***Code Together* Podcast TRANSCRIPT - Episode 29: *SciVis Unveils a Billion Cells, Covid-19 & Invisible Monsters*** [25:51]
Host: Brenda Christoffer, Intel
Guests: Teodora Szasz of University of Chicago & Donna Nemshick, Intel Corporation

**Brenda Christoffer** (00:04): Welcome to *Code Together*, an interview series exploring the possibilities of cross architecture development with those at the forefront. I'm your host, Brenda Christoffer.

**Brenda** (00:16): Scientific visualization is providing new insights in healthcare and advancing medical research in ways that we couldn't imagine a decade ago. It's an evolving and promising field where extracting visualizations from very large data sets can improve analysis, diagnosis and patient outcomes, while reducing costs. Potential usages are abundant, but there are still challenges to be overcome. Today, we'll hear from two guests who are working in scientific visualization in different ways. We have Teodora Szasz a computational scientist at the Research Computing Center of the University of Chicago, where she also manages its visualization lab. Dora leads several teams working in computer vision, deep learning and scientific visualization. She is also a Google Cloud research innovator, part of an international team of experts that use cloud technology to advance research. Welcome Dora.

**Teodora (Dora) Szasz** (01:15): Thank you for having me.

**Brenda** (01:18): Our next guest is Donna Nemshick. She's the performance validation lead in Intel's advanced rendering and visualization architecture group. Donna leads performance analysis of our company's oneAPI rendering toolkit components on many types of Intel platforms; be it clients, desktops, servers, HPC clusters, and even GPU-based systems. Donna, thanks so much for joining.

**Donna Nemshick** (01:45): It's good to be here Brenda.

**Brenda** (01:47): So let's get started. Dora, could you share how you got started in scientific visualization?

Teodora Szasz (01:53): Of course. And I like to start with how I joined the University of Chicago. It has been a interesting journey. So I joined the University of Chicago in 2016 after finalizing my PhD in France in medical imaging. And, here at the University of Chicago my role is to help researchers across different departments. You name it – economic, medicine, chemistry, public policy - there's no boring project, I promise. Now, there are also lots of challenges. I remember in 2017 when I joined the surgery department at the University of Chicago, they contacted us and they brought a billion simulated cells data. We were trying to visualize it and I was trying a lot of libraries, openCPU, OpenGL. In 2017 in May, I participated in this wonderful hackathon that was organized by the Texas Advanced Computing Center (TACC) on visualization.

**Dora** (03:22): And there I met Jim Jeffers from Intel, and Paul Navrátil from TACC, Jeff Amstutz who was actually developing on Intel® OSPRay Library, and we were thinking well let's give it a try. So, I have the billion cells data and we have a new library. It was like magic just plug and play. And I could see the beautiful visualization of a billion cells in multiple colors, multiple orientations. I called immediately the University of Chicago and I told them: "you should see this." The only thing that was missing was a very large screen to visualize this data. Now I was very lucky in 2018, I led and designed the installation of - we call it *the wall of knowledge*, which is a very large 25 x 9 feet screen that contains 18 displays and it's capable of rendering 38 megapixel resolution. It's also touch screen, so we are able to rotate and zoom in and look at different patterns of the cells it was just amazing.

**Donna** (04:53): Wow, that sounds incredible. So visualization has really changed your research dramatically. With the big data sets and being able to visualize them and manipulate them on this giant wall of knowledge, that is truly amazing. So you mentioned touchscreen and the ability to maneuver through the simulations of these objects. So, I'm curious, how has COVID impacted your work and not being able to stand in front of that screen and do your research and really understand the depths of the medical processes you're looking at, and how's COVID changed things for you?

**Dora** (05:30): I am very glad that you ask this question, Donna. I have lots of stories around how we transition from working in the lab to remote work. I remember the last day before Covid I was there and we were just able to reproduce this impressive visualization of radiation of a massive star. The massive star is a hundred to thousand times bigger than the sun. So I was there with Jim Jeffers and Joe Insley from the visualization team at Argonne. And they share with me the data and I was able to reproduce it with ParaView on our big screen. I was just getting starting to being very excited when pandemic hit I couldn't go anymore in the laboratory and in fact in the office.

**Dora** (06:36): So my transition was mostly towards projects that I started before the pandemic on computer vision. And one of the projects we started was measuring structural inequality in childrens' books. I love this project. It's basically we are – with my team - looking at the images in childrens' book and it's a lot of diversity there – in illustrations and pictures, different resolutions. But what we are interested in is - we call it the 'invisible monster.' And it is this discrimination that is built in social structures and institutions that we really don't really notice.

**Dora** (07:32): So we really want to look into how the images in the books – in childrens' books - really influence what children think that's possible, not only for them, but also for the others. So for example, if a man never sees a woman as president, then he might think that this is not even possible to vote for a woman as a president. This is really, really important. Now when Covid hit, I got really anxious and I was really considering at one moment going and studying on the front lines as a nurse maybe. But then I realized that will my skills on medical imaging, I could volunteer on supporting the Department of Pulmonary Critical Care at the University of Chicago that were working on Covid data. So I reached out to them and am still working nowadays with them. Otherwise my longtime projects are really on prostate cancer and breast cancer lesion detections using deep learning and on visualizing this data in more interactive ways.

**Donna** (09:02): It was great that you were able to almost reinvent yourself and take on some new projects still with visualizing and imagery in the context of diversity, that must be a really exciting project for you. And I was wondering, has COVID led you to research directly related to COVID itself. Were you able to latch into some projects there so that you can explore this disease that has really impacted everyone on the globe?

**Dora** (09:34): Yes. At the beginning of the pandemic, I reached out to the Pulmonary Department at the University of Chicago to use my medical imaging expertise. And I started working with them and am still working with this department. It is a project that uses artificial intelligence to predict the maximal oxygen support that will be needed for COVID-19 patients based on their chest X-ray and their clinical data. So why we are interested in looking at the maximal oxygen support? Well, if you remember at the beginning of the pandemic, it was a crisis on hospital beds, and I.C.U beds and not everyone was lucky to have the really good care that they needed when they got infected with COVID. So this team is the pioneer in placing the critical patients into we call them: oxygenated helmets instead of pursuing the more clinically-involved options of using ventilators, which requires going in intubation and taking a patient to I.C.U for a certain number of days. So the team looked at having patients on helmets.

**Donna** (11:07): Did I hear you say helmet? Like they wear a helmet on their head?

**Dora** (11:37): Yeah, I was imagining the motorcycle helmet at the beginning at the beginning of the project as well. But when I got closer to see how works actually. So this helmet that surrounds the patient's entire head to supply oxygen. And it fills with the nice soft airside collar that wraps around the neck. So, maybe they could go in the motorcycle with the helmet, but I'm not so sure.

**Donna** (11:45): That is incredible. What a novel idea, to be able to give acute care and advanced notification to patients who may need intensive care, and there's minimal beds and what a great research. Wow, that's amazing.

Let's circle back a little bit more on the visualization, and you mentioned big data sizes and clusters. Can you share a bit more insight on your computing infrastructure?

**Dora** (12:14): Of course, and I could talk for days about this. Our cluster is named Midway, but this year we launched Midway3, which is new high-performance computing cluster of over 10,000 cores and together with Midway, we make around 40,000 cores. What's special about Midway3 – and I'm very excite, because I can use it for all of my projects, it is dedicated for artificial intelligence extensive jobs. It's also the first time also we build a system that is both Intel and AMD processors. When I say dedicated for artificial intelligence, it means that we have top GPU capabilities nowadays. We have NVIDIA A100 and V100. I can really see the performance going super up.

**Dora** (13:20): The other system that we launched is Skyway. And because I am a Google research innovator this year, I really love Skyway because it allows people working in a cluster to connect to a virtual machine in cloud and it's just beautiful. We don't need to know or to provision cloud resources anymore to cloud, we can just do that through the supercomputer. So it's really improving productivity and these are basically the two main systems that we have at the university. Of course my favorite is still the wall of knowledge. And that's something that I really look forward to go back and keep visualizing more data.

**Donna** (14:16): How has performance impact your research, like the performance of this infrastructure?

**Dora** (14:23): Yeah, well because of Midway3 and because all my projects right now include deep learning, I can really see when I go and train a model, instead of waiting for training to happen in two days, now it will happen in several hours. New GPUs are really crazy. And it's also great that the fact that this year the University of Chicago team was the first to generate a computational model on the COVID-19 virus. This was used - they used Midway and Frontera system at TACC to create this model. So, it's really very exciting.

**Donna** (15:14): So then, on top of the infrastructure, we've got this layer of software, and you had mentioned here the oneAPI Rendering Toolkit. How do you use the ray tracing APIs in your research?

**Dora** (15:28): These days, we are mostly looking in the medical field. We're looking at simulating how cancer lesions might develop over time or how the tumor is growing in certain areas of the body. And using imaging, we know the location of the tumor but also using simulation and basically knowing the mechanical process on the cancer growth, we can visualize how the cancer is spreading in the surrounding tissues. oneAPI Rendering Toolkit would be really good to use in visualizing the structure and segmenting these shapes and watch over time by simulating how they could develop in surrounding regions.

**Donna** (16:29): Do you generate a whole lot of data and look at it at the backend, or do you use in-situ type simulation where in real time you're observing the data as it's being generated?

**Dora** (16:43): I'm very happy that you asked this question Donna. How I like describe in situ to people that do not know what in situ is... imagine you're buying a car, and give to the manufacturer all the specifications: color, seat, what the noise, what music system, everything; and the manufacturer comes back to you and six months later and says, here's your car, bye. Well, we don't want that. So when we buy a car we want to go and visit maybe every week and check on the progress, adjust some parameters, and see the progress. So, more and more we transition into having simulation and generating visualization realtime as the simulation is progressing and be able to visualize it. That's what in situ is. We can check on results every time and we can go back and change parameters and see what's the progress of the simulation – if it is in the direction we want. Right now, we use ParaView with Catalyst to do in situ visualization.

**Brenda** (18:07): So Dora, it's really fascinating how you've really opened up how scientific visualization has worked in the research that you're doing and really helped out the university in advancing healthcare. Now, you mentioned all of these tools that enabled you to turn those data sets into something visual and Donna, you work on a lot of those tools: the Intel oneAPI Rendering Toolkit that have helped enable that. Could you tell us a little bit about that tool and the work that you do on the teams to make them work for those who need to use them?

**Donna** (18:46): Yes, absolutely. Thank you Brenda. So what I do and the advanced rendering and visualization architecture team is those components that, you know and love out in Open Source, be it Intel Embree, OSPRay, Open Volume Kernel Library, OSPRay Studio, Open Image Denoise; I take those components and add a little layer of glue on it, I call it. To format them into the oneAPI Rendering Toolkit bundle. And what that all means is you get the same great APIs, no change. I pull the source code from the GitHub releases and I distribute easy-to-use binaries in a toolkit bundle. So you go to the oneAPI Rendering Toolkit search on Google and get it now button is right there on the page for you. You click that button, and you will have this toolkit install on your Windows machine, on your Linux workstation, and the installation will download everything you need to run those libraries all in that one button click. And that's the ease of use that we've brought to the user community through, the one API Rendering Toolkit and then the ability to access our tools in Yum and Apt repositories, there's a whole set of other repositories that folks can download them from. So, you can get them anywhere these days and make your development real streamlined, make updates of the components within your applications very easy at the click of a button.

**Brenda** (20:23): Super. So as we wrap up, Dora, could you tell us where do you see the future heading for scientific visualization in your organization?

**Dora** (20:33): Oh – real-time visualization and being able to get immersive experiences – that's where visualization will go. And as I'm hearing you Donna, you mention how the OSPRay library and oneAPI framework and toolkit are really easy to install and plug and play. I wish we could design in academia the same tools that are easy to install and reproduce. This year maybe because I had time to just sit and be more with the code, I got really into advocating for bringing more industry resources and best practices for writing frameworks and code and creating products into academic. So having reproducibility and great documentation as the OSPRay Library is when I look in GitHub and same level of collaboration between institutions. I think this will really – it's an area where we need to put more energy. And sometimes it's hard to see that. In academia great researchers are leaving with their code so nobody are able to continue and collaborate and there is not documentation. So yeah, I wish we can learn more from Intel and bring these best practices in our setup in academia.

**Brenda** (22:25): Donna, as one of our technologists, where do you see the future heading for visualization?

**Donna** (22:30): So I am thinking big, big, big, and big. I'm thinking big data, data getting even bigger, big compute power on HPC clusters, big storage capability, factoring in the powerful [Intel®] Optane™ Memory that can drop right into your machines and expand your memory footprint into the terabytes. And then there's this big screen displaying this incredible stunning high-fidelity images that enable this groundbreaking research, by way of using the components of the oneAPI Rendering Toolkit. I have to say Dora, I would really love to one day stand in front of that wall of knowledge and manipulate some of those incredible research images that you get to play with all day long. I've got oodles of experience in writing embedded code on HPC switching systems, and I've worked on mass storage performance and memory generation for IC's. And I tell you, working in this visualization space, this just brings it all together for me, and it's just one big playground. It's exciting. It's an exciting space to be in, and I think the best is yet to come in the visualization space.

**Dora** (23:48): I do agree Donna and please come and visit our visualization wall anytime.

**Brenda** (23:57): So, Dora, where can people go to learn more?

**Dora** (24:01): In the research computing center web site at RCC.uChicago.edu - You can find here resources on information about our systems and lots of workshops that we give on visualization, Python, GitHub and so on. Our visualization laboratory web site is: vislab.rcc.uChicago.edu where you can find some of our work on visualizations there. And of course, please connect with me on LinkedIn.

**Brenda** (24:31): Super. And Donna, do you have any places you want to recommend as resources?

**Donna** (24:45): Absolutely. Very easily google search oneAPI rendering toolkit, and you will be led to our homepage, which will get you access and deep dive into all the components that comprise the toolkit. Likewise, you can go to bitly.com/rendorkit, and it will transport you to the same page. And follow me on Twitter at Donna Nemshick @dnemshi, and do connect with me online and LinkedIn.

**Brenda** (25:12): Wonderful. Well, I greatly enjoyed this topic in conversation with you both. Dora, thank you so much for joining us today. Your passion really comes across and I definitely enjoyed it.

**Dora** (25:25): Thank you so much. It was a pleasure.

**Brenda** (25:29): And Donna, great to have you with us. Thank you so much again.

**Donna** (25:34): Thank you, Brenda. This was a lot of fun.

**Brenda** (25:34): And a big thank you to all of our listeners for joining us today. Let's continue the conversation at oneapi.com.