panelists

Aron Pilhofer: OK. So, I'm Aron Pilhofer. I'm the Editor of Interactive News at the New York Times. My God, this room goes up high. [laughter] This is a very weird, almost…. OK. So, Rosental asked me to come down. This is my second time to this conference. [gets PowerPoint set up] OK. So, this is the second time I've been here. And my first time here was — what was it, three years ago? Four years ago? It was a while ago. Three years. And at that point, we had just sort of started this team that I run now at the New York Times, which I call Interactive News. The team has now grown to about 14 people. At the time, it was about four. What we do is a radical departure, I think, from, at least in my experience, what I used to do, which is straight-up print reporting. Data is part of the reason why: the ubiquity of data, the availability of data, and the need to bring data and data journalism into the mainstream of newsrooms.

So, that's kind of the purpose of today's panel is to look at the spectrum of how newsrooms are using data, how data is changing the way we tell stories, and how the availability of data is providing new avenues for journalists to do different kind of work and actually changing, I think, the very definition of who a journalist is, as will be evidenced by my co-panelists who will be up here in a moment.

So, when we talk about data, we kind of think of this, right? A spreadsheet, columns, rows. How many of you just went to sleep? [laughter] Probably a very large portion. I will admit to you that you are probably not like me. I see columns and rows like this and the hair on my arms stand up. [laughter]

I couldn't get more excited by this. But I will admit I am not like you or most normal people. But let's think a little more broadly though. I want to change how we think about data. I want you to take that spreadsheet right out of your brain and think about data like this. Paul Bradshaw, who is an academic in the UK, [defines it]. I think, to me, this is the absolutely best definition of data out there. And basically it's this: It is information that can be processed by computers. Full stop. Full stop. Now, I'll give you some examples of how changing your thinking about data can impact journalism in a second, but this is my working definition for today.

So, you've seen things like this. This is our house map. This is from a wonderful work of our graphics team and my own team doing the data behind it. And this you can look at and say, "All right. That's data." But what about this? How many have seen this interactive graphic before now? If I had…. Yeah, I'm getting the…[chuckles]…from the audience. I don't have an Internet connection here, but do yourself a favor—and don't do this now, please, because there is sound involved—but Google this: Fractions of a Second: An Olympic [Musical]. To my mind, this is the most amazing data-driven graphic I have ever seen. This is Amanda Cox. And I had nothing to do with this, by the way. I'm just a massive fan, borderline fan boy of her work. And to take the problem of telling a story driven by data and turn it into something like this is truly amazing.

But data can also be this. This, I don't know if my friend Alistair is going to talk about this or not, but everybody's tried to use other sorts of, like, we'll call it non-traditional forms of data. In this case, Twitter. Everybody's tried to things with Twitter. Most of them stink. This does not stink. This is using the idea of a social network and bringing it into a news context and actually telling a story with it. And how did they do that? What they did is, there was human intervention here. They got all these tweets around the London riots, and they wanted to study and answer one simple question: How did rumors propagate across social networks? And how were they corrected? And they had to do that in a very both complicated technically [way] but also very simple way—they asked the right questions: What are the rumors? Are animals actually being let out of the zoo? Are they running wild on the streets of London? And how did that propagate across Twitter? It's an amazing interactive.

Data also can look like this. This is PolitiFact started by Matt White and Bill Adair at the St. Pete Times. Here, you're taking a very traditional form of journalism. This is basic watchdog reporting. This is fact checking. And you're blowing it up and you're rebuilding it into the web, into a web context, into a digital context, where every promise is a data point and every promise kept is a data point. And you can then go through here as a reader and find out what has Obama kept, what has he not kept, and keep track. This is an interesting approach. This also can be data.

This also can be data. Politics Verbatim. This is a project started at California Watch, which is doing some incredible data work these days; particularly, the project they did around earthquakes, which if you haven't seen it, you should look. This was an interesting approach, because it takes something that is not structured data, in this case, what politicians are saying—this is the Jerry Brown [and] Meg Whitman race from a couple of years ago—and where they are saying it. So, it's taking what they're saying on one axis and putting it on a second axis, where they are saying it. So, what Jerry Brown says in Northern California about immigration versus Southern California. What Meg Whitman says…. You know, you get the idea, right? Interesting idea. And there's more of that.

This can be data. This is the Word Train. This is a project we launched in 2008 at the Times. Gabe Dance, who's now at the Guardian, and myself worked on this. The idea here was to crowdsource, essentially, a news story. Anybody [who has] every covered politics knows that the day of an election is probably the most news free day of the election, right? [laughter] OK. A few of you laughed, and those are the political reporters out there, because you know on election day there's absolutely nothing to report. Right? The politicians show up, they kind of wave at the camera, they go into the voting booth, they wave at the camera again, that's it. That is what we have to report. But you also get this sense that there's this like kind of collective holding of breath, right? Because you know that as soon as those polls close, you're going to start getting some results, and then the news is going to happen. So, Gabe wanted to capture that in a kind of interesting way, and this was his idea. Just a simple question: What one word describes your mood right now? And then at every hour, we snapshot it. We asked, "Are you a McCain supporter or Obama supporter or neither?" And as you can imagine, you can imagine how this worked, but it turned into a really interesting narrative. And as the readers sort of filled in the data points for us, you could see about, you know, obviously, six, eight o'clock, how these things diverged.

And finally, I love this project. It's been an ongoing project of the L.A. Times for a very long time. It's the Neighborhood Project. I like this one in particular, because the idea of a neighborhood is kind of a state of mind, right? It's not a defined thing per se. And they wanted to sort of capture that. They wanted to gather that data, so they sort of crowd-sourced it. They said, "Hey, readers, here's what we think the neighborhoods around Los Angeles are. But you can take this map and you can adjust it and you can tell me where you think they are." And then were able to overlap these different boundaries and sort of figure out where people thought they lived versus where the so-called boundaries are. It's an interesting idea. Interesting way to tell a story.

So, those are just some examples. Again, I just want to leave you with this idea that throughout this panel, we need to sort of expand our minds a little

bit about what we think of both as journalism, as data, and who journalists are. And so, with that, I'm going to turn it over to Brian.

[Applause.]

Brian Boyer: All right. These slides are timed, so work with me. I'm Brian Boyer, News Applications Editor, Chicago Tribune. Now, I'm going to wait for the next slide. [laughter] All right. So, this is a pile of sunflower seeds, but they're not sunflower seeds. They are in fact made of clay. Each one was handmade and decorated. It's a beautiful, moving piece of art [that] I find incredibly inspiring and required great craftsmanship and skill. These on the other hand are cups. They are also made of clay. There were also handmade and hand-decorated by my friend Emily, and they require great skill. They may even move or inspire you, but they have something else going for them, which I'll get to in a moment. All right. So, I've worked in clay since I was in high school. It's sustained my spirit through many shitty jobs. But a lot of what I learned in the ceramic studio informs what I do today, so I hope you'll find what I'm saying interesting.

All right. So here are two beautiful mugs. I could talk for an hour about how they were made, the clay body they were made with, the throwing technique, the glaze chemistry and how it interacts in the atmosphere with the kiln, and the total insanity of the people who made these damn things, as evidenced by their textures. But look closely at the handles. On the left, the handle is angled so your finger supports the mug by the handle. On the right, the handle is round and the side of the mug rests against your finger. Now what happens—guess—when you fill these mugs with coffee? Both of these objects are useful. Both required skill and craftsmanship. But goddammit, when I use a mug, I expect to be able to drink hot liquids! [laughter] The blue mug is art and craft. The brown mug is merely art unless you're drinking iced tea. [laughter] So, a mug must be useful. Craft is a step above and beyond art: where your finger rests, how your lips meet the lip, the heft, the comfort of the handle. Usability is required for craft.

All right. So, if you think powders are sexy and you'd like to bag one, find a powdery booth at a craft fair. [laughter] Lots of people will pick up a mug. They'll feel its weight. They'll contemplate a cup of coffee. But you will know the potters when they flip it over and they ponder the foot. Because it's the foot where the craft is revealed. A course or sloppy foot is an indicator of poor quality. Ugly pots invariably have bad feet. But there are also a lot of beautiful, artful objects that will scratch your coffee table.

So, you're probably asking yourself, why the hell is Brian talking about mugs at a data journalism panel, right? And I'll tell you why. It's because what we do is a craft. News applications should be made for use. Used by people you care about, not your peers, not an awards committee, your audience. Now data visualizations are not on their own useful. They are art. Now, I like art. I

made this damn thing. But I believe if we only make art, we are doing our audience a disservice.

How many color blind people are in this room? Raise your hands. Not too many. This map sucks for those people. [laughter] All right. So, make your data beautiful. Make your news applications moving. Make them inspirational. But for frak's sake, please everyone, make them useful! [laughter/applause] All right. We're still going. Journalists want to tell interesting stories. Designers, they all want to make beautiful things. And engineers want to solve big, interesting problems. And we all must be stopped! You must have a process that controls for your urges. Because it's so easy to get carried away by a fun idea, right? Our team has a simple exercise we run at the beginning of every project. We ask, "Who are our users? What are their needs? And what can we build to fulfill their needs?"

Example one. Wait for it. I have nursing home data. Who are our users? People with loved ones in nursing homes. What are their needs? They just read a scary story about nursing home safety, and they want to know if their grandmother is safe. What can we build? Nursing home safety reports.

Example two. I have fire safety data. Who are our users? People who live in high-rises. What are their needs? They just read a scary story about fire safety in high rises, and they want to know if they're safe. What can we build? A thing that lets them look up if they're safe. (My mom loved this thing.)

Both of these apps had geodata, right? We could have made awesome maps. Both of these apps had time-related data. I could have made fancy, slick, crazy timelines, right? They would have beautiful. It would have been moving to see all those unsafe places scattered around a map. But no, goddammit, that's just not useful! Don't do it!

The written story will speak to the trend. You'll include three anecdotes of people who died in fires or got knifed in a nursing home, but what are the— that's not funny—[laughter]—what are the readers to do when they're finished, right? Our work sings when it helps readers find their own story in the data. How this issue impacts their own lives.

So, I hope you'll take this information home. Craft is useful. Fight your urges. Know your audience. And please make useful stuff. And also, show your work.

[Applause.]

Alberto Cairo: Hello. I'm Alberto Cairo. I would like to thank Rosental and the University of Texas for having me again. This is the third time, I believe, that I speak at this conference, and it's always a great experience. Actually, my short presentation is going to be like a follow-up to Brian's, because

some of the points that he made are really interesting and really important and I really care about them. Which is that for making a visualization or an information graphic appealing to read, first of all, you have to make it useful. One of the core ideas in my courses at the University of Miami, where I am teaching right now Information Graphics and Visualization, is that an information graphic is above all a tool, a tool for understanding, but it's also a tool for presenting information and also for exploring that information. Those are the two sides of information graphics. And there are certain trends, certain trends in information graphics and visualization that worry me a little bit today.

In the past, information graphics used to be mostly about editing information down, so you got a lot of data, a lot of information, and then you try to cram it into a presentation. You know, you dumb things down. You just edit it down. That's an example. That's a very recent example. Because that trend of dumbing data down, it's still in very good health, and that worries me, because this is the presenting side of information. This is just presenting information to readers. But we are not allowing readers to explore the information, to explore the data behind all those graphics that you are creating. So, this is a traditional kind of information graphic that I had to deal with when I was working in newsrooms.

Today, we have the opposite trend, which is basically that—and this is gaining momentum, as I say there—that because data is increasingly available and we have the tools to handle the data very easily and to process it, we end up having things like these that are basic. And this is an extreme example, which is basically data art. It's not data journalism. It's art. It shows a connection, the cross-references in the bible. And don't get me wrong. I think that this is extremely beautiful, but it has to do with what Brian said before. This is not really journalism, because it doesn't really help you understand anything. It's just a beautiful presentation of data. It allows you to explore the data if you know what the data is about beforehand. So, if you can bring some previous knowledge to the table, you will be able to understand what this is. But as long as I am not very — I am not a very good reader of the bible, I don't really know what this thing is about. I don't really care.

So, but we see this in the media. Today, if you go to the New York Times, for example, you will see things like this. This is an extremely beautiful, extremely nice presentation. And my first reaction when I saw it.... It's called Mapping America. It's basically getting all the data from the recent census and just cram it into a beautiful map built on Google Maps. This is very beautiful. And my first reaction when I saw it, it was that, you know, this is extremely cool. And yes, I started exploring it, etc., but after five minutes, my reaction was, so what? What is it useful for? I just wanted to look for my neighborhood. I saw the data for my neighborhood, but it didn't allow me really to compare with other neighborhoods or to find the structure of the data. It didn't allow me to rank neighborhoods or to compare one thing to

the other. It didn't tell me anything beyond what I was seeing about my neighborhood. I didn't get any story.

So, in some sense, there are many designers out there and also many programmers that are forcing readers to become their own editors. And journalism is about reporting. We are doing data reporting. But in data journalism, I believe that the editing process is even more important than the reporting process; particularly, in these times when data is increasingly available. So, we have to become editors in some sense. We should not just show readers data and throw data to the readers, but beyond that we should extract the meanings from those data and explain what those data mean. That's what journalism is about in the future. We have to help readers somehow.

And we have to create layers in our information graphics. We have the presentation layer and we also have the exploration layer. The previous infographic was all about exploration. You just throw those data to your readers and you let readers figure it out. You let the readers navigate. That's exploration. Readers extract their own stories from those data. But we also have to present some sort of summary of that data beforehand, and we have to explain to readers how to navigate that presentation beforehand.

We also must embrace complexity. And this is something that all time infographics artists don't really understand—that complexity is something useful sometimes. Showing all those data, the complexity of the data, is very useful and very interesting, but you have to arrange it in a way that the human brain can understand.

It's also appropriate to use novel graphic forms, interesting looking, you know, innovative kinds of infographics that you see in a computer science academic paper. You saw the other day and you want to apply it to your own work? Well, that's great. But then you have to think beforehand, is this something that adapts to the data? Is this something that adapts to the story? Is this something that will advance the story, that will get the message through? You have to think about that also as well.

There is a model or a mantra that was pushed out by Ben Shneiderman in 1996 called The Visual Information-Seeking Mantra. You will find it over the Internet. Shneiderman has written broadly about information visualization. And I think that this sentence summarizes what information visualization should be about and how to balance out the presentation layer and the exploration layer. [On the screen: "Overview first, zoom and filter, then details-on-demand."] First of all, you present a summary. First of all, you overview the data. What are the most important points? What are the main things that I should extract from those data? How [do] I use this application? That's the presentation layer. And once you have shown that to readers, once you have presented readers with the most important points of your data, of your story, then you allow readers to navigate the story and get to

the exploration side of the story. You let them zoom. You let them filter. You let them see all the details of the information. But you cannot have or you should not have one thing without the other. Both are equally important.

How to do that. Well, in the past, I'm not going to show this project, but this is a project that I made with my team when I was working in Brazil. I am in Miami right now. But until December 2010, I was working for Época Magazine in Brazil that belongs to O Globo, Globo Group of Communication. And I built a small data team and a small infographics team, and we put out a lot of very interesting projects, in my opinion. One of them is this. And this is a project. If you Google it, it's called Telephone Vermelho, right telephone. It basically summarizes how much Brazilian congressmen and congress-women spend a year on telephone. And those numbers are basically outrageous. Actually, when we published this infographic, it created a national roar, in some sense, because they spend so much. Readers were not really aware of how much money those people spend on telephone over a year. It's just a huge number.

But on this graphic, it shows you the two sides of data. It shows you the two sides of visualization. First of all, you have the presentation layer, the summary layer—how much those people have spent in total, and that's 13-billion reais, which is around $6-billion combined, all of them, in a year, which is a huge number. When you transform that number, and this also belongs to the presentation layer, when you say to readers, "Well, that's the amount of money they have spent through one year, in the last year. When you divide that by the average cost of a local phone call, a Brazilian phone call, well, it turns out that if you want to spend that amount of money talking on the phone yourself, you will pick up the phone today and you will end up your conversation 300 years from now." That's the reason why the headline, the main headline of the graphic is 300 Years of Phone Talking, because that's the main idea of the graphic. So, that's the presentation layer of the graphic.

And then after that, now the readers—and that's the part over there—now the readers [can] look for their own representative [or] for their own congressman. So, you can click on those balls. You can compare one congressman to another congressman. You can look for a specific political party or you can look for a specific state in Brazil to see just those congressmen. That's the exploration layer of visualization. But you should not have one without having the other. Both are equally important. And you have to find out what the right balance between those two layers is in your graphics.

So, thank you so much.

[Applause.]

13th Annual International Symposium on Online Journalism

Alistair Dant: Right. Hi, everyone. So, I'm here today from the Guardian in London. And what I wanted to do…. I'm actually going to just quickly say that I feel very lucky to have been sort of blessed with a surname that begins with D, because it means that in the alphabetical order I've now come after Brian and Alberto, who are actually kind of, I guess, sort of like covering a lot of the issues which this work that I'm going to show you touches upon as well. But I'm probably not going to talk so much about the sort of more philosophical aspects of it. I'll just show you the work, and I'll talk about three categories of sort of telling stories with data. So, the first thing is really plotting a path for events. And more to the point, I think it sort of coincides with the idea that you can actually make a visualization from the raw data directly and sometimes that data has a sort of emotional power in its own right.

So in this case, when we were covering the first WikiLeaks war log dumps, which is Afghan IED attacks. A discussion was had editorially. Julian Assange expressed a preference to give people something like a video, but giving people just a simple playback of many, many thousands of events in space and time was only sort of one part of telling the story. We wanted people to also be able to stop at any point on that timeline, add filters, zoom into the map, and actually get to the point where you could click on any single one of those individual events and find out the details of it. So, what we've got there is hopefully an encapsulation of some of the things that Alberto was talking about.

Here, down at the bottom, we've got this line chart which shows you the volume of attacks over time, so you get a sense that it's ramping up in addition to what's happening here in the chart. And we've also got a bit of an annotation layer as well there. And I think the idea of annotation layer is one of the places where this kind of editorial voice really comes out.

So, I want to sort of move on from this to a second example, which is, I think, characteristic of something a little bit different. We began a project with the World Cup in 2010. And I had this kind of notion in my head that I wanted to get the feeling of the roar of the crowd passing through Twitter. So, we built this interactive here, which I'm not sure how well you can see, because this kind of gray doesn't work so well. But basically we took a recording of all the tweets that occurred in every match, every minute of every match. We figured out what the most popular words were and then we created a playback where 90 minutes of a football game were turned into sort of 90 seconds of flying balls. And as these balls fall around, the places where kind of human energy is most present become reflected in the activity of the balls. So, there's something about the way that data itself is a record of something. But to make that record speak to people, sometimes you have to bring it to life.

Now, that alone is not the whole story. What we've got here is a follow-up project which we put together round about the time last year where the

Murdoch family was answering political questions. And what you can see here is the volume at the start of the day is quite pronounced. People are quite interested in what the Murdoch's are going to say. And it kind of gradually starts to tail off, because they're fending off all the questions that they're being posed quite effectively. So, as this sort of timeline progresses, you actually see that the words here start to drift away from what they're talking about in court, because people are getting bored. And then what we've added there on the left is a sort of sense of editorial annotation. We basically put in a kind of set of key moments that our editors had picked out. But here's the good bit. Here's a sort of money shot. [laughter]

And what I think is really important to just sort of note at this stage is that it's not only adding these annotations on the left-hand side that kind of makes this piece work. It's the fact that (a) we were lucky enough to be recording at the right time, and (b) we were then able to sort of make an editorial judgment as to exactly where it should start and end. And so that's almost a bit like editing video at that point. So, if you imagine that data is a kind of raw material, at the point where you're bringing it to life, you have to exercise the kind of editorial sense in terms of which bits of that videotape, which bits of that recording are most interesting to people.

[Here is an] example of how print graphics differ from online graphics. So, what you can do in print is you can give a sense of spread to something like this, and you've got an enormous amount of kind of—I can't remember what Tufti calls it—something like ink space, but basically you've got enormous amount of detail that wouldn't actually work if you tried to put it online all in one visual field. I think what Alberto was saying, which is really important, is the fact that when you create something online and you create an interactive interface, you can make that interface work in a kind of incremental way, so people can move through and explore—perhaps, it's sometimes in a guided way—and not necessarily have to deal with everything at once.

So, this is what we ended up putting online. It's basically something we called the Spending Challenge. And the idea here is that you can actually decide which cuts you want to make yourself. And you see a tree map here, which gives you an overview of the different departmental spending, but you can into any of these departments, and you can actually sort of see a rather kind of editorialized list of possible cuts. And then you can make a blanket cut here, because you really don't like the idea that we spend so much money on defense, so you cut half of that. And you go on and you go on. And what we also did for this project, which I think is actually a pretty useful thing for us to sort of revisit in the future, is we made it possible for people to share their cuts with others. So, we basically set a very simple kind of like encoding system that put all people's choices into a URL, which they could then kind of take this "arm-chair economist" pose and share with all their friends and say, "Here's how I would solve the budget problem."

And we opened up…. This is the way you do it, sort of down here. We opened up comments on this page. And what we found was that this actually triggered a debate. The ability to point people in at your kind of vision is a good way to stimulate debate. And I think this idea of creating a tool like this, which is almost a sort of like open-ended tool, it's something where you can kind of find your own path through the data and then you can share that. And I think that's potentially an interesting way that you can start to make these things work in the context of a public website, because you can basically invite people into a discussion. Anyway, getting to the kind of most interesting fact, nine times out of ten, people solved the 49-billion deficit by cutting Scotland, which is about 52-billion. [laughter]

So, this brings me onto this sort of main thing that I wanted to highlight today, which is a project that we've just started development on in January. It's called the MESO Project. And basically, what we're doing here is we're using some money from the Gates Foundation to develop a set of open-source tools. And what those tools will hopefully do is enable us to get to a stage where when we create an interactive piece, we'll no longer just be publishing it up in a way that sort of preserves our dignity, as the chap was saying earlier. We actually want to make the kind of code and the data that make the piece work public, and we want that to be a start of a conversation.

So, the idea with MESO is that we'll start using a standard set of tools to build all of our work, which we're also sharing with people, so that anytime you see…. In this case, we did a kind of remake of the cabinet office spending graphic using the new tools. And anytime you see a piece, assuming that everybody keeps their New Year's resolution of learning to code this year, [laughter], you can actually open that piece up and start making your own version of it. So, you know, obviously, if you don't live in the UK, you might have other spending office data to look at, and for that reason, I think that this is a sort of set of tools which could be useful to a kind of wide community. Because the more this data is opened up and surrounds us, the more we need a way of kind of engaging people in conversation about what that data means.

So, yeah, if you want to follow along with the project, we've got a sort of Twitter thing here, the MESO Project, and we'll be sort of sending updates as we put these libraries out throughout the year.

OK. Thanks.

[Applause.]

Angelica Peralta Ramos: I'm Angelica Peralta Ramos from La Nación in Buenos Aires, Argentina. We are in a very, very different context here. First, I want to talk about La Nación. It's based in Buenos Aires, Argentina with a Sunday print circulation of 360,000 that has grown the last four years at 4% per year. La Nación's unique dot-com visitors per month is eight million. We

have nine magazine titles, and we just bought 90% of ImpreMedia, that is a U.S. Hispanic leading publishing company. Most of you may know.

So, a little about Argentina. Transparency, you just say. We are positioned now 100 over 180 in the corruption perception index, so we have an issue with transparency too. And we think data can help. We have no FOIA, no Freedom of Information Act, but we have a decree. And I think this is outdated, because Brazil is now better than us. So, if we ask for information in Argentina, perhaps when it's answered, it comes in paper. We are not participating now in the Open Government Partnership that eight countries and 55 other countries are participating to work in open government and open data. And, of course, we don't have Data.gov portal. But there is hope. There are portals in Argentina that are opening. Data is opening from the activist's word more than from officials.

But these are…. Buenos Aires data is from the government of Buenos Aires. Our Judiciary System Open Government Initiative, too, we have. The World Banc has lots of statistics also from Argentina. And INDEC has census data and lots of data in good quality.

At La Nación, we found ourselves, we want to do data journalism, but we had no programmers in the newsroom, no team of 14 programmers, but we had a team that was willing to learn, and so, we moved on. And we just said, "What are we going to do? Are we going to sit back and wait for others to build datasets? No." We started doing it ourselves. We think that information is there, but we have to make an extra effort and convert it ourselves. So, we love PDFs. [chuckles]

We decided to start small, to "friend" our IT and BI departments, and to start from scratch with the IT and BI department, and to learn some tools for non-programmers just to have quick results, yes, like Tableau Public, Excel for Databases, and Qlikview to analyze large datasets, and to "build once and use many," have this approach for investing in developing datasets.

So, just one case study. We have two case studies. With Argentina's inflation data, in 2007, the government intervened in our national statistics initiative, and since then, INDEC has reported an annual inflation below 11%. Economists and provincial governments started reporting higher inflation rates. Private consultants, too. And in March 2011, Argentina fined firms over inflation data. They could not communicate this in their own CPIs. These firms teamed up with opposition in Congress and now they are releasing monthly inflation data. And also The Economist says, "Don't lie to me, Argentina." [laughs/laughter] And they removed the figures from The Economist.

So, one very smart journalist in La Nación, he took the simple example of, if inflation is relating to poverty, and we cannot speak about inflation, and we think this has effects in poverty…. This is a slum in Argentina ten years ago.

And with Google historical images, he shows, well, 2010, look at this. And the title says, "Will the government fine Google Earth?" This is another example of the same thing.

Another example to show data can show what we cannot tell. This is another example. We took a consumer association price index dataset that contains 28 products of the supermarket. It's measured since 2002. And it always comes in the newspapers as a press release. This is a press release. We love this press release that we converted into a dataset. And so we can show this price index in the two datasets, the Lider and the economy one. But also with the same dataset, and we keep it monthly updated, we have the 28 products. So, we have stories we can make with data of everyone and comparisons with the Tableau Public that has a dynamic interface with data. This is a share [data] story, for example. You can do it. You can steal these infographics, and you can do it, and compare it with another of your country, for example. The other thing we did, we put [in] official inflation, and every month, we update this with the private consultant's inflation that is released in Congress. So, these are many ways to show the same as a map also.

Another case, ah, and this is our Open [Data]. One month ago, we started opening data ourselves with our sources, our own sources for journalists. We made a team, and we are releasing. We are converting things to Excel to a CSB. We are releasing this in our Open Data Dashboard. It's powered by Junar, that is an open data platform, like perhaps you know Sograta Vas Data and these many, like the new YouTube's of data that are out there. We integrated that into our platform, and now we have more than a hundred datasets that we are keeping updated.

Another case is the subsidies for the Bus Transportation System that Diego Cabot suggested we transform 400 PDFs of more than 1,000 roads, each in three categories of subsidies. This was impossible to do, to process as a human, to analyze that information of each company that was receiving payments in cash every month. And subsidies were growing in Argentina a lot. For example, this is a PDF that has more than 1,000 lines. And this is coming without totals, but we have another that has totals. And I just wanted to show you the program running. We are developing scrapers to—as a video—to convert this public data and to build our own datasets. [Trying to play video. Technical difficulties.] We made the effort to make it in English, but it was the program working. [Tries again to get the video to play.] I have two minutes left.

Rosental Calmon Alves: Yeah, okay. [laughter]

Angelica Peralta Ramos: Inflation. Little inflation. [Tries again to get video to play.] Brian told me I have his extra minutes.

Man: His five minutes.

[Laughter.]

Angelica Peralta Ramos: We negotiated before. Inflation. [Tries again.] OK. It worked. In the meantime, this is not the video. But I'm telling you that we are also committed to opening data, but there is no open data movement in Argentina, so we just, as a medium, we feel we have the duty to do it ourselves, and we are doing it. And that's why we are building the databases ourselves. We are also.… We have launched a data blog, like, following the Guardian's idea. And all these New York Times and Los Angeles Times are wonderful places to learn with their blogs where they explain how they do their work, data visualizations, and everything. [Unable to get video to work.] OK. I'll follow with the presentation. We tried it yesterday.

Well, I'll tell you that we are doing also an explorer of this subsidies dataset. It's 235,000 rows. We found that these PDFs were changing and the payments were going backward, so they were updated backwards, and so we took this challenge to Hackathon in Boston. And Matt Perry is a programmer, a Python programmer. He said, "This is a difficult challenge." He made a code for this. It is called PDF Spy. And he solved this problem of looking backwards for PDFs that are overwriting themselves. And we copied that, and now we are monitoring the changes of PDFs two years from now. And also, we are working in this open-data platform. And I'm very sorry I couldn't show this video. We practiced this before, but.…

OK. Thank you.

[Applause.]

Ben Welsh: All right. We'll get by. So, my name's Ben Welsh. I work on a team called the Data Desk at the Los Angeles Times. And my talk is called *Human-Assisted Reporting: How to create robot reporters in your own image.* If that's not clear right there. And so, I come from, I was trained in a tradition of journalism which is called computer assisted reporting. It's decades old. The whole idea is that we can be more efficient and do cooler and better investigative reporting by using computers. Right? This is an idea that's not new. It's been around for a long time, and the name kind of tells you that. Right? My joke always is that everyone is computer assisted now. Photographers, architects, pretty much any job you need to do, you use a computer, but you don't call it a computer-assisted architect or a computer-assisted reporter, right? It's only in journalism that we continue to distinguish ourselves with the use of Microsoft Excel. [laughter]

But anyway, so it's a spin on that term is what I want to talk about today, is this whole idea of computer-assisted reporting and thinking about it differently, right? And all the URLs and everything I talk about is going to be available right here on Delicious. So it's just http://lat.ms/robotreporters, right? So, in my opinion, this is how computer-assisted reporting works today … in an image, right? So, your editor or you, the reporter, have an idea,

something out in the world that you want to investigate and get to the bottom of, right? You pick up your weapon, your computer, and you go out hunting for it, right? And that's the way that most computer-assisted reporting gets down. People already kind of have an idea or a field or a dataset or something they want to hunt, and they go out and look for it, right?

And the idea that I kind of want to put out there as an alternative metaphor or way of thinking about what we do is this, which is from the movie, Minority Report, which I love. And it says, "How it ought to be" at the bottom there. And in this scene, there's these robotic spiders that are able to crawl all over and do an operation on Tom Cruise. And I think that if we can up our game and what we do in computer-assisted reporting, we don't have to go out hunting for the story, right? The computer can go hunt for it for us, right, and bring it back to us. And where are we while that's all happening? We're back at the bar with Cary Grant, [laughter], our editor, talking about the next story, but also enjoying a drink, right? And this is a shot from the movie His Girl Friday, which if you haven't seen, do yourself a favor tonight.

And this is [how] the idea first kind of came to me when I saw a website right here which was created by Matt Waite, who's a leader in our field, now a professor at Nebraska. And he's famous for the website PolitiFact, but he made a lot of other sites in St. Pete, and this is one that's kind of gone now that he did about real estate. And it had a page for every neighborhood in Tampa Bay or the Tampa area, and there was a map, which is now dead. This is from the Internet archive. I had to kind of, you know, pry this out. And it had the latest home listings on a list, and then it had this paragraph right there. And I read this paragraph, and it was what Matt called a Mad Lib right? It was an automated paragraph written by an algorithm that said, "Based on this week's data in this neighborhood, here's the story." And for every neighborhood and every page, there was this same paragraph, but it had different information and it was up or down depending on the trend in that area. And I saw that and I said, "My god, that's news!" Right? That's not even…. You know, it's like an automated news story.

And that kind of inspired me to think about what I was doing in a different way. And at the Los Angeles Times, I've spent some time over the last few years experimenting with that. And I want to kind of show you how that process works and how you can create algorithms that write the news for you, right, and also find it.

So, what I'm going to do is teach you how to Dougie. [laughter] Right? So, all right, so here's how you Dougie. Okay, (1), you find a simple, repetitive, and moving data stream that updates every day or with some frequency, like home sales. In this case, this is an email that I received with a list of blacked out other people every morning from the Los Angeles Police Department at about 2:30 in the morning. It includes a CSV file, spreadsheet that has everyone arrested the previous day and booked by the LAPD. So, I have my

structured, simple, repetitive moving data stream that lands in my inbox every day, right? I then do a pull and a parse, and then I put it on loop, right? So, I write a script that goes in my inbox every day, looks for that email, looks for that attachment, pulls it in, parses it, loads it into a database, and then I set up a system so that that just runs every day. It's just an automated data pull, right?

Then, this is where the fun part comes in—the algorithm. You can then write code that will ask and answer the common questions that a reporter would ask when they were looking at that same dataset. There's so much of what we do, these questions we ask, "What was the biggest? What was the most recent?" These sort of basic journalistic questions we ask out of data that really can be turned into algorithms or code when you think about it. And these are just some examples that I thought of looking at this dataset, right? Then, that can turn itself into code, right?

So, is this the first code of the conference? [laughter] [Screen shows: #-Get-the-highest-bails-in-the-new-data…] Right? If it is, I'm pretty proud. So, this is an example of like, you know, every day you would want to know, "Hey, what were the most severe things that people got arrested for yesterday? What were the biggest deals?" And a proxy for that in the data is what their bail amount was set for. You know, the worst thing you did, the higher your bail, most likely, right? So, this is just a little bit of code that every day goes through that spreadsheet, sorts it by bail from highest to lowest, slices it off, and says, "Hey, I've got the list of the biggest bails," right? And what do I do with that? Oh, I send it out in an email to all the reporters who cover the police and crime for the L. A. Times, right? And we don't just get the list of the biggest bails and send them that. We also keep a watch list, right? We want to know anytime anyone who's a minister or a producer or a musician in their occupation field gets arrested, and that gets flagged, right? So instead of having to comb through this book every day as a reporter and spend all their time doing it, the computer can automate a lot of that process and do it for you and then send you a nice email.

You can see that's an alert. You also could make like a dashboard to drill down. This is an internal web page we keep at the L. A. Times that just has everybody arrested yesterday. But it also has some search features, so when someone is arrested for a major crime, we can go look for previous arrests, etc., etc. It's a research tool internally for us to use. And this is an example of another piece of code that then takes like a structured dataset like that and turns it into a sentence, right? It sort of diagnoses certain things about the data and then writes a mad lib, writes a little sentence that kind of tells you something about it. This actually isn't crime. This is census data that we did for a neighborhood site we keep. And it wrote this sentence. So, for every neighborhood in Los Angeles, we can write a sentence that tells you (1) the data point that's interesting to you, and (2) some contextual comparison along with it that links to other stuff, right? And so that way, I wrote, you know, 250 of those by writing it once with the template.

So, but really, what do you get out of doing this sort of thing? Well, 1) you get breaking news, right? So, this is an example where Puck from The Real World was arrested. Not the biggest deal, but we scooped TMZ and had the news first in the world, because the alert system caught it, right? We're watching closely through computer programming. 2) It's a way around PIOs. One of the biggest crimes in Los Angeles last year was on opening day of the Dodger Season. A man was brutally beaten, a San Francisco Giants fan. He quickly became a symbol for the decay of the Dodgers organization and our former team owner Frank McCourt. And the police arrested the wrong guy the first time. They screwed it up. When they arrested the first guy, there was a big press conference. Oh, man, everybody's got to know, right? We got the guy. Well, it turns out it wasn't the guy. And when they finally find the people who really did it, they tried to kind of hide who they were arresting. There was like this.... They didn't want to tell the media right away who they were going to go bust, but I had the data. I didn't have to ask the PIO. It was in my system and arrived at two in the morning. And we were the first reporters knocking on those neighbors' doors, figuring out who these guys were. Because the system got us a step ahead, right?

You also get instant analysis. So, Occupy L.A. was camped out across from the L.A. Times for three months. There was like a three-day standoff with the police, where they came and cracked down and rolled everybody out. When they did the big arrests, a couple hundred people were able to instantly do a census of all the people that were arrested and tell you something about them using this data. We actually published a list of all the people who were arrested and some other things about them that were in there.

You get the automated copy. This is from our blog, The Homicide Report, where we try to track every homicide that happens in Los Angeles County [and] have a post for every person. We don't have enough resources to do a lot of reporting on all of that, but certain amounts of information based on the coroner's data can be automated, and then we write that. That's the bare minimum for every post is the automated paragraph. And then as we gather more information, we then write through that and add more to it.

This is a similar thing we do. This is an automated blog post written by the computer that runs a couple of times a week when we get new LAPD crime data. It analyses it for trends, and it tells you what neighborhoods in Los Angeles this most recent week have had — are having an uptick in crime historically. Right? And here's another thing from our crime site where it's all automated news.

Same thing with earthquakes. My colleague Ken Schwencke did this. When an earthquake happens, everybody's going to the USGS site copying and pasting. "Where's the link? I can't find it. Where is it? Ahhh!" We don't gotta do that. It's structured data. We have a computer system that just sits.... Uh-

oh. I better hurry. [laughter] So, we have a computer system that just automatically writes a blog post and sends it in as soon as it happens.

Anyway, there's a lot of other stuff you can do this with, which will be fun. There's companies that are trying to make money off it. Like, Narrative Science is really good. They do some awesome stuff. I'm just making news. They're trying to make money. Who's smarter? [laughter] I'm out of juice, right? [laughter] And so I had — no, I had a big finish to try to make a more serious point and kind of talk a little trash. I'm sad I won't make it. But that Narrative Science guy, the next slide was a quote that he delivered to the New York Times, where he said, "In full visionary startup, I am the future mode." Right? Said, "Within five years, a computer program will win the Pulitzer Prize." [laughter] "And I'll be damned if it's not my software." [laughter] Right? Well, you know, I hate to break it to him, but guess what? Computer programs have already won the Pulitzer Prize. They've won a half-dozen of them, starting in 1989 with Bill Dedman's story in Atlanta, right?

Woman: The Color of Money.

Ben Welsh: The Color of Money. That's right. And my point is, is that what we should really strive for is not to automate for automation's sake or to save money, but what will really be great is if we can automate and make it easier and lower the barrier to do the kind of work that wins the Pulitzer Prize. That's already been done by people that come before us. And we need to be respectful and see what's in that tradition that's worth saving and worth automating and worth making more efficient, rather than just throwing it out acting like it doesn't exist. Because, you know, I work at an old-line media institution. I complain about it every day. It drives me nuts. But there's also a hubris in the startup community around this idea that, you know, that they don't need to learn anything from the past. So, there's nothing worth saving about journalism. There's a lot worth saving! And there's a lot worth doing. And we're the people who are going to do it. Gotta be the motion. [laughter] But it's just, you know, OK, some old newspaper guy didn't like your blog. Get over it. [laughter] Write a story that's worth reading. Write a story that's worth Brian's mom reading. [laughter] Let's fucking do it. I'm done.

[Cheers/applause.]

Question & Answer Session:

Aron Pilhofer: So, I think everybody saw the range. OK. So, I think it's clear, right, the range of backgrounds and skills and approaches here. This is, I think, just scratching the surface. I'm just curious. I have one quick question. I definitely want to open it up to the room here very quickly. But I just want to know, on this panel, who here would call themselves a journalist? [They raise their hands except Angelica Peralta Ramos is hesitant

and Alistair Dant raises his slightly but motions with his hand as if to say it's iffy.] OK. [laughter] Alistair, why not?

Alistair Dant: I wouldn't necessarily call myself a journalist yet, because I've been at the Guardian for about three years, and prior to that was very much just a software engineer. I'm learning about journalism. I feel like I'm some kind of alien observer in the world of journalism in some respects. But yeah, I think I'm kind of trying to find ways to facilitate journalism, but I'm probably more of a facilitator than a journalist myself.

Aron Pilhofer: I'm not sure I would agree actually. I think what you do is journalism, but I think we could discuss that.

Angelica Peralta Ramos: The same for me. I am a computer scientist. And I don't code, but I understand data, and I find data as a convergence point between lots of skills, so that's a new opportunity for journalism. But I'm not a journalist, so that's why.

Aron Pilhofer: So, what you're seeing is data people coming or people coming from the code side, from the technology side, journalists coming to the technology side. You're seeing an interesting convergence both on this panel and in the industry. Brian, why do you think that's important?

Brian Boyer: OK. [laughs/laughter] Um … because what we're doing is useful? I mean, golly, I don't know. Why do I think it's important? In what context?

Aron Pilhofer: Why do you think it's…? Where do you think this is taking the industry? Why do you think this is…? What's changing? I think as Ben emphatically pointed out, we've been doing data analysis for years, maybe decades. We've been involving data in our journalism for decades. What's different about this, what we're talking about right now?

Brian Boyer: Well, there's mountains of data available at this point, right? I guess one of the rants that I've been getting on…. I may be borrowing one from you that our friend Derek Willis at the New York Times gets excited about, is about just the sheer quantities of data that is available to, let's say, oh, the Obama Campaign or Phillip Morris or somebody. And they have data scientists on staff. They have really, really smart, incredibly well paid data analysts [and] quantitative sociologists. There are lots of names for these people. And they are using that data to sell you cigarettes or to get their candidate elected. And we don't know how to do that. I'm not a data scientist. I'm a programmer who knows how to work with databases and is slowly learning how to tell stories. I don't think anybody on this panel would call themselves a data scientist. And I don't think I know a single data scientist who works in journalism. It's a problem. It's a gap. And we have to…. We've all got up our games. And we also as an industry have to bring in a lot more people to do that sort of work, because we're ignorant.

Alistair Dant: I would add something to that, which is that at this conference Strata, where I spoke at last year, which is a data summons conference, there was actually a lot of attention paid to what journalists could bring to them. Because the one thing, I think, that flows through everything that we've been talking about today is the ability to find the stories, to find a headline. I mean, every time we start a piece, we sit down and before we go any further, we have to know what the headline is. You know, so, this is something I was saying to Brian yesterday that we're going to be like the Crystal Reports of the big data world. [Aron Pilhofer laughs.] But yeah, journalists are increasingly —

Aron Pilhofer: That was a joke, by the way.

Alistair Dant: Yeah, sorry, it was a bad joke, but yeah.

Brian Boyer: An enterprise programming joke actually. [laughs]

[Laughter.]

Alberto Cairo: I want to add something to that. I mean, I'm going to make a comparison between journalism — what is going on in journalism and what's going on in science. If you take a look at a scientific papers from 20 years ago, you only had, you know, one offer or two offers per paper. If you take a look at [a] scientific paper today, you will see that each paper is authored by 16 people. And what that means is that the world is growing more complex. We have more complexity to deal with. And our work as journalists is to make sense of that complexity. But for making sense of that complexity, we have to bring together skills from many different areas. That's what's going on right now in the newsrooms. You need programmers. You need data scientists. You need information visualizers and designers and pure journalists, so you have to bring them together to try to make sense of the world for your readers or for your audiences or help them make sense of the world in some sense by creating tools that will allow them to do that. I think that's the main difference between the past and the present.

Aron Pilhofer: And let me add one more community to that, and I think, Angelica, you mentioned this, the activist community, non-journalists, people who are outside the newsrooms. And I think this is true both of your newsroom—not as much mine, unfortunately—[and] Ben's, absolutely. So, maybe you can talk a little bit about how newsrooms need to reach out beyond the four walls of the newsroom.

Angelica Peralta Ramos: We are engaging with these communities in the local level in Buenos Aires. I recommend we go to hackathons. Last Saturday, I was in one. I went to four or five hackathons Sunday, Saturdays, and if you want to learn, you have to go and try to understand and self-teach yourself. This is very important, because who are there are programmers,

are data scientists, are people who want to help, but they don't perhaps understand the business or the content you know as journalists. So, this is very important to work in teams and to find the other part that you need, and ask.

Aron Pilhofer: Ben, you guys have done a lot of crowdsourcing projects. You've reached out to the community extensively. Why have you chosen to do that? And what value do you think this brings, in particular, to your reporting?

Ben Welsh: Well, for one, it's a lot of fun, which is a good start. But I just think, for instance, with the neighborhoods project we did, we knew there are no official boundaries. It's a long-time disputed thing in the city of Los Angeles. So at the outset, we knew that there would be no right answers. It's not like the L.A. Times could draw the lines and they would be perfect. If that was possible, you might try. But we had the humility, I guess, at the outset to admit that. And that's where bringing in the outside made it a lot more credible, I think, to our audience. And it resulted in a better product.

Aron Pilhofer: So ten years from now, I'm curious—this is a toss-up question for the panel—ten years from now, what's the newsroom going to look like?

Ben Welsh: Oh, fuck. Much, much smaller.

Aron Pilhofer: Five years from now.

Ben Welsh: Much, much smaller.

Aron Pilhofer: Much smaller. Questions, yeah.

Ben Welsh: Though, hopefully, leveraging technology much more efficiently than I do, but in that same department to get, you know, stuff done. There's this old maxim that's much maligned in newsrooms, "Do more with less." You know, "We've just laid everybody off, but we're going to do more with less." It was parodied by David Simon on The Wire in the season he did there. And it's bullshit. I mean, it's a lie. It's not true. You can't in most cases, but there is one hope in that department, which is through software, technology, open source projects like MESO Project, we can lower the barrier and increase our efficiency in how we work with stuff. And that really is one of the few ways you really can do more with what you have or, you know, sometimes, hopefully not less, but....

Aron Pilhofer: Cool. All right. So, we have people lined up. Why don't we start over here?

Rose Ellen Downey: Hi. My name is Rose Ellen Downey. And my question is for Ben. I was curious with the L.A. Times, I lived and worked in L.A. for

five years in the entertainment industry, and the entertainment industry is obviously prevalent in L.A., and I was wondering what you do with data relating to the entertainment industry and data mining that for the readership there.

Ben Welsh: Mm-hmm. Well, we just this year before the Academy Awards did a data-based investigation thing where we published for the first time ever an analysis of the members of the Academy. This was a data project we did where we were essentially leaked a list. Am I allowed to say this? Yeah. We were essentially leaked a list of people who are in the Academy, which is secret. And then through a very tedious process, we set up an internal database that had reporters then individually verifying the people, collecting their age, race, and gender, and then using that to do an analysis that found that the Academy is—surprise, surprise—overwhelmingly old, white, and male. [laughter]

Rose Ellen Downey: Thank you.

Gerald Rich: Hi. Gerald Rich. Another toss-up question for you guys. What was your hardest hack?

Ben Welsh: People.

Angelica Peralta Ramos: Skepticism.

Panelist: Yeah.

Alberto Cairo: In the newsroom, yeah, I would say skepticism sometimes, yeah.

Panelist: Yeah.

Gerald Rich: Hack like a project.

[Laughter.]

Alberto Cairo: Oh, a project.

Gerald Rich: Sorry. Yeah.

Alberto Cairo: All right. Well, in the one that I showed, I mean, it's just very basic. The data was in a website. They did the project about the phone numbers and phone calls and stuff. The thing is that, well, we talked a little bit about open data here and open data laws. And Brazil has one. But the thing is that the government is forced to release data, but it's not forced to release data in a usable format. So, they don't normally release things in PDFs. They do that, but they also release it in websites, and then you phone them and say, you know, "Give me access to your database, so I can get

your data." [They] said, you know, "That is dangerous. You should not do that." So, you're basically forced to hack them sometimes and get the data yourself. That's the main challenge that we face over there. It doesn't happen here in the U.S., but it does happen also in Spain, where I am from.

Brian Boyer: I'll throw one more thing into this. I'm a professional software developer leading a team of professional software developers. Making software is easy. We like it. It's what we do for fun. It's the editing that's hard. It's the journalism that's hard. It's figuring out what to name that column. It's figuring out when to use a map and when not to use a map. Making the right software is very difficult and requires many of the same though processes as editing a regular story.

Alistair Dant: I'll quickly come back to the subject of dealing with big data. We did a project just before Christmas, which was a pretty nerve-racking one, in that we had to get a sort of mixed team of academics, who were all very talented data scientists, to understand a three-week deadline, which was kind of a little bit foreign to them. [some laughter] But we basically had 2.5-million tweets, which pretty much covered the entire period of the London riots last year. And with the help of the academics, we were able to analyze those. And in that three-week period, we also devised a way of then basically playing back the evolution of certain rumors through Twitter. And I think that was a very hard project from kind of every angle, not least that a lot of the things I was showing earlier were originally done with Flash. That's actually a pretty mature set of tools for doing animated interactive visuals. And having to now start building those things with open standards has been a real challenge. I won't go into the details, but doing something that's kind of as complex as what we would previously build in Flash in three weeks is a stretch.

Aron Pilhofer: Question over here.

Woman: I'm curious about, you know, bringing together traditional journalists and traditional technologists. And I really see that that's what we are talking about here. We're talking a crossover. On your teams, do you have more technologists or do you have more journalists? Who do you bring to the table?

Alistair Dant: Our team, unusual for this field, is made of almost all people who were programmers first and then found journalism, either because they believed in the mission or the religion of it or they wanted to make things they could show their mom. [laughter] No really. We have to make an emotional appeal to hire software developers to come work in this stupid business. [laughter] But I love it.

Angelica Peralta Ramos: I think we are more journalists than technologists, as I've said. But from the technology side, you have to find the people who value data, doing things, building things, discovering knowledge,

sharing data as a way of innovation, and they have a different motivation. I know different technology profiles. And the ones who feel something for this, they are not so easy to find. And just like changing the world a little bit, that's what we want the to feel, and this is how we engage [them]. So, that's it.

Alberto Cairo: In my case and the team that I used to leave until December last year, it was one jack-of-all-trades, which was me—journalist and graphic designer also, but above all a journalist—four graphic designers, infographics designers, and two developers. We hired two developers in the last year that I was there, and it was a huge change for us.

Ben Welsh: Our team is called The Data Desk, and it sort of exists more in our imagination than in the org chart. We don't really have a boss. We all work for different departments. But we're a crazy team, and that includes the kind of older computer-assisted reporting team that preceded a lot of the Internet work. That's Doug Smith, data analyst, Sandra Poindexter. They write in SAS and do classic census, old-school car stuff. Malloy Moore, who's a converted librarian, who does a lot of work for us, too, cleaning data, working with data. She's getting better. She's learned some jingo. And then myself and Ken Schwencke, who are sort of journalism school students converted into computer programmers. And if you want to become one of those people, you can do what I did, which is go to the University of Missouri and the National Institute for Computer-Assisted Recording—pick up your chest—[referring to Alistair Dant's t-shirt with NICAR on it]—it's called NICAR, and that's a great — they have a graduate program, and that's a great way to spend some time doing real journalism and real programming at the same time and meet everybody in the field and get plugged in. I'd recommend it.

Alastair Dant: So, I've seven people on my team, and I could tell you about the fact that we have a more or less half and half male/female split. I could also tell you about the fact that we embrace a number of ethnicities, including cyber punk, [laughter], but I want to tell you what actually brings…. Oh, and also in terms of skills, we've got developers. We've got journalistic backgrounds, and we've got a wonderfully spirited Portuguese lady called Mary Anna, who is just a kind of force of nature. But what brings everyone on the team together is the fact that everybody is very pragmatic, everybody is capable of seeing…. Well, the Guardian campaign this year is all about the whole pitch, but everyone on my team is capable of looking at the kind of whole of a story that we're trying to get to and not letting details bog us down or get in our way. We tend to try and move quite fast. And having everybody able to just find a new way of doing something at kind of like last minute notice is pretty important.

Rasmus Nielson: Thank you. I'm Rasmus Nielson from the Reuters Institute at the University of Oxford. So, one very quick question and then a slightly longer one. The quick one is, does any one of you know whether the

things you are engaged in actually make money for the company that you work for? And then the second one, which sort of feeds off that one is, many of you come from organizations that dedicate large amounts of resources to what you're doing. What do you think are the takeaways for smaller newsrooms, which are the newsrooms of the future? But also most of the newsrooms of today cannot assign 14 people, cannot run a website like The Guardian. Amazing as it is, that is a loss-making website. I think The Guardian lost 40-million pounds last year. 20% operating loss is not a very attractive proposition for most news organizations. So, I think we all appreciate the value of what you're doing, that you are doing amazing innovative work, but I suppose there are two questions, one, what is the business of this? And what are the takeaways for organizations who don't have the kind of resources that the great newspapers and great news organizations that you represent have to dedicate to this?

Alberto Cairo: I think that I'm going to get started with that. I don't have a good answer for the first one. I don't think that anyone in this room has a good answer for the first question. But I guess one thing that can be brought to the table by this kind of journalism is that it can bring attention to your organizations by doing things that nobody else is doing out there, but that's not a very good answer. The second question, I mean, I can answer. And the thing is that you can do very good data journalism, and Angelica addressed that in her presentation, by using Excel and Illustrator. That's what I do. I'm not a good programmer. I know a little bit about programming, a little bit of Java script, Action script, but I'm not a good programmer. 90% of my work as an infographics designer and information visualizer is done in Excel and Illustrator. I usually call that low tech visualization.

And in fact, one of the things that I do in my courses at the University of Miami, at the beginning of each semester, is to try to take out the fear that students have for this kind of journalism. And I basically tell them, you know, "You don't really need to have very deep computer skills to do this kind of journalism, because it's about the story. It's about storytelling and extracting the stories from the data. And you can do that with pen and paper sometimes. Of course, if you know programming, that can help, and that will help down the road, but at the very beginning, Excel and Illustrator, and that's it." So, that's what small newsrooms can do. Start investing in tools that are easy to learn, easy to understand, and Angelica mentioned some of them—Tableau, Excel, Illustrator. Things that can be learned very fast and can be very effective. And down the road, you can learn the advanced stuff that The New York Times does or The Guardian does or everybody else on this panel does.

Ben Welsh: My personal opinion is that all these new, small, startup newsrooms, if they want to have big impact and do investigative work that makes a difference, they need to have at least one person who's a specialist at doing data analysis for investigative work. I can't tell you that our methods have a track record of making a lot of money, but I can tell you

they have a track record of getting great stories and kicking ass. And so if you're goal, even if you're like a non-profit, is to kick ass, you should have somebody who can do this stuff.

Aron Pilhofer: Brian?

Brian Boyer: I'm going to run this calculation when I get back to the newsroom so I can answer this definitively, but I feel that are team, man-hour-wise, produces just as many pages as any other team in the news organization. So the page views divided by man-hour number, I think our ratio is probably pretty solid compared to the rest of the newsroom. Does that make us any money? No! Because our whole business model is broken, right? [laughter] So, I'm not going to take the blame for that. [laughs]

Ben Welsh: Right. Yeah, and if you're just talking about page views, you can say, like, the homicide blog that we do, we automate all these blog posts that come out and then just do a minor amount of work to clean them up, so the actual time it takes to make the blog is next to zero. But nobody's going to buy ads.

Rasmus Nielson: Just on the business side of things, it's true that the old model is broken, but it's not everything is equally broken. And I think that's an important distinction, whether you're throwing away a lot of money or losing a little bit of money.

Ben Welsh: Right.

Rasmus Nielson: I mean, from the point of view of the business.

Ben Welsh: Right.

Aron Pilhofer: Angela, did you have something to add?

Angela Peralta Ramos: We really think this is innovation and the business for this is to stay in the business. This is the raw material of the future, and the audience is there. So, these data applications will transform into mobile services, and so you can demand more information directly from your news organization or from your data provider. We are convinced this is directly related to innovation and growth.

Ben Welsh: Can I ask a heretical question on that topic? So, we did the Academy's project. So, the L.A. Times is like on my computer. Not that one, but a different one. We have the exclusive list of everyone who's in the Motion Pictures Academy. No one else has it outside of the Academy. You could probably sell that, right, [to] people who want to market their movie for the awards. "Hey, I want to send a package to everybody in the Academy to let them know to vote for Kate Winslet," or whatever. If the business

people came to me, which they haven't, and said, "Hey, let's sell it?" Should I do it?

Panelist: No.

Carla: Hi. My name is Carla. And I thought of this question when I was looking at Angelica's data and Ben's data. What do you all do to make sure the data that you are reporting or portraying is accurate, understandable, and not taken out of context?

Angelica Peralta Ramos: Our data is taken from official public records, so this doesn't mean that the data is accurate, but that's somewhere we start with, because before we had nothing. So, that's it.

Aron Pilhofer: So, anyone else very quickly?

Alberto Cairo: Yeah. I mean, there are data that you cannot see if it's true or not, because it's from official sources and sometimes, many times official sources are not accurate. But that's what you have to deal with, and that happens every day in journalism. The other thing that you can do and that is done by everybody is that every time that you have the opportunity to talk with an expert about the data that you have in your hands, that's what you do. So, you run a data analysis yourself. And we did that with educational stories in Brazil when I was there. So, we did a story about education in Brazil, and we ran the data, we analyzed the data, and once we had the graphic ready, we showed the graphics and the data to specialists in education to contrast what we have to tell and to make sure that the story that we had in our hands was reliable. So, that's what you have to do. It's a very basic answer, but that's what journalists do all the time.

Aron Pilhofer: Super quick question over here.

Man: Sure. Yesterday, we heard about how journalism is becoming more democratic through the online medium, how pretty much everybody can create content and publish it, put it out there for everybody. What you guys are talking about today is how journalism is becoming much more specialized. That you need to have very special skills to be able to do that. So, do you see a dichotomy there? And how does it play out for the future of democratic journalism via online?

Aron Pilhofer: Alistair?

Alistair Dant: So, in response to that and also going back to the previous question, which was about the way that we can share what we do or the way that organizations like ours that are perhaps privileged, I'll get another plug in for our MESO Project. I think what we're trying to do there is democratized data journalism. I think we're trying to show the way that we do things and

we're trying to make it so that every piece we publish ultimately will be something that anybody can take as a starting point for their own piece.

Brian Boyer: I mean, that's kind of the point of the Show Your Work t-shirt. We blog, we explain exactly. We explain how we did the data stories. We release the code with documentation so people can go build their own sort of work like ours. And we release the datasets for every single thing we publish. So, we're trying to spread this thing out. We're trying to train both our peers in the industry and also the public about how to do this stuff.

Aron Pilhofer: OK. Rosental is going to yank us off stage. Thank you everybody.

Rosental Calmon Alves: Yes! We thank you. This was fabulous! Fantastic! Wonderful! Thank you!