



Season 3, Episode 1

Joe Insley: Big Data to Beautiful Images

SPEAKERS

Sarah Webb, Joe Insley

Sarah Webb 00:00

Welcome to Season Three of Science in Parallel, a podcast about people and projects in computational science. I'm your host, Sarah Webb. In this season, we'll be exploring frontiers in computational science: novel processors, research questions that require next generation computing and the practical challenges that come along with so much data.

Sarah Webb 00:26

We'll start with a conversation about visualization with Joe Insley, who is both the team lead for visualization and data analysis at Argonne Leadership Computing Facility and an associate professor in the School of Art and Design at Northern Illinois University.

Sarah Webb 00:41

You might have seen Joe's work without realizing it. His team's animations and images regularly show up in the Scientific Visualization and Data Analytics Showcase at the annual Supercomputing meetings and have graced science journal covers, including Nature's. Joe's work relies on his training and experience in both digital art and computer science. We talked about his unusual career path, how he approaches these projects, and the future of visualization with exascale systems and beyond. We have a lot of extras with this episode. Please check out our show notes at scienceinparallel.org for images and a narrated animation.

Sarah Webb 01:32

Joe, welcome to the podcast, it's great to talk with you.

Joe Insley 01:36

Thanks for having me.

Sarah Webb 01:37

What I think is so interesting about your work is that it's both so artistic and so technical. And so I want to take you back and tell me-- what came first: art or computing? And how did these interests come together?

Joe Insley 01:52

I would say that I've kind of come full circle a little bit over my career and education. Right. So I actually started out in undergrad as a computer science major, but I always had an interest in art. And so I learned early on in my undergrad degree that you only needed a handful of courses to get a minor in art. So I thought, sure, let's do that. And one of the first courses, I took just a design course, the TA asked me, he's like, What are you doing in computer science? Computer graphics? That's where you should, and I'm like yeah, how do I do that? And it turned out, this was at Northern Illinois University where the whole circle part will come in. So I switched to art pretty quickly, early on in my freshman year, even though I continued to live on the computer science floor, so figure that out.

Joe Insley 02:36

So I got my undergraduate degree, bachelor of fine arts, with electronic media sort of emphasis, doing like computer animation. And you know, and this was back in the early 90s. So there wasn't a ton of that going on. And then I went to grad school at University of Illinois at Chicago, UIC, the Electronic Visualization Lab there, which the two co-directors, one was a computer science professor and the other was an artist in school of art. And so, I originally came in on the art side, and started working there doing animations and whatnot and got turned on to doing scientific databases there as an art student eventually finished my MFA degree doing virtual reality environments. And some of that, along the way, picked up scientific databases as a research assistant.

Joe Insley 03:23

Around the time that I graduated, my first child was born. And so I intended to stick around and do the computer science degree full time, but needed a way to support my family. So I was lucky enough to get an opportunity to come to Argonne National Lab where I'm at now doing scientific data viz and continued to go to school eventually got my master's in computer science, that UIC while I was working at Argonne. Many years later, I got an opportunity to go back to NIU in the School of Art, essentially the same program that I graduated in, and I teach data visualization and intro to programming in the School of Art. And so I get to draw on my background in both computer science and art in both my day jobs.

Sarah Webb 04:03

I can totally imagine with this kind of, you know, computer graphic- type background that the first thing you tend to think of is gaming. Right? You know, how did you end up going into this path toward scientific visualization?

Joe Insley 04:16

There's a lot of the same skills in terms of drawing on the art background in terms of color and scale and, and design and that sort of stuff. And I think it was just an exposure to it. I met the right people at the right time that were looking for people that were looking at visualizing scientific data. One of the earliest projects I worked on my advisor on the art side at UIC, Dan Sandin, was really interested in mathematical Julia sets, math equations, and visualizing them and so I worked with him and others on projects associated with that. And then because we were doing a lot of so if you're familiar with the cave virtual reality environment that came out of UIC around that time, I was a few years after the initial CAVE, but did you know I was involved in a lot of the early work there.

Sarah Webb 05:02

Sandin was one of the original inventors of the CAVE in 1992. And these spaces projected high resolution images on walls and the floor, adjusting to a user's position and rendering 3D views. They were groundbreaking at that time and are still used in science visualization.

Joe Insley 05:21

And so there was a keen interest in looking at scientific data in VR [virtual reality] and immersive environments. So that's really kind of where I got my start. When I first came to Argonne, I was looking at putting scientific data in VR. So I did my master of fine arts degree in building environments for the CAVE, and then got to do the cyber stuff along with it, which became my day job.

Sarah Webb 05:43

Obviously, visualization and the tools of change, how has your job changed over the years?

Joe Insley 05:49

A lot of it is the tools have evolved, one of the big challenges in visualization is often especially looking at data at the largest scales that comes off large supercomputers. And so there's always been this sort of moving target of the data's getting bigger, what our needs in terms of memory on the computer, ability to render. In particular, if you're trying to do something exploratory and you want to render in real time versus you're rendering something for an animation where we can submit it to the queue and if it takes an hour to render a frame, that's okay. And so there's those tradeoffs. So a lot of it is sort of the scale of the data continues to grow. And one of the major challenges we're facing now is, it's not necessarily an entirely new problem. But as the rate of compute power of these biggest computers continues to grow and expand, we're able to compute much faster than the rate at which data can be saved to disk, right. So the I/O rates are not increasing at the same pace as the compute rates. So you know, this gap between how much we can compute and how much we can save continues to widen. And so somewhat out of necessity, the community has been looking at methods for doing in situ visualization and analysis-- so looking at the data, while it's still in memory. But with that big gap in how fast you compute, how fast you can save, there's going to be lost science. We can't save everything.

Sarah Webb 07:16

Well talk a little bit about that. And generally, at the stage that you get involved with some of these projects, a lot of times is this huge science, there's a lot going on the simulations are big and trying to capture all sorts of information. Are you involved from the very beginning of this process? Do you sometimes come in in the middle? How does this play out? Or does it depend on the project?

Joe Insley 07:42

It's project-dependent for sure. And I should say, it's always an iterative process, right. My team and I and others work very collaboratively with the science team. They know what they're looking for. We don't. We're scientists and visualization experts but we're not science domain experts. But that's actually one of the things I love about my job is that I get to work with all these people that are experts in their field. And it's always a learning experience. But to answer your question of when do we get involved? It totally depends. Sometimes they've already run their simulation, and they come to us and say, Can you help us figure out what's in there? And so it's completely post-processed. Other times,

we'll be involved while they're running the simulation, start planning out what they want to do in terms of visualization.

Joe Insley 08:29

One of the other challenges along the way is also that we're in order to make the best use of these large compute resources, data needs to be in a particular format, in terms of being able to leverage the specific hardware. And so the format that it's generally needs to be in, in order to make best use of the compute resource is often nothing like the format it needs to be in to visualize it. And so there's data wrangling, as we refer to it, in terms of converting things from whatever the native compute format is into something that we can visualize and exploring it. But then at the other end of the spectrum, we are starting to work more closely with teams like I said this because this gap is getting wider. And it's starting to become a necessity to start looking at data while the simulation is still running. We do work with teams to sort of integrate some of the in situ frameworks, essentially instrument the simulation code with these other frameworks that allow us to then sort of offload some of the viz and analysis either to a separate resource. In which case that's more of what we would refer to as in transit, or, or in some cases truly in situ, where essentially the simulation pauses for a period of time while we take over and do some analysis and visualization using the same resource and then we hand control back to the simulation.

Sarah Webb 09:57

You were talking about the growing importance of in situ visualization, and particularly with Aurora coming online, to what extent is that going to become the default rather than something that's kind of been--hey, do we need this?

Joe Insley 10:10

That's a great question. I don't know that I have the full answer to that. But I think increasingly, that is going to be the norm. Now, in fact, we're planning on that in some cases. So as a part of Aurora, there was a call for proposals for the early science program for applications to get early access to software and hardware that will eventually be Aurora in order to start preparing. And we work with one of those projects, specifically to instrument with in situ viz capabilities. And so it's starting to be more accepted, I think, and recognized that science is going to be lost if we don't make a change. Certainly, post hoc visualization will always be a thing-- at least I think it will be.

Joe Insley 10:54

But there's a growing need and interest in doing in situ. And of course, that comes with all sorts of additional challenges. Quite often during the simulation process, you don't necessarily know where the interesting thing is, that's what you're looking for. You want to explore and find those things. And so essentially, knowing where to point your virtual camera to visualize data, while it's being simulated, is a challenge, and so there's just other areas to do things like feature tracking, and those sorts of things to figure out where to point the camera. We know we don't know where to look, so let's look everywhere. Instead of capturing maybe a single image, maybe you capture images from a 360 view of a thing and capture multiple views that you can then go back and sort of play back from different angles after the fact.

Sarah Webb 11:44

I want to ask you a bit about some projects that are good examples of the work that you do.

Joe Insley 11:53

One of the things we've been pushing on lately is there's a lot of software and hardware out there that's enabling us to increase our rendering capabilities in terms of, for example, doing ray tracing, which makes it look more realistic. And so like I said, in both the hardware and software space, there's been lots of advances in that. And so some of the tools that we use, for example, ParaView, a widely used tool in the sci-viz community has backend renderers that can leverage some of this hardware software that been enabled in some cases. ParaView is an open-source tool. And so some of these other commercial renderers and such are able to take advantage of some of the advances a little bit quicker, right. So there's integration with those tools. But in some cases, it maybe lagged behind a little bit in terms of exposing all the controls that may have, for example.

Joe Insley 12:46

And so one of the things that my team has been doing is looking at how can we combine things like ParaView, which enable us to take the data and its scientific format, load it, convert it into something that's renderable, essentially, turn it into geometry of some flavor, and then export that geometry. And you leverage some of these other tools for actually doing the rendering. One of the other advantages or things that has been appealing to us for building that particular workflow is that we're able to use these other tools to build models to give context. So for example, one of the projects that we worked on visualizing combustion engine. So there, in particular, the science case was looking at flow in and out of, and the circulation of, essentially air inside the cylinder of an engine. In particular, they're interested in what's happening around where the spark plug is, in order to make more efficient engine design in order to convey that to a more lay audience.

Joe Insley 13:46

Just showing the cylinder with vortices and things mixing around is somewhat interesting, but to give it context, this is one of the places where I get to leverage my art background, and leverage some of my students actually, who did some of the 3D modeling of components of the engine that they were then able to bring together with the scientific data. So show the inlet and outlet exhaust and valves moving up and down and the various pieces of the engine to give it context, before we kind of peel those things away and look inside. One of the challenges there was that this is all really big data. And so I don't want to move it to my desktop in order to use these other tools to do the rendering. It's not practical. And so we worked on building a pipeline where you can sort of do key framing and take a couple of time steps of the data, move it to your desktop, look at it in Maya, which is one of the 3d modeling packages that we've used, combine it with these other models, add lighting and materials and whatnot. And then sort of save that state, push it back to the supercomputer where the data lives and then essentially run that same pipeline in a batch mode. Or basically you just swap out the data set, then you can run that, in parallel, run multiple instances across multiple nodes of a supercomputer. Or in this case, we were using our visualization cluster to do that. So that sort of sped up the process and eliminated the need to move data, which is obviously expensive timewise, but enable us to sort of speed up that pipeline and not have to move big data around.

Sarah Webb 15:23

That's cool. I had seen an animation of that online and just sort of imagining all of those pieces coming together. It's just wild. I also wanted to ask you about M interior, which is a supernova, right? Yes. Can you talk through that one, because I had seen the animation of that one, too. And there's a lot that goes into it in terms of the different perspectives and zooming in and out.

Joe Insley 15:45

Right, so that's an ongoing project. Adam Burrows from Princeton University is the collaborator on that he and his team, but basically, there's multiple stages of the star. And there's multiple things happening at the same time, where there's this supernova explosion, where it's expanding, and then at the core, proto-neutron star. So it's condensing into this neutron star at the, at the center. And in that image, you sort of see this, this cutaway where this material that's expanding, and then and these particles that are, well, the particles that are in there are artificial, but it illustrates how, at the center, essentially, there's the particles are sort of trapped in the center where things are shrinking and cooling off. And then there's this exterior layer, where things are then also expanding at the same time. And the thing that's interesting to me about it, is that it's at such different scales, both in terms of time and space. So that particular image, where it's cut away, you see this sort of central thing that's on the order of 10, or 20 kilometers, maybe 100 kilometers across, while at the same time, if you zoom back out, the shock wave that's expanding is on the order of 20,000 kilometers. And so there's these sort of interesting happenings happening at multiple levels at the same time. So that's part of what that animation tried to show.

Joe Insley 17:09

By the beginning of the animation, we're sort of at a medium scale, and then we zoom in to look at the core, and then we zoom back out to see the whole thing expanding. So this was one simulation that that the team did, and evolve to about two seconds. And so now they're at the point where they're able to speed up some of those computations, and they're doing stars of different scales. That one was a 25 solar mass star. They've done others, a handful lower in the 9, 10 range, and then others at multiple scales in between there. And so their goal is to essentially build a catalog of all different scales and different conditions. And so now they've been able to, because they've been able to speed up some of their codes, compute much further out. And so now, they've computed out to six, seven seconds, which is further than anyone's done before. And so there's interesting things to look at those later times.

Sarah Webb 18:05

We've added narration to one of these animations to help explain how Joe and his team think through visualization, there's a link to that, in this episode's show notes.

Sarah Webb 18:16

I want to ask you about the artistic part of approaching this type of problem because you have to grapple with all of the same timescale differences in scale that the scientists do. But you're having to put it in an image, I was referring to it as the zooming in zooming out, but the way that you take slices the way that you use color. How do you think through that process?

Joe Insley 18:39

Yes, we just do, right? Like I said, it's an iterative process. It starts with a conversation with the science team: What are you interested in looking at? And, of course, everything. And so something like this, where there's so much interesting happening at the same time. And so sometimes it's a matter of you focus on one thing. And then we do another animation that looks at a different thing. This was a case where we do have a series of these little vignettes or sort of pipelines of, we know we want to look at this aspect. And so I'll do an animation that looks at that one thing for the whole time series. And we have a number of different ones like that. But then this was one where we wanted to try and touch on all of those things, right, to illustrate that there's these things that are happening at the same time or within small points of time at very different regions. And right and so yeah, exactly. We zoom in and out, we look at peeling away some of the layers so that we could see what's happening at the interior. Of course, this isn't the first animation we produced. There was a lot of experimentation in hit and miss. So the first time I did it, you flew in. Well, it was big, you couldn't see anything, right. So that was when we decided, well, we really need to clip this stuff away to show things and not just fly in. Otherwise, everything is occluded by all this these particles flying around and you can't see anything. So a lot of that is it experimentation and just trying things and see what works.

Sarah Webb 20:04

There's a style to your images that I recognize when I see them. And I wonder if that's something that you've consciously developed. Is that years of honing your craft? What is that?

Joe Insley 20:16

Well, thank you. I don't know that it's conscious. I mean, I guess it is just having been doing this for quite a few years. Now, I guess maybe I do have sort of a style to it, I guess I hadn't really thought of it in that respect before. I mean, one of the things that I that I do try and do with a lot of the viz that we produce, in particular, for animations, if this is stuff that's going to be for a lay audience, they don't know what they're looking at, right? They don't know what's important in here. So for example, the animation we were just talking about with the supernova star, I worked with Adam, the science team, and sort of developed the story of what we wanted to do, right. We wanted to see the sort of middle ground, fly in and see what's happening there, fly back out, watch the thing expand. Okay, so then, we worked on the rendering, essentially had a more-or-less finished animation. And I played it for him and recorded that conversation. I said: Explain to me. What are we looking at? And so when we stopped and paused in places, it was similar to the conversation we're having, right? It was just a conversation. And of course, half of that went over my head, when I had to ask him to repeat stuff.

Joe Insley 21:25

But then I went back and listened to that conversation, and then looked at the animation and said, Oh, this is what he's talking about here. Let's pause, put some arrows in, some additional text. Explain why this is interesting. And we went back and wrote a voiceover that included in some cases, the exact things he said, in other cases paraphrased it. And that's sort of the approach that I tried to take. Again, some of the animations and visualizations we do are specifically for the science team. They're trying to understand what's happening in it. In other cases, we want to be able to share that knowledge with the general public and present it in a way that's understandable. I think that's an important aspect of it is explaining that not just the science team is going to be interested in it. And so that's definitely, again,

sort of this iterative process I talked about. It's a collaboration with the science team to initially understand what it is they're looking for, and what are they interested in.

Sarah Webb 22:23

So I want to talk about your work as a professor, and the next generation of visualization. It sounds like you kind of worked your way into this work, because of certain opportunities that came up. How much of a career path is there into what you do now? And how has that changed?

Joe Insley 22:44

So I guess the question is, How do other people get to where I am? I think it's a lot of people bringing whatever their background is to it. I mean, I don't want to say my path is totally unique, but somewhat other people that come at it from the purely computer science side, and they learn some of the aesthetics. Other times they maybe they struggle at some of that stuff. I don't know if it comes easy to me, that's a good question. In terms of my teaching in this space, one of the other aspects that I like about my job is that we have a visualization lab here, where we have a large tile display, and we show off lots of our movies and interactivity and things like that. Sometimes it's for VIPs. Other times, it's grade school kids. And we have a great education division that does outreach to camps and stuff for various age groups, but lots of like middle school, high school stuff. And so they have a couple in particular, Girls Who Code and as a couple of others like that, that come through. The members of my team and I get to show off some of the stuff that we do, and hopefully, I think, inspire them. I talk about the fact that you know, I came from an art background, computer science and art, but my initial degree was in art. And so I still have something to contribute in the science and computer science space. And so can they. And so hopefully they get that, that even if you're bad at math, you can still do computer science and science and visualization.

Sarah Webb 24:08

So tell me about the courses you teach at Northern Illinois and what it's like to work with students on some of these projects. What do you sort of learn about doing the visualization with people who are learning and interested in the space?

Joe Insley 24:21

That data viz class I'm teaching this semester is focused on what I would term as information visualization, but there's actually a lot of overlap between the class I teach, and actually my colleagues, Mike Papka. He's my division director at Argonne, and he's also a computer science professor. And so we bring students from both computer science and art together. But for many years, he's taught a similar course in sort of data visualization, but more on the info viz side with the main difference being that because I'm teaching in the School of Art, there's less emphasis on programming and more on leveraging existing tools. In terms of bringing students into the scientific data viz space. So the program that I teach, there's other professors that teach 3D modeling and animation and those sorts of things. Some of those same students have also taken my data viz-- and I also teach an intro to programming-- course geared at art and design, I've known that we've wanted to sort of bring the scientific data viz and the more traditional 3D modeling, animation together. And so initially, I started working with some of those students through assistantships at NIU and worked on similar problems, and then eventually brought them to Argonne for summer internships, for example.

Sarah Webb 25:37

And so that's how they ended up working on the combustion visualization, right?

Joe Insley 25:41

They were pretty instrumental in helping me figure out, you know, what the logistics of there's a lot of moving parts there. And so I had ideas on how to implement it, and then I task them with-- go try that; let me know how it works. Of course, in the process, they'll discover something I hadn't thought of and bring that to the table. And I was like, yeah, let's try that. And so that's been really rewarding to one of the students that worked on that project, I originally had her as an undergrad and then she eventually also did her master's in fine arts at NIU. So she worked with me for a couple of years. And coming into it, she just hadn't been exposed to scientific databases at all. And then when she saw the stuff that I was doing, and I started involving her in it, that inspired her and made her want to follow in that path. So that was sort of a nice anecdote of inspiring your students to want to go outside their comfort zone.

Sarah Webb 26:33

That's really cool when you have those kinds of relationships, that development, you feel like somebody had an experience they wouldn't have had otherwise. I'm wrapping up. And I have a couple sort of your words of wisdom for others.

Joe Insley 26:46

You make an assumption about my wisdom, but we'll see what you have to ask.

Sarah Webb 26:50

If you could give one piece of advice to computational scientists or others to improve visualization, what would that be?

Joe Insley 26:59

Start thinking about it early. Don't wait until you're done computing to say, kay, now what I want to do with it? For example, you may not know what sort of data you need to save unless you know what you're going to do at the end. So of course, scientists that go into it with-- these are the questions I'm looking to answer. These are the statistics I need to gather and whatnot. But they may not think about how do I want to explore and visualize this data? What do I need to capture in order to do that? And so you don't want to get to the end, like, Oh, if only I would have done this differently when I ran the simulation. I can't afford to run it again. And talk to us at the Argonne Leadership Computing Facility. We have a number of different allocation programs, and so we like to reach out to teams that are just getting started. and let them know. Hey, the viz team is around; we're a resource for you. So reach out to us, and what is it that you're thinking that you want to do? How can we help you with it?

Sarah Webb 27:55

What advice would you give to someone out there who might be listening who maybe has caught the visualization bug? Or is thinking that that's something that they might want to explore? What are the critical tools that someone needs to do this work?

Joe Insley 28:08

Well, I think that somewhat depends on the science. I mean, there's a variety of tools that are out there, our team uses things like VisIt and ParaView are sort of the standard, DOE-supported, open-source, high-performance visualization tools. And those are more sort of general purpose in terms of can use those same tools for a lot data from lots of different science domains. There's others that are more domain specific, like VMD, for example, for visualizing molecular dynamics data.

Joe Insley 28:36

But in terms of being the viz guy that interfaces with all these different science domains, it's getting exposure to all of it. Again, it helps to have some aesthetic skills in terms of design and layout. And in particular, if you want to try and do animation, having some idea of how to deal with time. That sounds silly, maybe, but it's not as obvious as you might think, at times, or to learn how to tell a story. There's lots of different skills, and you don't necessarily have to be an expert at all of them. I think just don't be afraid to get your feet wet and try it. Science is all about collaborative teams. You bring people with different expertise together. And so there's room at the table for everyone, I think,

Sarah Webb 29:18

Joe, thank you so much for your time. This has been such a pleasure.

Joe Insley 29:23

Well, thanks for having me. I had a great time.

Sarah Webb 29:25

To see more of Joe's still images and animation and learn more about projects you've worked on. Check out our show notes at [science in parallel.org](https://science.in.parallel.org).

Sarah Webb 29:35

Science in Parallel is produced by the Krell Institute and is a media project of the Department of Energy Computational Science Graduate Fellowship program. Any opinions expressed are those of the speaker and not those of their employers, the Krell Institute or the U.S. Department of Energy. Our music is by Steve O'Reilly. This episode was written and produced by Sarah Webb and edited by Tess Hanson.