Practical Significance | **Episode 38: New Year … New Initiatives**

Donna LaLonde: Well, welcome back to the February edition of *Practical Significance*. I probably say this every time, but truly this is a special edition for Ron and me. We're really delighted to have David Matteson as our guest, and we're going to talk about lots of new initiatives with David. But we'll start by just asking you to introduce yourselves to our listeners. And so, tell us a little bit about what you do in your day job(s)

David Matteson: Yes, it's many jobs these days. So first and foremost, I'm a professor at Cornell University. I serve as associate chair for the Department of Statistics and Data Science at Cornell, I also work in the fields of applied mathematics, operations research, as well as computer science. So, I do a couple of other things. So, with the ASA, I'm the new chair for the Business and Economics Section for 2024, and I'm excited about my continued work with that group in the section.

I'm a longtime member and longtime officer. I am the founding editor-in-chief of the journal *Data Science and Science*, and I'm excited to share updates on that journal today and in particular, highlight its new affiliation as an official ASA publication. The other recent news is that I'm the new National Institute of Statistical Sciences (NISS) director. We can talk a little bit more about NISS after we get to the new journal.

Donna LaLonde: Absolutely. So, yes, let's dive right in and talk about *Data Science and Science*. Can you tell us a little bit about how you see it differentiating itself from other publications? And then also, what are the key features that researchers can expect when submitting to data science and science?

David Matteson: Data science has evolved a lot over the last ten years, and different groups have latched on and taken a leadership role in that. I've seen statistical sciences more recently step up as well. It is kind of a combination of different domains that come together to compromise data science. And a big promise of data science and AI more broadly is to make scientific advances. And that's really what this journal is about, is new science, discovery, testing, and anything in the whole workflow or pipeline that can be enabled through these new tools.

I want to say a disclaimer that we define as a group "science" very broadly. So, it touches on the social sciences as well. And I work in all kinds of domains across ecology, hydrology, agriculture, physical chemistry, computer science. It's a lot of fun, and I think that's where my passion for this came in.

So let me say a little bit about what we're about. We aim to be a home for collaborative research that potentially spans multiple scientific domains and that's newly enabled through data science. We hope that we can recognize advances in data science itself, whether they're methodological or maybe new adaptations that are motivated or demanded by key scientific challenges. We also appreciate the importance of data visualization in scientific discovery and communicating those discoveries as well. So, some of what I think of as key features are about rewarding collaboration, and in particular, recognizing the type of advances that we make these days are largely driven by collaborative teams. It's a combination of folks who identify as scientists and others who identify as data scientists.

Ron Wasserstein: So, thanks, David. I'm going to dive down a little deeper, but when we were talking about your background and the various things you're involved in, you didn't think to mention it. So, I'm

going to be sure to mention that David is also the nephew of one of my best friends. So, this is a key factor. Really, if I'm going to be honest, David, the only reason you got invited to be on the *Practical Significance* podcast. Just kidding. Of course, I'm excited to have you on because I'm extremely interested in both *Data Science and Science*, which I'm going to ask you about a little bit more here in the next few minutes. And then also, of course, NISS, which, as you know, I've been involved with for many years. But with regards to *Data Science and Science* and those things that you were mentioning. So, what do you see as the opportunities in this new journal for our emerging scholars and researchers, people setting out at the start of their careers? Also, are there special issues envisioned for the future?

David Matteson: Yes. So, I think these do intersect. The special issue is kind of a way that the journal is taking some ownership of different collaborative domains, and it is kind of an ideal path for new researchers, in particular, to get involved with *Data Science and Science*. So, the current open special issues include data science and modern finance. The intersection of climate and the environment is the second one, and then more recently, data science and the brain sciences. And we've had a lot of strong submissions in each of these over the last year.

The interesting thing that the journal does in the special issue is the opening call to what then becomes a collection. So, a collection is kind of like a tag that helps label publications going forward. And for me as an editor, it helps provide some authority to our editorial board that we have good coverage and a good track record publishing in these different areas. So, some of the new initiatives in the coming year include the intersection of AI and the federal government. We're hoping to do a special issue on wastewater epidemiology.

That one's a little bit more specific, but I think that we can go a little bit deeper. High energy physics is on the short-term radar, a little bit longer, more intermediate-term. Something in space science is something I'm personally passionate about. And there's a big community, especially the astro statisticians, that could benefit from this new collaborative outlet. And then further down the line, I'd like to see some more work in ecology and the environmental sciences as well. In terms of new researchers, we're kind of an ideal fit because we see ourselves as very modern in our take on data science itself. I view myself as a data scientist from a statistical lens, but of course, some folks are coming from computer science or other fields, and we have an open call for a co-editor, and I expect that we get someone in place soon that complements my background and helps us fully encompass how broad the field is itself.

We are looking for new ways to partner, to kind of recognize and elevate new researchers as well. So, think about things like the awards that we give to ASA sections for new student research and trying to reach out to those award winners and invite them to submit to the journal. There are some other ways to partner with ongoing workshops and conferences, both with the ASA and, but then more broadly at other data science conferences as well.

Ron Wasserstein: So, you mentioned a co-editor, which is great and I know will be a tremendous help. But if listeners are interested in being involved on the editorial side, obviously they can respond to that open call for the editor position. But is there still a need for associate editors? I'm going to ask you if there is a need for reviewers, but I already know the answer to that question because everybody needs reviewers.

David Matteson: That's right. We do have an open call for a co-editor, and there's a search committee putting that together now and reviewing applications, but that'll stay open for the next couple of months. So do reach out. We are constantly looking for associate editors, and the ideal profile is someone who has a strong research track record of doing deep collaborations with scientists.

At the same time, I'm looking for what I think of as scientific unicorns. These are scientists that are also broadly recognized as data scientists, whether they're self-taught or otherwise, they're advancing and have full credibility as data scientists themselves. And I find those people through their papers.

Oftentimes I reach out and see where it goes, but we're looking for a good balance and try to reach both parts of the community. Junior folks, I think that once you've started publishing your own papers and then it's an appropriate time to start doing some reviewing yourself. We continue to expand our internal list of potential reviewers. I will say through the pandemic, the number of folks who decline review invitations has accelerated.

So, we do put out a lot of requests but we always strive to get at least two independent peer reviews for all of our submissions. For the associate editors, I am just looking for diversity overall. So scientific diversity and even more broadly, geographic diversity. This is a big world we live in, and we want to reach all the communities that are helping us advance it.

Ron Wasserstein: Thanks, David. I just want to make a quick follow-up. You mentioned junior folks, and you are eager to get those people involved. I was just thinking, because now I've known you for a few years, I remember when you were a junior folk, and that goes by really quickly, doesn't it? And now you find yourself in mid-career, and no definition of mid-career would define where I am right now. But yes, time does move by rather rapidly.

David Matteson: It does. And especially when you get busy, it flies. But I think you should keep an open mind throughout your career. A lot of us say "yes" very often. Eventually, you have to learn to say "no." So, I think what we're really about is building community, and if that's a good fit for you and you have some broad interests in collaborating, then this is a community that would be very welcoming to you.

Donna LaLonde: And David, just to follow up on that, if folks are interested in serving as reviewers, they should reach out to you.

David Matteson: Connecting with me is easy. And otherwise, on the web pages, there are links to the editorial office as well. I think direct to me is ideal.

Donna LaLonde: Okay, perfect. And we'll be sure to put all the web pages in the show notes as well, so folks can hopefully take advantage of those opportunities. I want to stay in the publishing world for a little bit. You will, in your new role, be joining the ASA Committee on Publications.

One of the conversations, of course, that that committee has had, as well as probably across the scholarly community, is how artificial intelligence will influence and shape the landscape of scholarly publishing. And I'm really interested in your take, both in the short term and then get your crystal ball out and say a little bit about where you think it might go in the long term.

David Matteson: Yes, it's wild times. It demands a lot of attention, whether it's just from the headlines or if it's in faculty meetings. For me, this comes up in the publishing world very acutely as well. So, initially, there was a bit of a panic, to be honest. Are we going to be flooded with submissions that are generated by AI? That didn't happen. And at the same time, of course, there are AI tools to screen submissions as well.

So, the initial take is that as a community, we have to have policies and best practices in place. And I think the ASA has shown a lot of leadership in this space, giving it deep thought. I think that we don't necessarily need to rush these, but as we roll out the policies, we need to continually review them, because this is something that is rapidly changing over time. There are pretty large ethical concerns both in academia, but more and more so, we see when things aren't held to the highest standards, the fallout in the business world, in publications. Of course, for the most part, it's not seen that AI is a legitimate collaborator, and the main reason is that it can't be held accountable for what's written.

And that's something that if you're using AI tools, I caution you to review the official policies, make sure

you're doing it the way that is currently accepted, and always disclose what you've done. And sometimes, as we try to push boundaries, it's the way to safeguard both yourself as the author and the reader as well, so that they're not misled. And that's ultimately, I think, what's most important because it's that path that advances science.

In the intermediate term, I could see AI tools providing something as a copilot, the way that they've started for writing, the way that they've started for coding. Through NISS, we're putting together a renewal for our writing workshop. It's a longstanding writing workshop that's run, it's got a couple of virtual sessions and then an in-person component at JSM each year. And AI is kind of the big new module that we're expanding to in the next three years, and it'll try to touch on what are the ethical implications.

But also, everyone's curious, how do we do this? Because it turns out you can also waste a lot of time trying to use some of these tools. It can generate nonsense pretty quickly, too. So, how do we get started using some of these things, and how do we wade into them cautiously and ethically?

Ron Wasserstein: As it turns out, David, AI-generated our entire last podcast, and nobody seems to have noticed. So, we're all good. I'm eager to talk about NISS. Thank you so much for taking on the role of director of NISS. And so, tell our podcast listeners about what NISS is, and briefly about your goals and visions. I only say briefly because I've heard you talk about these, and you have a wonderfully large set of goals and visions for NISS. But sum that up for us, and also, and maybe most importantly, talk about how listeners can be involved.

David Matteson: Absolutely. I will try to keep it concise. The first thing to recognize is that NISS has been around for thirty-plus years and it's part of our community, and I'm going to try to raise the visibility of it. So, what is NISS? NISS is a research nonprofit, and NISS was founded in large part by the ASA, the IMS, and the International Biometric Society.

So those groups got together in the very early nineties and through a report commissioned by the National Science Foundation, said that statistical science could make large advancements if there was a national institute. And this is the paradigm that was envisioned and was fulfilled. The other founding partners were the Triangle-area universities, the state of North Carolina, and other entities in that location. So NISS was an institutional partner on that with the three Triangle universities as well.

So back to NISS's broader mission. It aims to make advances across the profession of statistics and now the data sciences, with the hope of spanning academia, industry, and government. One of the key ways it does this is that it has a broad affiliates network, so different units, organizations, or departments can become affiliated with NISS. And what that signals is that they're part of the NISS organization. They're helping lead the activities that NISS does. So, what are some of those activities? One of the strengths of NISS recently, largely because of the pandemic forcing everything to go virtual and then maybe more hybrid now, is online programming.

So NISS offers a wide selection of not just research programming, but expert panels on hot topics and a lot of career development series. So, this could be for folks just graduating, or even folks before that are thinking about what kind of graduate school is the right fit.

And again, the hope is to serve the entire profession. Ultimately, I see NISSA's central mission, though, as a research institute. So, I want it to be a partner. And when there's a need, be a leader to help identify, catalyze, foster, and lead high-impact research along the way, we hope to offer exceptional training opportunities to supplement what folks get in the traditional academic setting. So, with that, a couple of other ways to get involved is through some of the new research priorities that are being rolled out this year at NISS.

So, the first one we've already touched on, and that's the intersection of AI and statistics. And this is something that I'm quite passionate about, enabling our statistical sciences community to take a leadership role. And part of this is through a new series, first virtual and then in person, on the use of AI in federal statistical agencies. Simultaneously, we'll be rolling out a complementary series that considers what are the core statistical concepts and principles that enable best practices for AI in the industry. A couple of other priorities. You'll be happy to hear that data science and science is a strategic priority for NISS in the coming years.

And a large part of that is trying to broaden the umbrella that NISS has historically touched and also move more into the sciences, broadly defined. I want to see NISS move into what I call (BML) biological machine learning. This is a big nod to our biostatistics friends and collaborators. They're maybe the initial folks who took the lead in data science itself. And I think that it's such a wonderful set of applications that we saw through the pandemic have such real-world importance, especially in public health, but more broadly in scientific discovery.

And then finally, data visualization has been a strength of NISS for a long time, and I know that it's been a priority for a lot of the events that ASA and many of its member sections have done. So, I want to see that be strong, and I'm hoping that we can partner on things like visualization contests in the future. And then finally, if I didn't emphasize it enough, I think research, training, consultation, and in particular collaboration are things that NISS wants to partner on in the future.

When I say training, I think about things that aren't happening in a traditional academic program. So, I mentioned the writing workshop. A lot of programs don't do technical writing. Grant writing isn't traditionally part of a lot of statistical programs, although I think biostat degrees have done a good job implementing that. A lot of coding is kind of outsourced to the students themselves. So, my students earning a PhD in statistics often spend evenings and weekends studying LEAP code and otherwise. But this is again another place where, as a shared research resource, NISS can partner and provide leadership for the entire profession.

Ron Wasserstein: So, David, thanks for that. I'm going to follow up with one thing. You talked about the importance of statistics being involved in AI, and of course, you were brief, just like I asked you to be. But in a recent meeting, I was in with you, you spoke very passionately about why statistics needs to be involved in AI and what your concerns are with regards to maybe statistics dropping the ball and not getting involved. So, could you just expand on that a little bit more?

David Matteson: So, this intersection of AI and statistics, I think, is something that NISS can provide real leadership on. And it's a little bit difficult because I haven't bought in myself all the way, but certainly it's not going away. There are good uses and there are ethical concerns and centralized leadership and drawing across our affiliates network, we could provide a lot to the profession.

I think the reason I'm so interested in it is really how much I felt statistics got left out of data science initially, and I don't want that to happen in AI. So, if I can help, really have statistics, have a big voice in the future, because this is going to affect all of society, not just research we had previously done. This was before my time, data science essentials for business, and there was a ten-part series, and we do something of that spirit, but maybe flip the script.

So, consider a broad audience of people who are doing AI, ML, and data science, but they maybe don't have statistics training. So, let's present a series that provides the core statistical essentials that provide the foundation for these things and give those folks who didn't have strong statistical training a lot more confidence in the way that they're implementing things and maybe some tips and tricks along the way. So, I think as a profession, that is a good service that we, as statisticians, think is important.

So, let's say why and target that pretty large community. A different angle is let's enable the new

researchers. They've got a clean slate. They don't have as much, I'll say, baggage in terms of "we do it this way or the highway." So, enabling them is kind of our core key to success, both through NISS, and more broadly as a profession. And that's what I hope to do with that.

I will say just another way to get involved with NISS, especially for young folks, is the graduate student network. This is a great network of students. It's a student-run network that has mid-career folks as the steering committee or advisors. They put on a conference every year to showcase their own work and it's just fantastic that they run this themselves. There are over two hundred members. It spans all kinds of areas of statistics and data science, master's students, PhD students, and folks who have just graduated or welcome us down. And I really would like to see this expand and just be more broadly a new researcher network to hit all those early career folks.

Donna LaLonde: Thanks, David. That's great. It's making me think that you and I need to find a time to talk because there's a community of really young and energized postdocs that are starting a new ASA community on postdocs and so see lots of opportunities for synergy. So, more to come on that. And then the one final thing on statistics and data science and AI, I might lose my job if I didn't shout out to the board statement on the role of statistics and data science and artificial intelligence that was approved in August.

Okay, well, as we come to the end of our conversation, we are going to shift gears and recognize that with all the jobs that you have, you probably don't know what free time is, but we always like to ask in your free time, or little, as that probably is, what are you listening to? Watching? Reading? What's on your TBR? Share some good things for us to do.

David Matteson: It is pretty limited free time these days. Believe it or not, I read statistics and machine learning papers for fun sometimes. But one thing that does help me unplug is science fiction, and somehow, I need to go to another reality to shut off my brain. So, I've been rereading some of the early Asimov – the *Foundation* series in particular, and those were things I got into in middle school and high school, and it's fun to kind of rediscover them. And some of this stuff is in movies and a TV series now.

I have a rule with anything like that, I have to read the book before I watch it. The other thing that I've been reading with my kids is *The Lost Art of Reading Nature's Signs*. And this is a fun book. It's something I got for Christmas just recently. It's by Tristan Gouli. I get a kick out of it because I'm big into nature and we're a scouting family as well. It brings in a bit of mathematics and that's fun for me to share with the kids. It gives me another perspective when I'm out going for nature walks.

Donna LaLonde: Oh, that's fantastic. Thanks for sharing both of those with us. And thank you for joining us. This has just been so much fun to talk with you and learn more about *Data Science and Science* and NISS. And now, as we come to the end, our traditional conclusion is for me to turn it over to my colleague Ron for Ron's Top Ten. So, Ron, back to you.

Ron Wasserstein: Thank you, Donna. About three years ago, I prepared for the Washington Statistical Society a "Top Ten List of Dubious Statistical Awards." As this is the time of year when many statistics committees will soon be reviewing nominations and making selections, we have updated that list because the *Practical Significance* podcast is always trying to help its listeners.

So, here is a "Top Ten List of Awards We Hope You Will Never Receive":

10. The Evidence-Free Decision-Making Award
9. Most Ambiguous Survey Question
8. Outstanding Grant Underbidding
7. Acronym Development Excellence (ACE) Award
6. Anecdotal Evidence Merit Award

5. Big Data but Small Insights Medal
4. Commendation for Lowest Response Rate
3. Most Inefficient Use of Administrative Data Prize
2. Lack of Transparency Using Statistics (LOTUS) Award

And the number one award you don't want to receive:
1. The National Prize for Massive Privacy Violation

Well, that's it for *Practical Significance* for this month. Thanks for listening. Join us next month as we continue the conversation.