

Intro: When you take a magnifying glass and look at a picture, you see that it's made up of thousands of dots and when you pull back, those dots become the details that create the picture. There's real power in the details, because when you have more detail, the bigger picture becomes sharper and wider, and a story emerges.

Hi, this is Amie Moreno and you're listening to "Seeing the Big Picture: Conversations on how Data and Artificial Intelligence can add the Details that Fuel Deeper Insights in the Life Sciences Industry."

Amie Moreno: Hi everyone, this is Amie Moreno coming to you to explore electronic health records and natural language processing. Today, I'm with Marguerite Dresser, a Director on the Data, Advanced Analytics and Tools team here at Optum. Marguerite is going to share her experiences working with the EHR and NLP from a life sciences and academic perspective. So Marguerite, thank you for joining us today.

Marguerite Dresser: You're very welcome, thank you.

Amie: Would you mind giving us some background, both in what your current role is and also some of your background?

Marguerite: Sure, hi everyone, I'm a peer of Amie's in Advanced Analytic and Data Tools group. We service a lot of clients across the life sciences industry and my background is in epidemiology originally and I moved to working in quality measurement in the provider and payer space for quite a long time and then also worked in the life sciences space looking at sizing, really, disease burden both in the U.S. and internationally.

I am here in this role, focused on Data and Advanced Analytic Tools after a few years ago, taking on a sales role, so I really better understood our clients' perspectives, needs, and senses of urgency. So, I'm very excited to be learning more about claims, EHR, and artificial intelligence and machine learning techniques, and I'm looking forward to our discussion today.

Amie: Great, thank you for that. Your background is quite extensive, so you understand this from a lot of different angles. Can you tell me a little bit about the data that you're currently using and the data you've potentially used in the past, how that's changed, been enhanced, etc.?

Marguerite: Sure. Years ago, claims was really all the "big data" that was out there, in addition to a lot of large data assets in the U.S. that were available federally that you could sort of download and explore. But in terms of something that was a very contained asset that was used in different

industries to address questions, claims was all there was. And then, I'd say, maybe 10, 15 years ago, EMR data started to be available, but it was really, represented individual EMRs rather than the continuum of care that is represented now in the best EHR data sources and certainly, there were a lot of provider notes that were taken down, but that was text and there was no way to really mine it at the time and learn from it. And now, it's very exciting to see that some of the data sets that are available allow you to really have a lens into the provider/patient dialogue and understand things that, before, we really just weren't able to glean at all.

Amie: And so, I know you've worked on an EMR system in the past, I believe, in more of a technological role. Given what you saw back then, 10 years ago, versus what's available now, how would you describe those enhancements or how much more, not only the breadth and the depth, but what specifically has been enhanced or how can we better use that data?

Marguerite: Sure. I think the thing that strikes me the most is that we, coming from a claims world, at the time, I worked on a pilot in the state of Massachusetts, actually, focused on trying to see if EMR data could be used to create quality measures that had been created for years via claims data. And what we did is that we tried to create technical specifications that really paralleled what we had done in claims and now what we've moved to is really starting to understand how EHR data has unique properties where we're starting to develop methods that are appropriate to EHR data, not just trying to duplicate what was done for claims, which was really a different animal, it's like apples and oranges, so the power of the learnings that we can have going forward, I think, are going to be much more substantial than we've done to present.

Amie: Sure, and would you agree that the ability to integrate not just EHR and claims data, but other data sources out there -- we've been hearing a lot about, wearable devices, for example, that give you a lens into not just the clinical perspective, but also motivation that comes from your surroundings, that come from your social networks, etc. -- do you agree that that integration is crucial in order to really understand holistically patients?

Marguerite: Absolutely. I think what we all want when we have a research question to address or a commercial question to address is to really make sure we have an accurate understanding of the patients that we're focused on. And to the extent that all data sources are really just one picture and EHR data is, I think you mentioned, is a much more complete picture than a lot of the data sources that have been available to present, but any additional pictures for where we know, either the data hasn't been mined yet from provider notes or that it's a unique aspect that simply isn't captured within the clinical experience, the more that we're able to knit things together, the

better understanding we can really have of patients and meet them where they are to develop therapies, to produce better health.

Amie: Sure. We use EMR and EHR interchangeably a lot. Can you explain what the difference is from your perspective of a single EMR versus an electronic health record?

Marguerite: Sure. EMR, historically, though, you're right, we use them interchangeably. Historically, an EMR was a system at a particular site that was built as an electronic medical record and initially it wasn't necessarily a complete medical record, but there were components of the medical record that were "electronicized," I suppose -- I'm not sure that's a word -- but, and then eventually, a full medical record was created at these individual provider sites and used as the medical record. But let's say you had an ambulatory care site where an EMR was built and that was eventually used as a medical record and the paper records were either used as a backup or no longer used at all. That would still only provide you with information about ambulatory care and you might not know anything about any possible hospitalizations for the patients of interest. So, EHRs are really the joining of electronic medical record systems across the continuum of care, anywhere where patients might receive care that would be relevant.

Amie: That you're able to see both ambulatory, but also that detail that occurs within the hospital setting.

Marguerite: Right and even, let's say, if somebody's in a hospital and they are discharged for short-term rehab to a nursing home. That kind of detail you could see in an EHR because they're all connected, but you wouldn't be able to see that if you were just looking in a site-specific EMR.

Amie: Right. So, I understand you're doing some interesting work within the oncology space using artificial intelligence, which we'll refer to a lot as NLP, natural language processing. Can you tell us a little bit about that experience and exactly what you're able to do and how that's going to impact the research?

Marguerite: Sure. We had a client who is very interested in a biomarker that was recently approved by the FDA. It's the first biomarker to be approved rather than a location in the body where a tumor originated. And so, information about that biomarker isn't currently available and codified, big data assets. And so, we are working with them to look and see when those biomarkers are -- because there are actually two of them -- are mentioned in provider notes and the context around when they're mentioned. The objective is to understand when these biomarkers are tested, why providers decide that certain patients will be tested, and

maybe other patients aren't tested and really build a data asset that provides information around the target population, which is patients who have been tested with a biomarker, then, of course, to understand how often are the patient's positive for this biomarker, what kinds of provider specialties typically order this biomarker, and in which parts of the country, etc. Any kinds of more granular information that will allow our client to understand how they can best serve both providers and patients in terms of recommending their therapies that are specifically work for patients with these biomarkers.

Amie: So, they can use that to educate physicians to not only find the right patient, but at the right point in their disease progression?

Marguerite: Absolutely. And since this is all so new, as I said, with the FDA approving these biomarkers and that being a new approval, this is going to iterate. We're looking to see what has happened to present, but we know that even in terms of the guidelines around these biomarkers, there are new things that are going to be published probably in the next six months that we can keep iterating and looking and refreshes of the data to see how much the population is growing, whether the distribution of which patients are actually having the test and which providers are ordering the test, whether it changes.

Amie: Great. So, in terms of this process, what have you found to be a potential challenge or what are some of the limitations versus the value -- given your experience in this exercise, but just generally that you find with natural language processing mined with data?

Marguerite: Sure. I think for many of us who are not necessarily data scientists as much as I've been an analyst for years, I wouldn't call myself necessarily a data scientist, the difference between a sort of text-based searching for a term, which I think a lot of times when people think of NLP or artificial intelligence, we come to it much in the way that we have created other forms of information in the past and really, NLP is broader than that; it represents not only the initial sort of looking to see if what you're interested in exists in the data, but then the process -- if it does exist with sufficient frequency -- to then see what context. Let's say you have a term that would be of interest, even just a tumor type term. By finding it in the notes, that doesn't mean that it necessarily occurred, you have to understand what the attributes of the term are, what the sentiment is. Maybe it's a rule-out diagnosis, maybe a provider was mentioning tumor type because they wanted to say, no, I don't think this patient has this. And so, understanding the whole context around it, is really what NLP develops. Having clients understand that it's not just a Google search of the provider notes is something that sometimes is a little hard to digest initially. So, there's a lot of learning together and partnering with clients to

understand sort of where we are in the exploratory phase and then we can move to an annotation phase where we invite in our clinical experts to look at the notes and see what needs to be tagged in terms of the nuances that are most important for this concept and then we move in to the artificial intelligence, machine learning, NLP days where the machine is able to scale things in a way that humans simply can't, and able to generate from the input and the clinical expertise that we put in on the left end, iterate and generate output on the right side that is at scale.

Amie: Yeah, and I think that could be potentially a misconception, meaning that there's that first phase, as you mentioned, involves humans, it involves experts both at the clinical perspective from data scientists, from natural language processing because, ultimately, it's a technology that's not a human, so there's a lot of that upfront work that happens to make sure the results are as accurate as possible.

Marguerite: Right. I mean, I'm sure you could actually do NLP, I suppose, without clinical expertise as part of your process. I don't think that would be as accurate or as precise as what we need in this industry and certainly, I'm lucky enough that the work that I'm doing has not only incredible data scientists, but clinical experts who have both provided care and lab experts that really understand how labs are recorded, what they mean down to the detail that is honestly somewhat beyond me, but really necessary to the process.

Amie: Great. So, is there anything in particular that's coming up that you're excited about or any ways that you think about to potentially use the data that you haven't had the opportunity to and could potentially in the future?

Marguerite: Not thinking about the project I'm working on specifically, but as a former epidemiologist, I think one of the ways that EHR data and NLP could be used is to really understand some of those confounding clinical problems – I'm especially interested in pediatrics -- a family who has a child who's sick, but there's not really any diagnosis and we don't really know -- or things that are difficult, where we may have diagnoses like chronic fatigue, but there's just not precision around what is this? Is this all the same thing? Are these different things? And that's where I think using NLP methods and really getting into those billions of notes that are available to understand, you know, how do we diagnose these patients and how do we then, by that diagnosis, hopefully, we'll be able to then create therapies that address some of these chronic conditions that are really confusing today.

Amie: So, Marguerite, we've done some research and we find there's a lot of questions from clients around natural language processing. Does it work?

Is it accurate, etc.? How would you address those questions, from your experience?

Marguerite: Sure. First of all, I think it would be really important to start with what a client means in terms of “does it work?” If they're looking to use NLP to develop information that they can point to as data that is the same, you know, an apple to an apple in terms of like the claims data, for example, they've been used to using, then I would need to help educate them that that's not, in fact, what NLP can produce. That said, given the depth of potential information that really could be unleashed in provider notes and depending on the question that's being posed of those notes, I absolutely think it can work to inform business questions and clinical questions that, heretofore, we haven't had information to address. So an example I became aware of recently is that a project that I was not involved in, but some colleagues of mine were involved in a project where they were asked, how good was the NLP that you produced? So, they did a validation study to look and compare the NLP pipeline process to a human, clinical annotation process. What they found, happily, is that the results, the true positives, the rate was really high, like greater than 90% across the different domains that they were looking at and they also found good recall in terms of the true positives that were identified, those were also able to be gleaned from the data, so able to be extracted from the provider notes. So that kind of validation exercise, I think, depending again upon the kinds of questions that are being addressed of the data -- sometimes you don't need that level of validation to occur, because you're really just exploring to better understand a potential patient population. But, certainly, if you're going to use something to develop a therapy, you want to make sure that the information is very precise. So again, I think the question of whether something works, it depends what you mean by works and what the bar is for evidence.

Amie: Well, thank you, it's been a pleasure chatting with you today. I appreciate your feedback and your expertise.

Marguerite: Thank you, Amie. I look forward to learning more from others that you've interviewed.