**Season 1 Episode 6 Dr. Julia Lane: UMETRICS and the Survey of Earned Doctorates**

Dr. Nancy Potok

Hi. I'm Nancy Potok, and you're listening to Show US the Data. Today our guest is Dr. Julia Lane, a well-known NYU economist and researcher and also the co-founder of the Coleridge initiative. Our machine learning tool has identified her as a top researcher using data from the NSF National Center for Science and Engineering statistics in the UMETRICS program. We're very excited to learn more. Welcome, Julia.

Dr. Lane

Thank you, delighted to be here.

Dr. Potok

Julia, you've done a lot of work with the Survey of Earned Doctorates. Tell us what was exciting to you about this data? That is, what was the real-world effect of your research?

Dr. Lane

It really helped answer an important question, How and why does federal research funding shape scientific research? There hasn't been good data to answer that simple question. Which means that the federal government has been flying blind about decisions about where to make those research investments. They're effectively dealing with a black box without data. I wrote a piece for Nature called “Let's Make Science Metrics More Scientific”, and argued that new types of data, combined with the survey data from NCSES (National Center for Science and Engineering Statistics), could answer that question. Our paper showed how to link federal funding to the people who create ideas, who create innovation. And the basic idea was, if you could tie research funding with who is being funded, in which field, and with which results, you unpack that black box. The results showed for the first time which agencies were funding the next generation of scientists, doctoral recipients, in which research field, and what the funding was doing.

Dr. Potok

Wow, that sounds exciting. Maybe you could elaborate for us a little bit more on how your work helps the American taxpayer?

Dr. Lane

Well, the federal government spends over $160 billion a year on research and development. Congress, the President, and a recent Nobel laureate, Paul Romer, have emphasized that the importance of that investment is not just $160 billion. It also is critical to stimulate economic growth, innovation, and high wage jobs. But we didn't have the data necessary to inform how those investments should be made. I think about it in sports - we've seen how data about baseball can transform losing teams into winning teams, businesses who have data out-compete businesses who don't make use of these data. So what we're able to do now is just like in the private sector or in sports, we can help the federal government use their data more effectively, to transform the productivity of our economy. That's a big deal.

Dr. Potok

Yeah, surely. I love that Moneyball example. Tell us, based on your hands on experience, what would be helpful for other researchers to know about these data, particularly if they're considering using them in their own research?

Dr. Lane

Well, they can go to the Survey of Earned Doctorate site at the National Science Foundation. Or they can go to the Institute for Research on Innovation and Science at the University of Michigan. We ran a terrific conference on the value of science in June. And Paul Romer gave a keynote. He's the Nobel laureate. Toby Smith represents the American Association of Universities, and explained how these data enabled us to go beyond anecdotes like, well, if you spend more money on research, you get Tang, which is kind of the classic joke. And there's also a paper on building a data mosaic of the type we're discussing. I should also call out the two papers by Kaye Husbands Fealing. It's on the value of such data to understand diversity and equity and research funding.

Dr. Potok

Great. So are there other datasets beyond those in UMETRICS you think would be helpful to access or maybe to link to the Survey of Earned Doctorates data? Or, based on your hands on work, is there something you'd like to see improved about these data?

Dr. Lane

Well, you know, the wonderful thing about the new types of data that we have in the new technologies is that there are endless possibilities. One of the foci given the concern about diversity and equity is bringing in more minority serving institutions. NSF and Michigan staff are really focused on making that happen. And we're also working with the United Negro College Fund and excellence here to provide richer information about the pathways that black and brown students take as they get research funding. Building links with other users is critical. The Show US the Data Conference highlights how others have used the data so we can learn from their work using the machine learning tools that you referenced at the beginning.

Dr. Potok

Julia, thanks so much for taking the time from your research to share these valuable insights with us. More information and relevant links are in the show notes for our listeners. Until next time, I'm Nancy Potok with Show US the Data.