

Privacy Sigma Riders Episode 6: Data Mashups and Missteps

What Norwegian Death Metal, Food Pictures, and My Morning Commute Can and Cannot Tell You About Me

Host Michelle Dennedy discusses big data and privacy issues with Hal Stern, vice president of IT engineering at pharmaceutical giant Merck.

Michelle Dennedy: You can learn a lot about a person from the digital trail they leave behind, but what can Norwegian death metal, pictures of food, and the mundane details of my next guest's morning commute possibly tell you about him? What are the implications of this type of big data mashup analysis on personal privacy? Stay tuned, or you may just get the wrong impression.

Cybersecurity, data protection, privacy. You like to stay ahead of the curve, and listen to experts who are leading the way in deriving greater value from data with a more organized approach to data privacy. You're like us. Just a few deviations past the norm. You are a Privacy Sigma Rider.

Hi everyone, Michelle Dennedy, again... chief privacy officer at Cisco. Today, I am honored, delighted, and excited to talk to a very dear, old friend. The friendship is old, not the man.

Hal Stern: Well, both, actually.

Michelle Dennedy: Hal, who rarely is stern. Hal Stern. Hal, will you introduce yourself?

Hal Stern: Sure, and thanks, Michelle. I am the associate vice president of IT engineering for Merck Research Labs, which is the research and development arm of Merck, known as MSD, outside the United States and Canada, a 125-year-old pharmaceutical company. And I'm responsible for helping us build software systems to go explore chemistry, biology, product safety, and product development.

Michelle Dennedy: So you guys probably encounter just a little bit of data every now and again?

Hal Stern: Yes! Absolutely. And, I would say, increasingly, it's a business which is being defined by, and driven by data. But, at the core of it, I think it goes back to

tenets that George Merck had in mind, when he founded the company, which is, it's really about patients, and it's about respect for the patients, and the patients' privacy, and treating everything about the patient with the respect to which it's due. I think that has certainly permeated our view of the data as it has grown in size and scope.

Michelle Dennedy: I think you can all tell, this is exactly, Hal, why you're here. I mean, Hal is one of the original Privacy Sigma Riders. You know, I can't even remember, when did we actually meet? It was like, a thousand years ago, I think.

Hal Stern: The first digit in the year might have been a one, so, a thousand years ago, or so, literally.

Michelle Dennedy: Something like that, and I remember, very early on, so we were, at that point, calling the cloud, "the grid." And Hal said, we were in a business meeting, once upon a time, and we were talking about a business model of a very, very large, multinational retailer, actually, and what you said to me has always stuck with me, that, "What you see on the shelf, isn't product placement. It's an auction." And it's an auction that's paid with data, where people will look at where you sit on the shelf, and what level, and how tall human beings who are walking through the retail meat world reality will place products will depend on how they are pulled through the entire system.

And it really informed my thinking then, and still now, of, like you're saying, Hal, is there's a lot of data, and that flow, and how it actually originates with the outcome of patient outcomes, in your case, and retail outcomes, and business outcomes. So that's kind of the big, messy world of transnational, and transgalactic data, isn't it?

Hal Stern: It is, and I think that the physical space world of shelf space, now, is what shows up in your Twitter feed, and what ads you're shown, and what stories are sponsored, and what search results get promoted to you, what products are recommended for you. And as we conduct more and more of our business online, I think, we want to be sensitive to, essentially, the way in which we show up.

Michelle Dennedy: Yeah.

Hal Stern: And it's becoming more complicated, because, like it or not, personally, we are generating an awful lot more data. Sometimes explicitly, and sometimes implicitly.

Michelle Dennedy: So, what does that data say? I mean, so, Norwegian death metal. Like what has all that got to do with all of this stuff?

Hal Stern: So let me start with a very simple example. I have about a 40-minute commute to work every morning. So, if you look at my phone, it's a fairly accurate GPS

sensor, so it will tell you where I am. And if you measure position over time, if you took physics, you know that that's velocity. Now you know how fast I'm going, on average. And you can tell, the same way Google Maps does, with its traffic analysis, tells you roughly how fast I'm moving.

So, by the way, a lot of people don't realize, if you are using Google Maps, and you're looking at traffic, it's getting that by tracking how fast your phone is moving. And so, obviously, what happens with a lot of people is, you're in traffic, you're stuck, you're angry, and then traffic breaks up, and the first thing you do is you toss your phone back on the passenger seat, obviously, because you're a safe driver. Which is why hysteresis of where the traffic actually breaks up, because you stop sending valid signals back to Google, in terms of your speed increasing like normal.

Michelle Dennedy: So what's a hysteresis? I like this. It sounds like a medical procedure, or a rash. I need a hysteresis!

Hal Stern: That things don't respond, think of gas prices, right? They go up in a hurry when there's any sort of tension in the oil and gas market, but they come down very, very slowly. You see the same thing with traffic, right? We tend to think that traffic's a little bit more exaggerated, simply because it's limited by the precision with which we can measure it.

So if I look at this, I have my phone, that's telling you, pretty much where I am, and how fast I'm going. It's also a media streaming device, so you can get a sense of what I'm listening to. Oh, and it also acts as a gateway for my personal fitness trackers, which give you things like my pulse oxygen level, my heart rate.

So now you know that I drive 70 miles an hour on interstate 287, while listening to Meshuggah, and Opeth, and other great things, and some things that are in eleven-seven, and other-

Michelle Dennedy: Can you sing a couple tunes?

Hal Stern: No, no, I am legally not allowed to sing. It's very, very bad.

Michelle Dennedy: But he is a musician, just for all you fans out there.

Hal Stern: If you put those things together, now it paints a picture of me.

Michelle Dennedy: Yup.

Hal Stern: Now, the question is, am I, with an elevated heart rate, listening to this rather angry music, driving faster than average, is that a dangerous thing? Or am I singing at the top of my lungs, and having a wonderful time in my car? When I get to work, I'm slightly more relaxed, and fortunately, for people in meetings with me, I'm a little bit talked out.

- Michelle Dennedy: Right.
- Hal Stern: And a little bit hoarse. And it's actually a good thing. I think that's the question that I'm generating all this data. Is it interesting about me? And part of it, I think, very much depends upon the context for the data.
- So, if you have a fitness tracker, you can get a variety of different views of what people do with that data. So I'm an over-50, overweight, over-caffeinated, over-sugared person.
- Michelle Dennedy: He looks good, don't listen to him.
- Hal Stern: I have a body built for radio. The data on my fitness tracker will tell you how many steps I've taken, and how many flights of stairs I've done, and, for me, it's, am I basically just getting up off the couch enough? For someone who is a health enthusiast, who is a runner, it's much more of a competitive situation. It's really, are they giving a proper view into what they consider a, not just a minimum level of activity, but a rigorous level of activity? Same device, same data, but interpreted in two very, very different contexts.
- For me, wow, I did 8,000 steps! And for someone else, it's gonna be, "I only did 8,000 steps, it was mostly walking." I'm thrilled; they're upset.
- Michelle Dennedy: Right, so it's collection, and context, and application.
- Hal Stern: Yeah. And so the question, then, is, if I hit 10,000 steps, do I say, "Great, I'm gonna have a donut?" Or, I hit 10,000 steps, that's really good, I actually feel better.
- Michelle Dennedy: Right.
- Hal Stern: And, I think this gets into the view we have, now, of just how much personal data do we have? There have been some industry proposed studies, I would say, that try to gauge the overall volume of data we generate, and I tend to look at just a couple narrow categories.
- If you're fairly active on social media, you generate 100 megabytes of data in some reasonable period of time, that's a lot. Even if you're a responsible, and fairly avid user of Twitter, at 140 characters at a time, it's really not that much data.
- Michelle Dennedy: Right.
- Hal Stern: If you use a public email service, and you have 100 gigabytes of email, that's a lot.

Michelle Dennedy: Right, and this is just what you're putting in. As you're saying, there's a lot of observational data, and machine data that's coming along with this, for the ride, right?

Hal Stern: Right. Oh, my genome is probably 300 gigabytes. So I can go and get a large portion of that sequenced, and looked at. And again, the part of it that's unique to me is maybe two-and-a-half to three percent of that. So it's a few gigabytes. If you believe that your gut microbiome matters a lot, argument that that's probably 10 times the size, so maybe you're looking at three terabytes of data, there. Your electronic medical record may also get into the terabytes as, again, it becomes more rich with images, and video, and, essentially, digital measurements that have been collected over a period of time.

But we're still talking small numbers of terabytes. I mean, this stuff fits on a disc drive, today, and you could carry it around with you, if you felt the need to go do that--

Michelle Dennedy: And it sounds like a lot of these things are doing a couple of things. They have state, they have volume, they're going over networks, and they're being interpreted somewhere, somehow. This is a lot of complexity.

Hal Stern: Right.

Michelle Dennedy: So you're a technologist, you're coping with a lot of this complexity. How does this come back to healthcare, and really dealing with, where is the patient in all of this? Is there still a human, left in a pile of metadata?

Hal Stern: Oh, absolutely. I think that is, at some level, that is the critical thing. If I look at the largest volume of data I produce, it's pictures. And, in my case, it's concert pictures, and pictures of food. So, you could just look at my phone, absent me, and get a sense that, first of all, given that I go to a lot of Phish concerts, there's probably chemistry involved that I shouldn't be doing, which I don't do, but you would infer that, by looking at the concerts I go to.

You would also realize I take a lot of pictures of salad with bacon on them. You'd say, "Wow, [he] really eats way too many cured pork products." Well, again, hopefully not enough to be a negative impact to my health.

Michelle Dennedy: And there is salad involved.

Hal Stern: There is salad, there are vegetables in there at some point, usually hidden by that fine cured pork product.

Michelle Dennedy: Those are the ones for your wife that, "Look, honey! I'm eating well."

Hal Stern: But let's put this together. And, I think about this, again, in a very patient-centric way. Healthcare, historically, has been something that happens to you under

duress. It happens to you when you're sick, you're injured, or a family member is sick or injured. You're upset, you're in pain, and you have to go somewhere and have something done.

Very rarely do we think of healthcare the way we think of media today, which is, you're in the middle. And there's this range of options of things you can use to inform yourself, and explore, and become self-aware, and consume it. And I think that that is the option we have here, is to figure out, how do we use the context of all this data to put the user in [inaudible], to really design a better experience. [crosstalk]

Michelle Dennedy: So, if you're well, you can be more well, and if you've got something that's unusual, you can at least find a community.

Hal Stern: And certainly understand other things, like there are sites now, Wiser Together, things like Patients Like Me, that allow you to aggregate your experiences, the way you experience healthcare, the way you experience symptoms, with other like-minded, and again, people in a like context. So I think that that's one of the great options ahead of us, now, is we think about how we apply this wide variety of digital technology to thinking about consumers in healthcare. It's going to be around, really applying some design thinking to it.

Michelle Dennedy: So how, when you're designing in, I've heard you talk about the three P's. Can you tell us a little bit about the three P's?

Hal Stern: Again, so you look at this data, we say, "Wow, there's a lot of it, and it's really, really rich, and it's all over the place," and people like to talk about volume, and veracity, and velocity, in terms of how much data, and how quickly it's moving around-

Michelle Dennedy: There's a fourth V, too, right? What's the fourth?

Hal Stern: Volume, veracity, velocity, and the other V. We're gonna, that's gonna get edited in post.

Michelle Dennedy: I can't remember the other V, but, victory?

Hal Stern: But I tend to think of, instead of thinking about volume, and velocity, variety, and veracity--

Michelle Dennedy: There's the other V.

Hal Stern: How rich is it? How much of it is there? And how fast is it moving? I prefer to think of this in a contextual way, which is three P's. It's providence, where did it come from, and how do you know how certain you are of a source? Second thing, it's propagation. Where is it going, and do you really want it to go there? And do you want all of it to go there? And those two things, together, take us to

the third P, which is privacy, which is to understand that this is not an all-or-nothing thing. That we may selectively disclose pieces of information, where we think they're gonna be used for our benefit.

Michelle Dennedy: But I think there's something that's running through all of your work, here, Hal, which is over time. So, Patients Like Me, great example, and not to call out any one company, but you go on there, when you're first diagnosed, and you want to share everything. I've got this, I've got this, I've got this, I've got this, and it could be, over time, and has been, in my experience, that, over time, once I've dealt with my condition, I don't want those early panic days, really, shared. Over time, my consent has changed, my feelings about the community might have changed, and my awareness of my own condition has changed.

So, I like your analogy of my fear of, and you'll have to explain this one, cause I love it. "My fear is that I end up on page 41 in a book of disgusting diseases." This is a quote from you, Hal. Explain.

Hal Stern: So, I developed a highly unusual form of psoriasis, after being exposed to strep throat--

Michelle Dennedy: And I don't want to call you out on your specific privacy disclaimer.

Hal Stern: No, it was the kind of thing where you look at it, and go, "Wow, that's weird." And I went to see the dermatologist, and I've never seen such a look of glee on the dermatologist's face. It's always a worrying thing, when you show up to the doctor, and it's like, "Wow, I've never seen that before!"

Michelle Dennedy: You thought our lives in technology was tough, I mean, a skin rash guy, come on, you gotta give him some joy.

Hal Stern: And, what was amazing to me is that, through his knowledge, and his execution of clinical practice, was able to say, "This is exactly what it is. It's this rare form of psoriasis." And a week or two later, I was fine. To me, the interesting point is that we talk a lot about technology and the digitization of healthcare. At the end of that story, it was the particular predictive capability, the particular clinical knowledge of this practitioner, that prevented it from getting a lot worse. Because the end state of that particular form of psoriasis, is that strep throat virus ends up in your heart and kills you.

Michelle Dennedy: Wow.

Hal Stern: Left untreated for a long period of time. So, I look at that, and I say, the fact that I actually dealt with this in the real world, in the analog world, had a very, very good result for me.

- Michelle Dennedy: So what you're dealing with, when we're systematizing these things, and creating, and curating this data, is you're basically digitizing kismet. He's giving me the look, of like, "What did you just say?"
- Hal Stern: Yeah, that's some cultural reference.
- Michelle Dennedy: Well, I just think that, like, here you are, happening to walk into the right doctor, that happens to have this piece of knowledge in his brain, or her brain, based on happenstance of information, being in the right place, in the right time, with the right data. It's like, once we get this going, and we have data that has privacy privilege propagation, provenance, all covered, isn't it possible that we can have more physicians that have access to these, kind of lucky breaks, or what I would call kismet?
- Hal Stern: Well, I think so. This is certainly, I think part of the goal is, we think about, how do we use digital tools to improve diagnostic capability. To do parametric, and I think that this is the risk, as we talk about all of our personal data, that we leave in a trail behind us, in our digital wake, if you will, is, how do they get put together in deductive ways? And, if you look at, I would say, the state of machine learning, today, it's very, very good at making deductions. And sometimes, it will take leaps that we didn't see. And sometimes they're obvious, and sometimes they're less than obvious.
- Michelle Dennedy: Now, is that better prediction of the past, or is it a better prediction of the future?
- Hal Stern: It's deductive reasoning, and I think that's part of the issue, is that, a lot of cases, you want to find out, what is the most important facet of it? What are the things you're looking for? So, I think the opportunity here, for machine learning, in this sea of data, is to help us understand what's useful. And help us understand what is most telling.
- I mean, you've certainly seen this case where you Google for a particular product, and, next thing you know, you see Amazon ads for it in your other social media feeds, because, clearly you have an interest in it. What I find annoying is, after I buy it, I still get the ads--
- Michelle Dennedy: Yes, it still stalks me!
- Hal Stern: Well, because there's no way of saying, "Okay, I'm no longer interested in that." It could be construed as a latent intent, right? That, the fact that you were searching for it, you must have some intent to go acquire this. You can't withdraw your intent later on, there's no obvious way to go do that, aside from using incognito browser windows.
- I'll draw a stranger conclusion, here. I mean, if you're on social media, and you're sitting in a doctor's office, you've just given a hint that you're in a

particular place. Now, again, there may be other things. There may be a Starbucks in the building, and there may be a lawyer's office upstairs, if it's professional building, and it doesn't necessarily mean that you're at this particular doctor's practice. But, the fact of the matter is, you're somewhere that's not your usual spot. You're not at work, you're not at home, you're not in your usual corner café where you do your posting.

What does that tell someone who, again, is interested in monetizing that data? And I think the critical issue, here, is to really, first of all, understand what data you're generating. And the second thing is to, in a larger sense, to really understand, how do we use the data in a proper context, with the intent that was given with it? And I think this is the root of, we talk about privacy, and consent, is really tying those things together. It's respect for the patient, it's respect for the patient's data. It's respect for all the data that we happen to collect, and to use it, very much in ways that are, in a lot of cases are regulated, they're regulated by things like HIPAA, and are controlled.

Michelle Dennedy: Yup. Yup. And I think, worldwide, I mean, the one element of data that crosses every single culture that I've ever encountered, is health data is always a sensitive, and obviously deeply personal category.

I want to pivot a tiny bit, before we close, because I know that you are always on the frontier of new technology, and I think one of the things that, if you are a technology junkie, as am I, you would think that Blockchain cures cancer, gets rid of cellulite, it makes you prettier, and solves halitosis. So, what is Blockchain, a little bit of a primer, and where do you see it fitting in this magical world of healthcare information, and patient centricity?

Hal Stern: Oh, great question, cause I think Blockchain, right now, is certainly one of those things, it's getting an awful lot of press right now, whether it's people using it as a funding mechanism, or, a lot of its historical associations with bitcoin, digital currency, that--

Michelle Dennedy: And we trust our currency to that sort of platform, it's interesting, but would we trust our health data?

Hal Stern: Well, the idea of Blockchain is, quite simply, it's a publicly accessible ledger, as anybody can go look at it, where the transactions are private, but they're verifiable. So, if you and I were to conduct a transaction, I mean, I'm gonna give you five dollars, I can then prove that I gave you five dollars, and you can then prove that you got it from me, and, because the transaction chain is public, anybody else can go in and say, "Yes, person A gave person B five dollars." We don't necessarily have to disclose our identities.

It becomes interesting, I would say, in cases where there's value in having a public ledger, there's value in having the identities hidden on the public ledger, and there's value in being able to split a transaction, as to offer partial sharing of

information. I think there's been a lot of interest, now, in things like, how would you use Blockchain as a healthcare data exchange? How would you use it to, again, propagate data with an amount of certainty around it, and then, again, what sorts of new, and interesting consent mechanisms come about as a result of this, because you are now able to go have a measure of privacy, and secrecy around identity, there?

At the same time, there is a set of new challenges that rise up out of this. One is, now, do the people who are really, really good, today, at using machine learning to assemble our location, and our queries, and our purchasing preferences together, to build a picture of us as a consumer, do they start to build network models of where our data is flowing? And that, a network analysis model is always interesting, so do we accidentally start leaking information that way, I think that's something we have to be able to address.

Second thing is, I think we have to translate this back to the vernacular. You say Blockchain, and bitcoin, and you know, cryptocurrency, to people, it's not instantly clear why that's a good thing, or why you're, it's a trust enabler. And a lot of this is about engendering trust with people who are usually very willing to share information, if they believe that they're getting value from it, or they're doing good.

We get people to recycle. And recycling is not obvious, right? You have to separate things, and put them in special cans on certain days of the week, but you do it because it's generally a good thing.

Michelle Dennedy: Right.

Hal Stern: We manage our energy usage under peak demand times, because it's a good thing to do. It's part of the social contract of being a good citizen. If we want to be a good citizen in the world of healthcare, obviously, people have been, for a long time, giving consent to share their data. If we're going to go do that with better, or new mechanisms, how do we translate them so people can understand them, and not giving them a very long disclosure of what that means, and what the rules are, but to try to make it as transparent as possible.

And again, I think there are opportunities here. Interesting challenges to get solved. Certainly, one of the fun technology projects to go work on is to understand how this will create new opportunity, but, you know, it's far from magic, and it's far from well understood, and solved, right now. But it is, much like I think our evaluation of digital rights management technologies in the media space, you know, something will go in and explore, and figure out just what the implications are, for the consumers of, the producers of data, like you and me, and all of the interested and, in some cases required, consumers of that data.

- Michelle Dennedy: Yeah, I think it's an area that attracts signal writers like no other, because I think it's also, if you look at some of the evolutions, and revolutions in medical, and bio-sciences, some of these people have been material sciences people, or people who were in the plumbing industries, and talking about flow. I think that, similarly, technologists have a lot to do with what's gonna happen next in longevity, living well. Wellness, as well as curing sickness.
- But it's also always been one of these areas that I'm very attracted to, because it's like privacy, so many people have come back over the years, our mutual boss, in the past, had said that privacy was dead. I disagree veraciously on that, and I think it's the same thing with healthcare. We have a finite time, maybe, on the planet, but we can live as well as we can, and technology can help lead the way.
- Hal Stern: Absolutely, and I think that, to go back to Scott's comment about "Privacy is dead, get over it," as he said--
- Michelle Dennedy: Thanks [crosstalk], we love you.
- Hal Stern: I also, yeah-
- Michelle Dennedy: We do want Scott on the show.
- Hal Stern: Also completely disagree, because I think, it's not that it's dead, it's different.
- Michelle Dennedy: Yes.
- Hal Stern: It's not, as [crosstalk] used to say, Ben Franklin and Thomas Jefferson sitting on a park bench in Philadelphia, discussing revolution. Where they were fairly certain they weren't going to be overheard, provided there were no other people within earshot of them, and they had a guarantee of physical privacy. As we have become more networked, and more connected via, at first, analog, and then digital channels, we have to be sensitive to that. And I think the challenge is to reduce it to that same vernacular of two people, sitting on a park bench, making sure there's no one around, laying in the bushes.
- Michelle Dennedy: Yes.
- Hal Stern: And how do we do that? And, again, this is the, it's the hope, it's the opportunity, it's the fun technical challenge, is to do that in a way that allows us to create mechanisms that are good for individuals. That, they put you in control of your data, they give you a sense of where your data is going, and why, and they return value to you, in exchange for that.
- Michelle Dennedy: I love that. I think you've already answered the question, but what gives you hope in this space, Hal, I mean, how are you looking to the future, and continuing the drive, and the energy that you've got in this area?

Hal Stern: The fascinating thing about data in healthcare, is that there's an awful lot of data. There are some huge problems to solve, and you sort of look at, what do we do with the data, and how do we interpret it? But the other thing is, I think, increasingly, we're seeing that it's an interesting mix of automation, and visualization, and expertise. And that there are, eventually, people in the middle of that. Whether it's individual participants in the healthcare system, or healthcare providers, and practitioners, or research experts. And that it tends not to be the big machine, into which we feed all the data.

But, again, the thing that gives me hope, and also makes my job fun, is that it really is about driving understanding, and we do that because we have people who are experts in the area who can synthesize in new ways.

Michelle Dennedy: Yeah, I love that, and it evokes the Mr. Rogers quote, "Look for the helpers." The helpers always come in. Doctors walk into situations that seem hopeless, and they help. Technologists like Hal Stern comes in and looks at something as complex, and important as healthcare, and you dive right in, and you've stayed in there, in the field, and that gives me a ton of hope.

So, thank you, Hal, this is amazing, as always. I'd love to have you back on, when you're next in the Bay Area.

Hal Stern: Thank you, Michelle, thanks for the opportunity to be a Sigma Rider. Being, you know, five sigmas to the right of anything considered normal, it's certainly a great honor.

Michelle Dennedy: Excellent, we're abnormal together.

Hal Stern: Absolutely.

Michelle Dennedy: Well, thank you very much, and thanks for listening to this data mashups and missteps, what Norwegian death metal, food pictures, and my morning commute can, and can't, tell you about me. My favorite title to date.

You've been listening to Privacy Sigma Riders, brought to you by the Cisco Security and Trust Organization. Special thanks to Kory Westerhold for our original theme music. Our producers are Susan Borton and David Ball. You can find all our episodes on trust.cisco.com, or subscribe wherever you listen to podcasts. Then please take a moment to review and rate us on iTunes. To stay ahead of the curve between episodes, consider following us on Facebook, LinkedIn, and Twitter. You can find me, Michelle Dennedy, on Twitter @mdennedy. Until next time.