

ICPSR Data Brunch Podcast Episode 10: Data Brunch Live! Extremism in Social Media

Original Air Date: October 2021

Transcript

Dory Knight-Ingram:

All right. Hey, everyone. Welcome to Data Brunch with ICPSR. If you love data, this is going to be food for thought. I'm Dory.

Anna Shelton:

And I'm Anna.

Dory:

We are recording these episodes, well, I'm going to say live from the 2021 ICPSR Biennial Meeting. So please excuse cameos from canine colleagues, kids in class, and other unexpected moments. And if Facebook has come back online, no, I'm not going to go there, we might go there later. But if you are posting about us, please use the hashtag #DataBrunch and also the hashtag #ICPSR.

Anna:

Today, like Dory mentioned, we are excited because we're coming to you live from the ICPSR Biennial Meeting. This has been three days of incredible data-related training and resources. You are also hearing all of the bloopers that Dory and I usually have live, so you'll know how much our producer extraordinaire, Scott Campbell, means to us. Thank you, Scott.

Dory:

Thank you, Scott. In just a moment, we'll get to talk to Dr. Libby Hemphill, who studies social media, civic engagement, automated moderation techniques, and the connections between them. We are incredibly lucky because we get to call Libby a colleague. Can't think of a better way to close out the Biennial Meeting.

Anna:

We do have few things we want to let you know about first. First, if you haven't seen this yet, we are hiring. As of this recording, we have five job openings, including a data project manager and many more. We will link to those in the show notes. We do have quite a few webinars coming up, and I'm going to link to those in the show notes and also here live for folks attending. Those webinars include topics like healthcare, health reform, bone fragility related to disabilities. There's a webinar coming up on student voting behaviors, on restricted data, on GIS data. We have a lot of training, and it's free. For everyone joining us here live, again, just head over to icpsr.umich.edu. You can click on the events listing that's over there on the left to register.

Dory:

Thanks, Anna. As you know by now, we have a constant stream of new resources, and they're not all related to extremism and social media. For example, our newest current events in the bib, AKA the bibliography, shows how exposing students to a political science class in community college helps create an informed electorate. Frank Fernandez recently published an article in Educational Research called Turnout For What? Do Colleges Prepare Informed Voters?, which uses the community college civic outcome survey found at ICPSR.

Anna:

And now I have that song stuck in my head. I don't know if we're allowed to use that song for copyright reasons, but you know I'm all singing it. If you are interested in digging into data related to extremism and social media and that's what brought you here, we have a new study for you. This study is called The Role of Social Networks in the Evolution of Al Qaeda Inspired Violent Extremism in the United States. It's from 1990 to 2014, and the purpose of this research is to address the question, how do foreign terrorist organizations mobilize Americans to carry out attacks on their behalf? There's very interesting data available there, so we will link to that right now in the chat and also in the show notes.

Dory:

Thank you, Anna. And without further ado, we would like to introduce you to our guest for today, Libby Hemphill.

Libby Hemphill:

Hi. Thanks for the virtual claps. Those feel pretty awesome. Thanks for having me.

Dory:

So Libby directs the Resource... Sorry. This will be a blooper where I stop and say what I'm saying again in our normal workflow. So Libby directs the Resource Center for Minority Data at ICPSR and holds a joint appointment as an associate professor in the University of Michigan School of Information. She studies politicians, nonprofit organizations, and television fans to understand how people use social media to organize, discuss, and enact social change. She also developed automated mechanisms for moderating and classifying content in social media in order to reduce toxicity in online conversations.

Anna:

And just as a reminder, this is a live taping of this podcast, so our attendees who are here with us, you are welcome to type in questions at any time. There's a Q&A button down at the bottom of your window. We will get to as many of those questions as we can.

Dory:

So, Libby, first, congratulations on being a co-recipient of the University of Michigan School of Information award for outstanding conduct in instruction, leadership, and community-engaged research.

Anna:

Woo-hoo! Well-deserved.

Libby:

Thank you.

Dory:

So in the School of Information's news release about the reward, you were praised for acquiring data sets and developing materials, tools, and research data about historically marginalized populations, appropriately accessible and reusable. That'll be a great lead-in to the first question that we usually ask our Data Brunch guests, and it's that we love to see the stories behind the data. So what makes your research a great story?

Libby:

Oh, wow. Well, I think there are a couple of different stories that we can tell from my research. The award from UMSI is actually... The way I read it is that it's an award for doing the work that I get to do at ICPSR, where it's for everything we do at RCMD, which is the Research Center for Minority Data, where we've had some big successes in the last couple of years on expanding our collections to include gender identity minorities and sexual orientation minorities and then to shore up other collections with racial, ethnic, and religious minorities as well. So that's my administrative role. In my research role, the one that I think we're going to focus on today is that I study white supremacy online and ways that we can use human AI moderation techniques, where AI means artificial intelligence, which is a pretty broad term for anything computers can do to help out, is how I'll use it today, so how we can use human AI collaborations to try to address some of the problems of white supremacy online, especially white supremacists' speech specifically.

Dory:

Thank you.

Libby:

Mm-hmm (affirmative).

Dory:

So can you talk about your research into using machine learning technology to moderate hate speech and how that can help social media companies adapt to the spread of hate speech on their platforms?

Libby:

Sure. So part of what I do, I consider myself a computational social scientist, so that means I'm a social scientist who uses computational methods. In this case, when I'm studying extremism online... And I'm going to say extremism and not hate speech in part because I think that hate speech is a triggering word for lots of different sides, where it's contentious, I think would be even a more accurate description, that what counts as hateful is not universal. Then, if we're talking about hate speech, we often end up in a fight about what counts rather than what can we do about it, where if we use a moniker like extremism, that gets us into more behaviors than just speech and gets us away from some of these definitional distractions that get us to only look at what is overtly hateful.

Libby:

So what do I do? What I want to try to understand is whether or not we can suggest ways that platforms could do a better job of following and enforcing their own rules about extremism and hateful speech on

their platforms. I'm assuming that platforms have policies against hateful and extreme speech, which the mainstream platforms like Facebook, Twitter, Reddit, Pinterest, they do. So we're going to set aside should they. They do. They have them. How do we make sure that they are enforced and that the standards are upheld? One of the first things we have to be able to do is to identify the speech that's not allowed or the activity or behavior that's not allowed anyway. So we have a detection problem, and then we have a mitigation problem. Assuming that we can find it, we need to know what should we do, and I do research at each of those stages. So at the detection stage, how do we know what's problematic, and at mitigation, assuming that we can identify something that's problematic, what can we do about it?

Libby:

At the detection stage, I'm interested in ways that we can use the expertise of marginalized populations to identify problematic behaviors online. Traditionally, machine learning uses a crowd, and I'm using air quotes. The people who are watching can see me, but you can't hear my air quotes in the podcast. I'm learning, through some experiments on Mechanical Turk, a crowd is overwhelmingly white and male, and you will not be surprised to find that what white men think is hateful is not the same as what those more personally impacted by hateful speech find as hateful. So when we use that narrow definition of what's bad to train computational models, we miss a lot of things that are really hurtful. Everything from microaggressions to overt misogyny or racism get missed if we only have one set of labelers.

Libby:

So in the detection, we work on both expanding the role of annotation, so human annotation, in telling us what is problematic so that we have better data to feed into machine algorithms to teach them to recognize lots of different kinds of potentially problematic behaviors. But in order to overcome what might be a greedy or over-reactive or overly sensitive system, we bring the humans back in, and we say, "Hey, the machines think this might be a problem. Can you take a look at the context and see if it is?" Some examples of where context might matter, there's often languages that groups use to talk within the group. I'm a queer person. I say queer. I say queer to other queers. That might be okay. But it's not okay for straight people to use potentially derogatory labels for queer people when they talk about me, even if it's okay for me to use them about myself.

Libby:

So we can't look at just the speech. We have to be able to look at the context of the speech, and the context is what humans are good at it, and the speech itself are what machines are good at it. So when they work together, we can do a better job of understanding is this particular exchange or behavior in keeping with the values of this place or is it not, and that's what we can do for detection. I feel like I'm rambling, so I'll pause there because I see there's been some chat that maybe I can check in on questions about. But I can also talk about the mitigation piece.

Anna:

Yeah, no, it's fascinating. I want you to just keep rolling because it's all... What I feel like is happening is I keep getting questions and then I get more questions. So one thing that you've brought up for me is all of this conversation you're talking about is happening on these social media platforms, which are private by nature. Do those social media platforms, do the private social media platforms, have a responsibility to the public?

Libby:

Yes, they do. They're actually public companies that are privately-owned. There are varying levels of private behavior that can occur on these platforms, but I have no problem saying that privately-owned companies have public responsibilities and publicly-owned companies have even more. So some of those are about what do private businesses own citizens under a system of capitalism, but some of them are also ethical concerns about when you occupy a space like Facebook or Twitter does in the world of public discourse, what moral requirements do you have. I think that they do have moral requirements to the public and that they haven't been keeping them.

Dory:

Thank you. Thank you so much.

Libby:

Sure.

Dory:

I'm still wrapping my mind around what you said earlier when you said that what a white supremacist thinks is harmful is not necessarily what someone who's affected by their work is harmful. I mean, mind-blowing. I'm sorry. I think I'm going to be chewing on that all day. Yeah.

Libby:

I mean, I think that this relates some to... There's a question in the Q&A about extreme and whether or not there's a shared definition of extreme. So there's a statistical definition of extreme. We can use the statistical definition to tell us how far out of the norm is a particular behavior, and if it's really far out of the norm, then it's extreme. So that's one way to do it. Whether that counts as, quote, political extreme or, quote, cultural or social extremism is different. But, statistically, there is an extreme.

Libby:

To Dory's question about the people who are doing harm or speaking and doing harm and those who might experience it differently, I think the... Oh, I want to look up what the name of the book is. I'm going to tell a story in just a second about how I learned to pay attention to this, and it was in my own personal anti-racism growth that I was like, "Oh, this is a thing that I did not recognize." So let me look up what the name of the book is. But, yeah, the idea that the people who are harmed, that you may not... So white supremacists often are not worried about whether or not they're hurting someone else's feelings, but all of us are so deeply soaked in... Well, all of us in the US are so deeply soaked in this white supremacist culture that we might be doing harm and not even notice. So we may not recognize the racist undertones or histories of common sayings or things that seem innocuous but are not really.

Libby:

I'll give two examples of anti-Semitism specifically, in part because I think among... White supremacists care a lot about race, but they have racialized Jews and Muslims also that are not races necessarily but they are cultural and religious monikers as well. From anti-Semitism, when President Trump came to Michigan to speak at the Ford plant... I don't even remember how many years ago this was. But he referred to the workers at Ford as... Or was he talking about the workers at Ford? In his remarks at Ford, he used the phrase good bloodlines to talk about Southeast Michigan and Ford workers, et cetera.

Without any context, you're like, "Okay, bloodlines. He must mean we're good workers and always have been."

Libby:

In the context of Ford, where if you're paying attention to anti-Semitism and the history in the US, you know that Henry Ford was a devout anti-Semitic speaker and writer, so you have Ford and his legacy of anti-Semitism, and then you have the essentialism of white supremacy that is about our genetic bloodlines and what they mean, and then you have a US president speaking in this historically anti-Semitic place saying this historically anti-Semitic phrase and then being able to say, "But that's not what I meant." But that's not what Jews heard, so the context of where the speech is occurring and who says it and what historical references are embedded in it matter, and you can't just show up at Henry Ford's place talking about good bloodlines and not expect somebody to push back and say, "That's pretty anti-Semitic." So that's one of the examples. You say, "Well, but it was innocuous," and on one level it might be, but, also, it brings with it a set of hateful baggage that we have to acknowledge if we're trying to be inclusive and respectful.

Dory:

We have some questions rolling in. Let's see. I'll take one from our participants. Here's one that asks, "Are machines able to detect sarcasm, and do your human crowds ever find that sarcasm can be offensive?"

Libby:

So, yes, machines can detect sarcasm, but not very well. And do humans find them offensive? Absolutely. So this question looks like it's, maybe I'm not supposed to say names out loud on the podcast, but from a participant. You know who I'm talking to because you're listening. So, yeah, machines are getting a little bit better at sarcasm, but this is actually one of the challenges for platforms, is that in all of their terms of service, they carve out an exception for humor, satire, parody, et cetera. When people make mistakes about stuff that they've posted, whether they knew it or not, they're often like, "Oh, I was just trying to be funny," and sometimes funny hurts anyway, and then we need to look at whether the benefit of the funny outweighs the pain of the hurt or the cost of the hurt.

Libby:

Often, the way that I would look at this is that sarcasm that punches up is probably fine and sarcasm that punches down is just unnecessary. If you're familiar with the punching up and punching down, it's about looking at who has power in a particular situation. When you're criticizing somebody who is more powerful, that is a speech act that I'm willing to protect, where someone who is in power punching down or hurting people who have less power is a speech act that, while I recognize it as important under First Amendment law in the US, I don't find it as important to protect in other private spaces. I don't want the government making a rule about that, but I'm totally fine with my immediate community saying, "We don't punch down here."

Anna:

I'm wondering, can we go back just a little bit? And this is me showing my own lack of knowledge. Can we talk a little bit about... So I know that social media data is notoriously hard to crunch, hard to get access to. Can you talk a little bit about what data are you actually looking at and how does that... Can

you just talk about the data package itself? That's not the right word, but can you help me phrase this better?

Libby:

Sure. We could talk a little bit about what is in social media data and why is it hard to work with. When I'm talking about social media data, I'm thinking of the content the users have posted to social platforms, where platforms are places that exist to host user-generated content. That used to be a much narrower space. Now, it's so big, it's everything from Facebook to Reddit to Parler to Substack. But I'm focused on places like Facebook or Twitter, where there are potential for following relationships and for commenting on each other's content relationships.

Libby:

For a report that I did with the Anti-Defamation League Center on Technology and Society, I pulled all of the comments that have been posted to Stormfront, which is the most successful white nationalist discussion board on the Internet, their term for themselves is white nationalist, I'm not assigning a term to them, and then from Twitter users who, in their bio or in their tweets themselves, claim an affiliation with the, quote, alt-right and then all of Reddit. So I used Reddit as sort of the Internet baseline and then Twitter alt-righters as a sort of extreme in a public space and then Stormfront as extreme in a niche space because Stormfront is public but you and I are probably not going to comment there. I'm assuming, yeah, that you're probably not going to comment there. So what that data looks like. There are a couple of different ways to look at the data. Technically, if you're reading your Twitter feed or your Facebook feed or your Reddit, the homepage, or looking at the Stormfront boards, you're looking at data. You're just looking at it in the presentation layer that they have defined.

Libby:

Usually, when I'm looking at data, it comes in JSON, which is the JavaScript Object Notation, where it's a structured format that is represented in plain text but has key value pairs that can be nested, and I'm pulling them from the APIs that platforms provide, where API is essentially the way that computers talk to each other. So there's a set of queries that I can make, and I get a set of results back, and those results are in JSON. Then we put the JSON through some Python or R scripts to try to pull out what we're interested in. For me, it's usually the text and the date because I'm looking at language over time. I'm not particularly interested in individual users but rather the overall language that gets used and over time. Then we represent it in a table, just like a survey, just totally different data but same format. Does that help, Anna?

Anna:

Yeah, that does. For anybody who's following along, is there a website you can shout out real quick that somebody might go to, to see some of the stuff that you're talking about as they're listening?

Libby:

Maybe.

Anna:

Just to really throw you on the spot.

Libby:

You mean to look at raw data or to look at...

Anna:

Maybe a report or something like that.

Libby:

Maybe. Social media researchers don't often talk about their data, in part because we might get in trouble. So I don't have one off the top of my head, but I'll get you one to put in your show notes.

Anna:

Okay. Cool. That's perfect. And maybe this is a good spot to mention that going to someplace like ICPSR to get the raw data... Obviously, you all, who are here, know about that, but something like the Resource Center for Minority Data, which is part of ICPSR, is a really valuable place to be able to find data that would be useful here. Then hopefully for other folks like me who are not as good at the raw data but could definitely use some of that data that has been packaged and reported out, we'll get that to you in the show notes.

Libby:

Okay. I did put one example of a study at ICPSR, the #MeToo dataset from some researchers at Northeastern University. They were studying #MeToo, the hashtag, on Twitter for about a week in 2017. But if you use that data, you'll see a bunch of the challenges that we have in archiving and sharing social media data, that Twitter's terms of service, which are the rules that Twitter sets for using its platform, prohibit us from sharing tweets with one another. So what the team at Northeastern did was share the tweet IDs so you can look up each individual tweet at Twitter or through the Twitter API. But we're not actually allowed, under Twitter's rules, to share the data. So we can share metadata but not data itself.

Dory:

Let's see if you answered this question with what you just said about Twitter.

Libby:

Okay.

Dory:

Has the change in Twitter policy on research data that kind of snuck in mid-2020... Is it now possible to do more with archiving Twitter data and research sets? Has it given you more access?

Libby:

I love that you think it snuck in. Among academics who study Twitter, it was like, "Oh my God, now we have access." But it didn't actually change the rules. So the only thing that changed is that we can get more data than we used to be able to while still following Twitter's rule. The prohibition on sharing still exists. You can listen to other talks that I have given about this to get a sense about how I feel about terms of service and whether or not I have any moral or legal requirement to follow them. Google that. The answer is shorter than you think.

Dory:

We will link those, too. We'll try to get all this good stuff in the show notes.

Libby:

There are things I probably shouldn't say on the ICPSR podcast, but I've been public about it somewhere else.

Dory:

All right. Let Scott know.

Libby:

Okay.

Dory:

All right. Let's see. Let's take another one from our participants. I'm trying to see. Why is it so impossible for people to apologize for the harm, even if it came unintentionally or out of ignorance?

Libby:

I love this question, in part because we... So I mentioned I would talk about mitigation, so I'm going to tell a long story about a mitigation setting to answer your question about apology and harm. We did an experiment on Reddit where we used a human/AI combination, such that every time somebody personally insulted someone else on Reddit, so a very narrow subset of unacceptable behaviors, where when somebody personally insulted somebody on Reddit, either a human or a bot would respond with one of... I think we have six different strategies for deescalating that conversation, where we use things like a humorous meme or a confronting, "Hey, man, that's not how we act here," or we say, "Oh, I'm sorry, this looks like there's some confusion." Of all of the things that we tried, only apologizing mattered. All of the other ways that we tried to deescalate, they didn't make anything worse, but they didn't make anything better, except for apologies.

Libby:

An apology, in order to be effective, doesn't even have to come from the person who did the harm, that it can be effective at reducing the rate of personal insults even if a third party says, "Hey, I'm sorry that that happened to you. I'm sorry that this occurred in our space." So apologies are really useful because they do acknowledge that someone hurt, and it doesn't actually matter whether we meant to hurt someone or if we didn't, that they still feel hurt and so it's important for us to apologize. The question that was asked is why is that so hard? Well, in order to acknowledge that someone else is hurt, almost always, like if somebody stubs their toe... This isn't true, but usually it means that someone else messed up, and it's hard to say, "I messed up," or, "My friend messed up."

Libby:

I'm also a parent of a young child, and a lot of what I study in moderation and content moderation is just like parenting, that you have to be able to say, "I messed up," and to acknowledge that everybody's working on something and everybody messes up sometimes and that it still matters that we say, "I'm sorry that I hurt you, I will try not to do that again in the future," and that there are actually these two parts to apologies, so the recognizing that harm occurred and taking ownership of that harm but then also an effort to repair, and this comes from... Restorative justice framework is the legal framework

around it, but I think it's also just from toddler training, that apologies require repair, not just acknowledgement. Both of those things are really hard to do because they threaten our social standing to say, "I messed up."

Libby:

I think that we put a lot of pressure on each other that raises the stakes of an apology and that if we just get a little bit gentler to each other and acknowledge that everybody messes up sometimes, which is also hard to say... But there's even some research about how we are all the trolls on the Internet. If you catch us in a bad mood on the right topic, we will all be jerks, and so we have to acknowledge that sometimes people are jerks on purpose more often but everybody messes up. We socially have to be willing to both accept and enact apologies for that to get easier, but to recognize that it's hard to say, "I screwed up," is an important first step.

Anna:

I want to say, I am here for the podcast on Libby's life advice. I would also go to that. So a couple of things. One, we did get a question from someone about the Media Bias Chart that comes from Ad Fontes Media. I want to just give a quick shout-out to that because we actually had... Our last podcast episode was with the creator of Ad Fontes Media, and we got to talk about the-

Libby:

Oh, right on.

Anna:

Oh, it was fascinating. I wish you could've been there for the whole conversation. It was wild, just knowledge bombs dropping. So good. But there is an interesting connection to the conversation that we're having right now about how you're saying helping... You can do so much by even saying the apology, even if you didn't do the thing. Vanessa Otero was talking a lot about the way that we move people from non-evidence-based media towards evidence-based media and how that happens over time, and I wish we could do a mash-up of these two episodes because there are so many good nuggets here between those two for how we could move the needle on making that better for our world, talking about data doing good.

Anna:

So next question that's coming up... And I'm realizing that we're getting short on time. I wish this was a four-hour podcast, but sadly it is not. So a question that we have is do you feel like your work relies too much on the assumption of good faith from social media companies? I will address that that was a larger question that I have combined into a smaller question. But, again, do you feel like your work relies too much on the assumption of good faith from social media companies?

Libby:

I have not experienced good faith from social media companies, and I still get work done. So I don't think I'm relying on their good faith. In fact, I'm assuming that they will not act in what are the best interests of society and, instead, what we need to do... So when you look at how do I try to have public impact in the work that I'm doing, I do engage platforms. I'm happy to talk to them if they're willing to meet. But I talk to the press. I do podcasts like this. I work with advocacy organizations like the Center for Technology and Society because I don't trust platforms to do good things with the research that we

do or that they do. As Dory mentioned, earlier this week, we saw Facebook knows a lot about what its impacts are and it doesn't act in society's best interest for them or even perhaps in the individual interest. So a lot of my time is spent getting around what platforms want and what platforms want to be able to do.

Libby:

I think I don't expect them to behave well, in part because they're profit-driven companies. We've told them that profit and growth matter, so they're going to optimize for profit and growth. Until we change the stakes, why would they change their behaviors? So part of what we can do to change the stakes, if those are the levers I have to pull, the profit levers, then we can do things like educate the public to be less extractable on their platforms or we can raise a stink in the press. These are things that still work. If you want change in the world, you need to bring hearts and minds and the media, and so that's what you got to do. Then platforms will come along when they have no other choice, just like any other company, that they'll come along when they don't have another choice.

Dory:

A lot to think about. All right. I'm looking to see if any other questions come in through the Q&A. But I'll ask you while we're waiting, can you talk about one of the biggest challenges that you've faced in your work?

Libby:

Getting data in the first place is definitely the biggest challenge. Some of it is because there are technical limitations, like one of our questions was about the change in the academic researcher API that Twitter made last year. There didn't use to be a separate way for academics to get data from Twitter. We were bound by the same rules as individuals and companies. I mean, many of the ORs who are watching today are parts of academic organizations, and you know that we don't have millions of dollars around to just buy up some data for one study. So to put us in the same pot as a brand like, I don't know, Audi USA, they have a lot more money to spend on figuring out what social media has to say about them than an anti-racism researcher has to study anti-Blackness online. It wouldn't make sense to put us in the same pot. Of course, we weren't going to have data.

Libby:

So some of them are the rules, policy limitations. Some are technical, that the amount of data that we're talking about... Oh, I just pulled up... I have 95 gigabytes of data that are compressed that are the tweets from US and Indian politicians over the last three years maybe. So super compressed data, 95 gigabytes, that's more than ICPSR's whole collection used to be, so thinking about how much data that is and how you move it around, and that's just one project. I had a conversation with a potential depositor a couple of weeks ago who had 20 petabytes of data. The physical storage requirements for 20 petabytes is more than our server room. You can't hold this much data. And the cheapest... I'm pulling up a chart in my notes because I just had this conversation. Moving a petabyte of data from place to place in the cloud, the cheapest one I can find. Does anybody whose audio is on want to guess how much it costs to move a petabyte from place to place, so if I wanted to move a petabyte from ICPSR to a user, what that might cost?

Anna:

\$600.

Libby:

Anybody else?

Dory:

Someone says 10K in the chat.

Libby:

Somebody says 10K. Somebody says 600 bucks. The cheapest I can find is \$52,500.

Dory:

OMG.

Libby:

So a petabyte... And this is going from Amazon Web Services' Glacier archive, which is a petabyte object. Moving it from place to place, Amazon will charge you \$52,000 to do that, which is... Yeah. Right? It's a lot of money. Yeah.

Anna:

It's a lot of money. For anybody who's listening, you are hearing the speechlessness and you're not getting to see the wide, open, gaping mouth visuals that are just... It's wild to us, and we work with data, so this is an incredible amount of money that we're talking about, just for the ability to get access to one piece of the puzzle.

Libby:

Yep. And so if you're wondering, ORs, where is the social media archive or SOMAR, there it is. We're still trying. And the storage for a terabyte a month is about a dollar, so I'm paying \$95 a month of my own money to pay to store some data before we deposit it. But when you're thinking about we need to have two or three copies of data, you're up to three bucks per terabyte. The scale is just outrageous. It requires a completely different funding and storage and infrastructure than the data that we as social scientists are used to and that ICPSR is really good at. We're learning how to do that, and I'm impressed with how much we've learned as an organization and what we're able to change. But, yeah, massive scale data is coming for social science, and it's going to be expensive.

Anna:

So for everybody listening, ICPSR is working on this. We are working on... You heard the reference to SOMAR. This is going to be our social media archive. It's coming, but it's tough. So we're working on it, and hopefully by the time posterity listens to this, there will be an answer to that question. Thinking about time, I do have one last question that I wanted to ask. What is it that you wish people would ask you?

Libby:

Oh. I get this every time somebody does a live interview with me, and I feel like I always fumble it. What do I wish that people would ask? I mean, I think I would ask, can individuals ever do anything or are these problems too big? I think the answer is that there is no problem that is too big for us to solve together. We talked about apologies earlier, that that's something that we can do. We can say, "I'm

sorry someone else hurt you." We can say, "I'm sorry this space isn't what you need it to be yet." We can be more thoughtful in the way that we engage. But even those individual behaviors, if they're still just from individual to individual, aren't going to matter.

Libby:

We're going to have to do things together. We're going to have to put pressure on platforms together. We're going to have to put pressure on bad actors together. We're going to have to hold each other accountable, even when we're in bad moods and even when somebody says something that sets us off, and that when we're able to do that, whether online or off, then social media will have the chance to be the inclusive and representative and important place for social connection that I still believe it can be.

Anna:

Yes, Libby! Shout it from the rooftops!

Libby:

There was a question about if there's any data as part of a future congressional effort, so I put a link to the legislation currently sitting at the US Senate Committee on Commerce, Science, and Transportation. That's where you can find out what is Congress doing about this. There's a sister committee in the House that has a lot less legislation but is holding similar hearings. So if you want to know what is Congress going to do about data and social media companies, those are the committees to watch.

Dory:

All right. So I think we might be down to our last two questions. I'll take one of them. This is a question that everyone can answer, but first you, Libby, is what is your favorite brunch?

Libby:

My favorite brunch?

Dory:

Yes.

Libby:

Like the event or the food or-

Dory:

The food. The food that you eat.

Libby:

You heard me say I have a small child, right? I haven't had brunch in 10 years. Well, what would be appropriate for me to answer? It's been so long since I've had... The essential tension in brunch is sweet or savory, and I'm having that problem right now. I'm going to go savory. I'm going to go chicken and waffles with gravy because it's morning.

Dory:

I was just thinking chicken and waffles because you're right. You get sweet and savory at the same time and dessert and breakfast.

Libby:

You do. Yep. Yep. Without having to eat sausage. I'm not that into sausage. But I'll take a chicken and waffle, for sure, but it has to have gravy because gravy is a food group.

Dory:

Someone says, "Is there an inappropriate brunch?"

Anna:

I don't know that there is. That's one of the great things about brunch, right? It's like just feed your body when it's ready to be fed.

Libby:

Yeah.

Dory:

I'm hearing hash browns and hot sauce, depends on how many mimosas are involved, scrapple. Now, okay, I'm going to need to know what-

Libby:

What is scrapple?

Dory:

A Monte Cristo.

Libby:

Oh, a Monte Cristo. There's actually a restaurant in Ann Arbor that has a really legit Monte Cristo. Yeah, that's a good one.

Dory:

Someone tell me about scrapple, please.

Anna:

Oh, it's Philly food. I like it.

Libby:

What is it, a day-old cheese steak?

Anna:

I like us guessing what scrapple is.

Dory:

Uh-oh. Someone-

Libby:

[crosstalk 00:44:39]-

Anna:

All right. So narrating for the folks who are listening, scrapple is also known by the Pennsylvania Dutch name, Pannhaas. It is pork scraps and trimmings combined with cornmeal and buckwheat flour and spices.

Libby:

Is it like a hash then, or is it like a patty?

Anna:

This person says it's weird ground mystery meat fried, which sounds amazing.

Dory:

Right. Doesn't it?

Anna:

And for everybody who's listening, a big part of this is... So when we created Data Brunch, part of why we do this is because there are so many incredible people doing incredible things with data, and we just wish that we had time to sit down with you on Sunday morning and have a mimosa. I don't know if that's allowed to say, but that's what we were saying. How do we have these conversations where it's just a conversation between folks talking about things that we care about and things that are important? So thank you for joining Data Brunch. Thank you for being part of this and for sharing your expertise. If people were interested in learning more about your work or contacting you, where would be a good place for them to go to find out more?

Libby:

Sure. So libbyh, L-I-B-B-Y-H, on Twitter and dotcom and at UMich.

Dory:

Thank you, Libby.

Libby:

Thanks for having me. This was fun. And thanks everybody for coming to the Biennial Meeting. You're the highlight of our year.

Dory:

That's right. Thanks, everybody. Thank you for the awesome questions. This was a multi-sensory experience, I have to say.

Anna:

It was.

Dory:

It's awesome.

Anna:

All right. Well, that is the end of our episode. Thank you so much to everyone for being with us live and also for everybody who is listening to us.

Dory:

Oh-

Anna:

And if-

Dory:

Oh, sorry, Anna. This is one of those things where... Yeah. Go ahead.

Anna:

This is a live moment in a live episode. So if you aren't already, please do subscribe on Apple Podcasts or wherever you get your podcasts, and tell us what you'd like to hear. We'd love to hear from you. We have a feedback form on our website. Anyone who's live, message us in now, and do share your thoughts on social media using #DataBrunch. We love to get to see stuff pop up there. As always, thank you to the over 700 members of ICPSR. This podcast and the Biennial Meeting would not be possible without our ICPSR members. A shout-out to some of the representatives who are here live. ICPSR representatives make the world go round, and we appreciate you so much. You can get in touch with us by visiting our website, which is icpsr.umich.edu. You can email us at icpsr-podcast@umich.edu. And I want to give a special thank you to Scott, our producer, who makes us sound less bloopery than we really are in real life.

Dory:

Thank you so much, Scott. For everyone here live and for those listening at home or in your car or however you podcast, we'll be giving away some ICPSR swag to someone out there. Take a picture and tag us on social media using #DataBrunch, or send us an email at icpsr-podcast@umich.edu, and we can't wait to see it.

Anna:

I'm Anna.

Dory:

And I'm Dory, and thanks for joining us at ICPSR's Data Brunch.

Anna:

Woo-hoo!

Dory:
Yay!