CERAMIC TECH CHAT

Episode 14

Title – "Materials Genome Initiative 10 years later: James Warren (E14)"

INTRO

De Guire: "I'm Eileen De Guire, and this is Ceramic Tech Chat.

This month is a major milestone for the Materials Genome Initiative, an initiative launched in June 2011 by the United States federal government that aimed to enable a paradigm shift in materials development by leveraging data science, informatics, and modeling tools in new ways. Now, those involved with the initiative are looking at how to leverage the significant advances of the first decade to create a new strategic plan for the next 10 years."

Warren: "So, we're trying to be very careful to figure out what is the government's role. Certainly the government's role is not to say that this kind of research is important without plumbing the consensus for what the community thinks is important and trying to understand what industry needs. So, in that sense, we're there for them."

De Guire: "That's James Warren, director of the Materials Genome Program at the National Institute of Standards and Technology. Jim has been involved with the MGI since the very beginning, and he continues to play a central role in guiding the initiative into its second decade.

So, what was the genesis for the MGI, and how has it developed over its first 10 years? And how does Jim expect the MGI to evolve through the 2020s?"

(music)

SECTION 1

Warren: "When I started at NIST, I didn't know almost any materials science. My training was in statistical physics, [University of California] Santa Barbara, and I did my Ph.D. on solidification. It was sort of fundamental problem in dendritic spacing and trying to predict the, you know, essentially the space between arms of the snowflake-like structures that grow when you crystallize a metal, for example. And that was fun and interesting for me, pattern formation from the physics community point of view, and it's very subtle. And then I got fortunate because of that work to get a postdoc within the metallurgy division at NIST at the time, in 1992, which is almost 30 years ago, which is mind blowing to me.

I got there at just the right time. Some amazing people, some brilliant minds where all there. I had some amazing mentors. A real, just, the place was bubbling. One of the real reasons was the presence of John Kahn, who's one of, you know, the great materials theorists of the 20th century. And so John was there as a fellow when I showed up, and of course there were a number of other colleagues as well, I don't want to sell them short. John was a busy and famous guy, but he created a space there for fundamental work—still high engineering impact, but still truly fundamental and long range—that didn't necessarily exist everywhere in government, let alone at NIST. Really ahead of its time. And because of that, in that energy, essentially, I was able then to stay.

And we created something called the Center for Theoretical and Computational Materials Science with several other postdocs at the time, all of whom have gone on to great things. I'm the only one left at NIST of the four of us that founded it. And that really got me interested in the issues around software and sharing of software. We really founded that Center around the notion that sharing one's software that one used to simulate a material was much more valuable than writing a paper. So when the MGI comes along, I'm like, 'Oh yeah, we've got to do this, because this is the right thing to do.'"

De Guire: "So what drove the idea behind the MGI, and how did the materials community react to the white paper?"

Warren: "Right. So that's a lot to answer, so let me see if I can sort of lay it out. So, the MGI, when it was rolled out, was a collection of ideas that weren't terribly new. There had been a large number of reports over the last few decades that preceded the rollout. Looking at how one could accelerate the design, discovery, deployment of new materials faster by tightly integrating modeling with experiment and better management data. What really started to change was that these ideas, while they seem sort of obvious at some level to a lot of the practitioners, really were starting to bear enormous fruits. If you look into the early 2000s to mid 2000s, then you'll start to see reports coming out calling for integrated computational materials engineering, you'll start to see a lot of the databasing efforts that were occurring in computational regime, mostly around density functional theory, that actually were yielding true payoffs. And so, the idea for the initiative had been sort of bubbling in, as it were, in the firmament of materials science and whatnot and all of these related disciplines like chemistry, for a long time. And so, when the Obama Administration approached the National Science and Technology Council saying, 'Hey, we think something like a Materials Genome Initiative would be a good idea,' there were a lot of people in government were like, 'Yeah, we can make that work.'

And I'm laughing now because, of course, the one thing that we didn't love was the name. There was a lot of discussion in the rollout of the white paper, when we were sort of writing it. Nobody really loved the name, but we realized that it was… the people that didn't like it, of course, were the materials researchers because it's a metaphor. Everybody else kind of gets it, right? They kind of see the analogy with the Human Genome Initiative, they knew that it was a metaphor, and they were comfortable with that. But scientists don't like metaphors because they're not precise. So in terms of perception, if you asked me what the worst part of the reception was, it was the name.

I think there was in general a great deal of delight that there was a major initiative in materials coming out of the government. The only other one really at that point was the

Nanotechnology Initiative, which was very substantial. And so the notion that there would be something that went beyond nano and really also had more of an emphasis on computation, I think was very exciting."

(music)

SECTION 2

De Guire: "So when we think about the Materials Genome Initiative, materials, that word alone, has a huge span. So how did you guys, who were crafting this white paper, how did you put some borders on what the materials scope was going to be for the Materials Genome Initiative?"

Warren: "We didn't want to put too much of a scope on this. In fact, we really understood that materials really did span anything that's made out of atoms, basically. There is a challenge of knowing sort of when to stop. But we did really view the materials development continuum was going from discovery of new materials all the way out to deployment and manufactured products, even, in principle, to service. Because, you know, we're working a lot. One of the big partners is DoD, for example, where they really, really need to have working things that don't break, or if they do, they want to understand how. And also the NSF, which funds a lot of fundamental research, more at the discovery end and the basic research end. So we really did want that impact and the influence of, say, integrated computational materials engineering as a design philosophy for materials that get incorporated into real products was influential. So we wanted to see this sort of foundational engineering problems kind of approach as well.

So we were trying to work across government and across all these technical research, you know, TRLs [technology readiness levels], to have something that everybody could be in the MGI umbrella. So we didn't draw a lot of boundaries. Certainly no materials boundaries and certainly no TRL levels. Once you get into actual manufacturing design and things where the materials playing, at best, an ancillary role, okay, so maybe we don't want to go that far. But if you asked me today, we are, my main focuses would be more outreach to manufacturing, more getting these technologies translated to the manufacturing level, and so I think we're really ready for that at this point."

De Guire: "That's tremendous, within the space of just 10 years to be already thinking about that translation into manufacturing. One of the goals of MGI, right from the start, was to build an infrastructure that would support these ideas of integrated computational materials engineering and you mentioned DFT [density functional theory]. So, what progress has been made on building some of these computational tools, the experimental tools, the collaborative networks, the digital databases, and access that was part of the dream?"

Warren: "So, I hadn't mentioned it yet, but of course in the end the MGI is a bit sneaky compared to a lot of these other initiatives because the focus is really on the evolution of this infrastructure. So it's in that sense a meta initiative, which is we are trying to build the things that allow us to make the materials. And it really is, as you say, it's very generic.

And because of that, it was kind of hard to explain, right? It's a little bit abstract because it's very generic. A lot of these tools are about managing data or how do you do a computation. And it's not like we want to make the next great battery. We want to make the technologies that allow somebody to make the next great battery. And so like that, the documents can be a little opaque and things like that.

So, in terms of specific examples where there are infrastructures, they're all over the place. Of course, probably one of the marquee examples would be the DOE's materials project, which originally was called the Materials Genome Project and the MGI nakedly stole for the name of the initiative. And we really appreciate that, and then they renamed themselves. And so there are, as I've said, many databasing efforts beyond just the materials project, but that's an exemplar. And there are just a lot more resources, like the Materials Data Facility and Materials Commons, which NIST and the DOE fund, respectively, which are more sort of generic data hosting efforts that have made a great deal of progress. There are a lot of efforts at NIST and other places, and just trying to think about better ways of curating and managing data so that other people can find that data, discover that data, and just reuse that data in ways that are more efficient and robust. How do you merge data sets, how do you gain extra value from that information? So, a tremendous amount of effort.

And then you mentioned software tools, computational tools. The DOE funds a lot of these sort of sustainable software efforts, of which the MGI is happy to build upon for computational research and predictive materials research. And then there's also this sort of whole community building activity. And that's, in my mind, almost a whole separate conversation about how we engage broadly to build out this infrastructure."

De Guire: "I think it would just be nice to talk anecdotally about some of the community efforts that are sprouting up, driven by the researchers."

Jim Warren: "Right. So, community efforts. I mean, there's this sort of broad problem, particularly around data, which is certainly where NIST has focused most of its energy, and it's been an interesting puzzle. So when the MGI rolled out, NIST as an organization looked at what the initiative was trying to do, and each agency sort of stares at it and goes, 'Okay, well what's our role?' So how do we make the work of the agencies become more than the sum of the parts? And NIST was like, 'Well, you know, we're the deliverers of awesome data, that's sort of our role, so let's figure out how to frame our support around data,' and we ended up with three sort of main areas. They were ensuring the quality of the data, so thinking a lot about getting the best information possible out there, and that also means thinking a lot about the models that are used to interpret experiment and measurement. And then we were also thinking a lot about the data exchange problem, as we framed it, which was how do you move the information around, how do people discover it, things like this. And then, in principle, once you have those two, you'd like to then hopefully enable data-driven science.

Now we were talking about all this in 2013, so it was a little before you started hearing a lot about artificial intelligence and machine learning as applied to materials. There were only a few sort of brave researchers at that time. But that's, of course, completely exploded at this point, so we're giving ourselves a little pat on the back for sort of anticipating that, more than I certainly did. I said, 'Well, this is the future, but who knows when.' And on the data exchange part, we thought that was sort of technically boring and, in some sense, just, you know, it's solvable. And while that's sort of true, that turns out a) it's not boring, it's very interesting, and b), it's hard, it's just very hard. And a big piece of that problem has to do with incentives. And that has to do just mostly with the way that academia rewards publication. And it also, industry, you know, how they make money, right? It's all of a piece, fame and fortune. So, academics get rewarded by and large only through the publication in high-impact journals. So, the publication of data is not rewarded, impacts associated with that publication is not rewarded.

So, there's this chicken and egg problem that sits there. And everybody just talks about it, there's no incentive to do this, I don't want to do this, why should I do this. And the answer, everybody knows that it would be valuable if we did, and the notion that there's a difference between a paper and the data that supports that paper is goofy. It's a completely arbitrary line, it's an accident of history that we have that line. So, what we really need to do is, in my mind, it would be nice if we could just say, 'Hey, everyone, you have to publish your data now.' But if you make it just a mandate, basically you'll get pretty bad compliance. You'll get people who barely, they'll do the minimum.

So what you want instead is what I like to think of as this asymmetric model, which is that if… I like to use social networks as an example. If I said to you, 'Hey, I want to create a system where I can target ads to you based on your personal preferences and interests,' you'd say, 'No thank you.' But if I said to you, 'Hey, I'll create this thing that allows you to say hi to your friends on their birthdays and arrange meetings, and, you know, various things like that, and see funny cat pictures,' you'd be like, 'Oh, that sounds pretty good, I want one of those.' And now, of course, it turns out that's because, you know, the social medias' goals and yours aren't the same. But it solves the problem for both of you, as it were, and so you're willing to make that exchange in certain cases. Similarly, I think that when we create platforms that allow scientists to achieve their goals and in the case that would, if it's a research goal, they want to write a new paper, they want to solve a problem. And the data sharing and data management piece is going to be a side effect of that. That's when you start to see this work because then the incentives are in the right place. For a company, it will be the same thing. They want platforms that allow them effectively to do whatever it is they're trying to do so they can make money, so that they can make the discoveries or resolve the engineering problems that they need to solve.

So the infrastructures that the MGI is really trying to build, in particular on the data side, have to acknowledge that and start there. And that's where the community becomes so essential. Because the right way, in my opinion, to ask the question is if the infrastructures of the MGI are not, if you feel like you have no incentive to use them, what's missing? What would it take so that you have the incentive to use them? And I'm assuming that we'll get back answers, and then we need to think about what, how do we get to there, right? So we're going to end up with a roadmap that these communities can engage at the right level. So that's the way I tend to think about the problem. And this has been a long thought process for a large number of us on this."

(music)

BREAK

De Guire: "For more than 60 years, ACerS and the National Institute of Standards and Technology have collaborated to offer the Phase Equilibrium Diagram Database, containing more than 30,000 relevant, critically evaluated phase diagrams with commentaries for inorganic compounds. Learn more about the ACerS-NIST Phase Equilibrium Diagrams at ceramics.org/buyphase."

SECTION 3

De Guire: "What does the future of the MGI look like as it turns that corner of 10 years and looks to the future?"

Warren: "So, to me what it looks like is at least two-fold, two ideas that are sort of in the front of my mind. One of them is something I mentioned at the beginning, which is this deeper integration with manufacturing. We need a stronger poll from the manufacturers about what they want. And we don't want to make them work too hard for it because that's not something they want to do we, but we need to figure it out, we need to figure out the engagement models and the discussions that we need to have to get them these tools. Because it's not just enough to say, 'Hey, we've got these great tools,' we've got to figure out what the barriers are again to adoption, what are their incentive problems beyond the sort of stuff we were talking about with academia; that's not their set of issues. So it's complicated, and it's very company dependent, and it's just a lot of the outreach, and sector dependent. So we just sort of need to sort of figure that out. So that's going to be a big focus I think of the MGI moving forward, is getting us all the way out on that TRL scale.

Beyond that, if you asked me, 'So what is the MGI going to start to look like?' I guess, I would have to say that I want to see a lot more focus on the integration piece. It's always been at the heart, but there's a lot of gaps. The sort of distance between the gaps, I think, is now starting to become small enough that we can really start to knit this thing together. And as we start to see more interoperation of various resources and scales, I think this is going to start to accelerate the MGI. If we were to look at sort of the Human Genome Initiative, there were some very nonlinear moments in how the cost of sequencing changed. You know, it started at nearly a billion dollars for the first one, now you do your cat for 100 bucks or something like that. So, you know, a lot of progress there. And I would imagine that we're going to see similar kinds of changes, where suddenly something that's going to drive the cost of certain pieces way down and then you start to attack some other element in the structure. But also, as people start to see the value proposition in these kinds of approaches, it becomes obvious to people that maybe it wasn't obvious to before, and we start to see real disruptive, rapid change in the way that things get done.

There is no question in my mind that, with the exception of probably biology and healthcare, materials is one of the most likely lucrative aspects of the application of AI that you can imagine because you're going to make stuff that people want. It's really that simple. And in the end, people want things a lot, right? So it makes up your life. So you can get food, and then you've got stuff, right?"

De Guire: "Well, and the application spectrum is infinite."

Warren: "Yeah, it's infinite. So, the potential here is so enormous, the economic potential is so enormous, that I don't think we, most companies have been able to really grapple with it yet, although you're starting to see it. I mean, the easiest one to make the argument for would be pharmaceuticals, which is sort of at this great boundary layer between chemistry, materials, and healthcare. But drug design is a classic materials problem. And those companies that are involved in that used to be giant pharmaceuticals. And now it's gone through many cycles, and a lot of them thought the forward modeling community, you know, your molecular dynamics and things like that, was going to solve their problems. Eh, it wasn't quite good enough at the time, and this was in the late 90s, 2000s, I'm thinking of.

I think with the AI again, we're seeing a whole other route. They're all of these startups with huge amounts of venture capital. And for small molecule drugs, my guess is we're just going to build robots and make them all. So we're going to see huge change and some of it will be driven, as I said, by these sort of areas where the profits are relatively clear. Whereas, you know, when I say, well, better batteries or better this or better that, again it can often be harder to see the ultimate application, or you worry about the marketplace, or there are all of these confounding interests. That's still going to always be the case, you never know what's going to catch on. But the capacity to do this more cheaply and easily, which is what the MGI is about, is got to be at the center."

(music)

CONCLUSION

De Guire: "As the Materials Genome Initiative stands on the threshold of a new decade, the opportunities to curate and manage materials data will continue to grow—and the opportunities to discover new materials, processes, and applications will grow along with it."

I'm Eileen De Guire, and this is Ceramic Tech Chat."

(music)

“Visit our website at ceramics.org for this episode’s show notes and to learn more about the Materials Genome Initiative in the June/July 2021 *ACerS Bulletin*. Ceramic Tech Chat is produced by Lisa McDonald and copyrighted by The American Ceramic Society.

Until next time, I’m Eileen De Guire, and thank you for joining us.”