Tech Leadership Standing Panel: Solving the Data Problem for the Climate Crisis Part 1, Data for Global Climate Knowledge

Thursday, August 10, 2023. 1:00 - 2:30 pm (ET)

Satellite data has been the cornerstone for observing, modeling, and predicting the course of climate change throughout the globe. This panel will discuss the state of earth observation data, the limits of current knowledge, and strategies for improving our knowledge and models. In addition to reviewing the state of well-established earth observation programs in the U.S. and Europe, the panel will discuss innovative programs from government, academia, NGOs, and the private sector, the potential for satellite observation at a granular local level, and the potential to supplement satellite data with on-the-ground or in-the-sea observations.

Moderators:

- Theresa Pardo, Ph.D., Associate Vice President for Research and Economic Development, University at Albany, State University of New York
- Joel Gurin, President, Center for Open Data Enterprise (CODE)

Panelists:

- Robert S. Chen, Ph.D., Director, Center for International Earth Science Information Network, Columbia Climate School
- Priya L. Donti, Ph.D., Co-founder and Executive Director, Climate Change AI
- Jed Sundwall, Executive Director, Radiant Earth
- Cecile S. Rousseaux, Ph.D., Research Scientist, NASA

Transcript

Dr. Theresa Pardo: Hello everyone. Welcome to the *Standing Panel on Technology Leadership* from the National Academy of Public Administration. My name is Theresa Pardo. Can I have a slide, please?

I am pleased to be here today cohosting this important panel on behalf of myself and my co-chair for the Technology Leadership Panel team, Alan Shark. Alan and I have worked together for a number of years as the co-chairs of the Technology Leadership Panel of the National Academy of Public Administration and we are thrilled to be able to present the first of a two-part series on climate action and data.

Before we move into that program, the core of our program, I wanted to give a little bit of an overview of the National Academy for those who may be less familiar with our program.

NAPA was established in 1967. It is an organization that focuses on solving public management challenges. We bring together thought leaders to look at key issues of the day, key challenges, key opportunities and I think today's session and the first part of this two-part series represents all of those things.

It is a key issue of the day. It is a challenge and we will speak to a wide range of opportunities that are presented to us from our speakers as we think deeply about how data can be more fully, more robustly and more effectively applied to our efforts in climate action.

NAPA is a membership organization. There's a group of about 1000 elected fellows and our goal is to bring together independent thought leadership and in-depth analysis. Slide please.

One of the key initiatives of NAPA is the identification of these 12 grand challenges and I think you can see very easily, very quickly that the topic of this new series that we're presenting maps across many, many, if not all of these grand challenges, certainly going right to the end of the list having to do with AI, social equity, resilience, communities, et cetera.

So this is I think such an important conversation related to so much of the work that's happening at NAPA but really even more so, so much of the work and the thinking that's happening across the organizations represented here on the panel today and by so many of the people who have registered and will be attending this session.

Next slide. So a little bit about how we actually carry out this independent thought leadership at NAPA under the direction of our president Terry Gerton who's here with us today.

So a series of initiatives, a center that focuses on inter governmental partnerships, an agile government center that's helping us think about the – how to be more agile in our efforts to innovate within the governmental context. Forums, active forums, events and in particular for today's purposes a series of standing panels, a set of standing panels, which the technology leadership panel is one.

So you have a chance, go in and look at the NAPA website. Each of the standing panels has a long series of different sessions that we've held. Our panel in particular has had sessions on AI, on heat, data, all of those topics relevant in particular for our session here today.

Slide. So this session, it brings us to the work that we're doing here today, which is focusing on the data problem, in particular in terms of the climate crisis and in a minute, I will introduce you to Joel Gurin who is my colleague in organizing and presenting this series.

But our focus, Joel and I for many years, have worked together on questions related to data. How can we more effectively, more fully understand or use data? What does it require us to do in terms of organizational innovation policy and process innovation? What are the capabilities that are required? How do we bring data and data governments to conversations where it didn't exist before?

So there are so many more topics related to data and solving the climate crisis and I think you're going to hear from our panelists today. We designed this as a two-part series, *Data for Global Climate Knowledge* and *Solving the Local Data Problem for Climate Action*.

So we're so excited to have tremendous interest in today's session and registration already for the September session. This is a topic of great interest to so many people, to so many organizations and as we think about the grand challenges of NAPA, this is one of the grander that branch out that NAPA has identified. So many of us are thinking about data and the climate crisis and potentially the role of AI in leveraging our data to take specific and impactful action.

So slide please. So as this slide says here and as you've seen before, my name is Theresa Pardo. I'm at the University at Albany. I also previously was the director of an organization called the Center for Technology and Government where I'm currently a senior fellow and in these various roles, as I said, I've worked with my colleague Joel Gurin.

Joel, let me introduce you as the President and Founder of the Center for Open Data Enterprise. He's going to talk a little bit about open data enterprise and then begin to introduce our panel and our panelists. Thank you very much Joel for being here. Joel, the floor is yours.

Joel Gurin: Oh, Theresa, thank you so much. I'm really delighted to be able to work with Theresa again on a great project and all of us at CODE are really honored to be able to co-host this session with NAPA on such an important topic.

I will just tell you a little bit about our organization. We're a nonprofit based in Washington DC. We started in 2015 with a mission to maximize the value of open and shared data for the public good and we do this through a combination of research, conventional white papers, developing online tools and resources and very much by doing convenings, round tables in particular that

bring together data providers, data users and all kinds of stakeholders to really figure out how to put data to the best possible use to solve important public problems.

Doing this work, we've collaborated with the White House, with about 10 different federal agencies, with a number of international organizations including the World Bank and the UN and with different national governments as well and over time, we've come to focus on three major areas, although we cover a lot of topics.

One is improving health and healthcare. One is the use of data for improved equity and the third is the use of data to address the climate crisis which is a huge priority for us as it is for so many organizations and so many of us.

We've been privileged over the last few years to partner with NOAA, to work with the OECD, to work with the Bezos Earth Fund and others. You can read more about our work if you go to our website opendataenterprise.org and click on the "Impact" page. You will see a lot of summaries of the work we've done in climate and other areas.

Next slide please. Just to set the stage for the conversation we're going to have today, we really hope it will be a conversation. We're focusing today on global climate data and as Theresa said, the next session in this series in September will focus on local data as well because we really see the problem of using data for the climate crisis as a very – actually somewhat of a wicked problem of how to bring together the kind of global knowledge that we have with local data that's also necessary.

A large part of this is because we're now moving from large scale understanding and prediction and modeling of the climate crisis to the need to take action. We have global data that is increasingly helpful in identifying what the issues are that the world is going to face, even identifying that at a fairly granular level where you can see what's happening in different localities.

But when we talk about how do we mitigate climate change, how do we adapt to it, how do we become more resilient, you need not only that kind of data but you need data on very local topics like the strength of housing infrastructure, natural disaster preparedness, financial issues, demographic issues, all these kinds of things.

We see this as a two-part session because there are at least two parts of this big puzzle and the one that we're going to focus on today is what can we really learn from data that is collected, managed, distributed, analyzed and used for modeling on a global level and most importantly, how can we begin to transition from using this kind of data for prediction to really using the data for action.

We have a fantastic panel. We're going to be talking about all kinds of innovative programs in government, academia, NGOs and the private sector and we're going to be talking not only about satellite but about ocean data and other kinds of global data as well.

The next slide shows you who our panelists are. We have Dr. Robert Chen, Dr. Priya Donti, Jed Sundwall and Dr. Cecile Rousseaux and why don't we take the slide down and we can actually see all of our panelists in person. I will just take a minute to introduce them all.

Dr. Robert Chen is Director of the Center for International Earth Science Information Network at Columbia University's Climate School. He managed the NASA Socioeconomic Data and Application Center for more than 25 years until last month and he co-leads the Data Distribution Center for the Intergovernmental Panel on Climate Change, the IPCC.

He has worked on climate impact assessment and related data issues since late 1970s. He received his PhD in geography from the University of North Carolina Chapel Hill and his MS and BS degrees from MIT.

Next Dr. Priya Donti is the Co-founder and Executive Director of Climate Change AI, which is a global nonprofit initiative to catalyze impactful work at the intersection of climate change and machine learning. She's also an incoming assistant professor at the MIT Department of Electrical Engineering and Computer Science and she was previously a US Department of Energy Computational Science graduate fellow, a Siebel scholar, an NSF graduate research fellow and a Thomas J. Watson fellow.

She received her PhD from the Computer Science Department and the Department of Engineering and Public Policy at Carnegie Mellon University and she attended Harvey Mudd College for her undergraduate degree.

Jed Sundwall is the Executive Director of Radiant Earth, an organization dedicated to increasing a shared understanding of our world by expanding access to geospatial data and machine learning models.

Before joining Radiant Earth, he created data sharing and sustainability initiatives at Amazon including the AWS Open Data Program, which hosts more than 100 petabytes of open access data.

He has helped create data sharing best practices that have been adopted worldwide by NASA, the US Geological Survey, Google, Microsoft and other institutions around the world. He serves on NatureServe's Board of Directors and he has a Master's Degree in Foreign Policy from UC San Diego.

Then finally Dr. Cecile Rousseaux has been a research scientist at the NASA Goddard Space Flight Center since 2011. Her research focuses on the role of oceans in the carbon cycle, the

effects of climate variability and trends in the ocean biogeochemical cycle and the support of upcoming field and satellite missions such as the PACE mission which will focus on ocean ecosystems beginning next January.

She received her PhD in Environmental Engineering from the University of Western Australia. Her bachelor and master's of science degrees from the University of Namur in Belgium and a Master's Degree in Oceanography from the University of Liege in Belgium. So we're really, really honored and delighted to have all of you here and I also want to acknowledge and thank Jillian McGuffey, Senior Research Associate at NAPA and Matt Rumsey, our Research and Communications Manager at CODE who have done a great amount of work to plan this event and are here today to help this all go smoothly.

So without further ado, Theresa, I think if you want to begin the conversation, we're going to start with some questions for each of our panelists individually but really I think very quickly evolve this into a conversation and as Theresa said, with a lot of opportunity for Q and A at the end.

Dr. Theresa Pardo: Great. Thank you very much Joel and let me reiterate my thanks to the team for getting us to this point into our panelists. So we're going to start with Bob and Bob, we've talked for many years about the work that you do all over the world. All the scientists that you're working with who collected new satellite data to try to understand the climate crisis.

As you're working with these folks across the international community, I'm sure that there are numerous priorities that emerge. Tell us what you see as the key priorities for the application of the data to action on climate change. So understanding what's happening, great. Translating, understanding into action. What are the priorities that we need to focus on?

Dr. Robert Chen: Thanks Theresa and hi to everyone. I would have the opportunity to talk about this issue and yeah, I think it's a really important transition that really the whole world is in, of thinking about climate change sort of in theory or as a problem of the future to one that's certainly happening all around us.

Coming out of the science community, I think as maybe you're familiar with in other areas, there is a big difference between the kind of data and information that scientists need and use to kind of understand and establish a problem, try to narrow the uncertainties and think about the implications and the kind of data you might need to take action.

I mean this series is about that transition but the things that we've done for many decades to try to – first of all groups like NASA launching satellites to take observations and NASA is a research agency. It's not an operational agency.

So the satellites that were launched were considered to be research tools collecting data to support climate models, to support all different types of interdisciplinary analysis to better understand climate change and human interactions.

But easily more than 10 years ago, NASA did realize that these data have practical applications and a core series of transition between agencies like NASA and NOAA which is more of an operational agency and that handoff between research and meeting every day, operational needs of the weather service or of other businesses and practical things.

That has been evolving I think over the last couple of decades but there's still a gap. So just to take one example, the climate research community really focused on trying to understand long term impacts of climate change. They use what they called "equilibrium climate model runs" to understand the future state of the climate and when carbon dioxide is doubled, they're tripled or whatever.

There was less emphasis on things like what scientists would refer to as "transient climate model assessments". It's not only what's your starting point and your endpoint but what does the climate look like on your way there. It's actually not safe to assume, as I think we're seeing, that going from one state to a future state is this linear, smooth thing.

The transients allow – reflect the fact that the climate is very – system is very dynamic that you can have a lot of variability both over time and across space and extremes can occur that you might not expect if you just assume a nice trajectory and I think the kinds of phenomena we've been seeing of drought and heatwaves and fires and glacial changes reflect some of that uncertainty about how transient changes in climate might differ from what scientists originally studied.

I think you've identified the issue of needing local data to make decisions, not just kind of global or data for large regions. The one thing I would add is some scientists haven't been all that good at linking the projected environmental changes with what's happening to specific populations on the ground and part of that is when you're looking at 100 years ahead, I would say the climate, it's hard to know who's going to be living where and under what conditions.

But now that this is a much more imminent set of issues, I think it's important to do a better job of linking the understanding of climate change and variability with who's living where and under what conditions. That's an area that I've worked in a lot, which is how to use our occupations to better understand things like the built infrastructure and the resilience that society is based on, networks and structural protections provided by our infrastructure and so forth.

So I will stop there but I think those are some of the points that illustrate the difference between scientific data and the kind of practical, actionable data that decision makers need.

Dr. Theresa Pardo: Great. Thank you very much for starting us off so well Bob and I think this question about who is living where and under what conditions and how and in what way are the - certainly how are those conditions changing both in the short and the long term? How can we do a better job at understanding that?

So let me at this point open it up and invite your fellow panelists, if anyone would like to respond either to the initial question having to do with the priorities for applying global data to climate change, action climate change or to respond or comment on Bob's remarks. So Priya, Jed, Cecile.

Dr. Cecile Rousseaux: I'm happy to add a little piece of information or maybe it's just a rephrasing, I'm not sure. But I think we are really moving to a direction where the people that are producing the data and the people that are making the decisions, including local communities, need to have a better dialogue and need to have a dialogue that opens up way earlier than before. We at least – here at NASA, we are trying to make that effort already but that communication needs to happen before we even launch a satellite or before we go out in the field and collect data because we need to know what type of data they need and what type of data is missing to make those critical decisions.

So I think we can all do a better job at that and it can be uncomfortable and it can be hard because there is a difference of language here basically. But I think this is something that we all have to try and push yourself to be a bit uncomfortable so that the data at the end are very useful and we can translate what we are doing or what we are collecting with what's actually needed.

Joel Gurin: You know, I think that's such – what you're both describing is really kind of a sea change. I mean given the history of satellite data and the collection of this kind of data is really being seen as a scientific research enterprise, the idea of really creating new data and collecting new data through this kind of interaction or with this kind of input from the people who are affected is – you know, is a major change but I think a very positive one.

Jed and Priya, I saw you both kind of nodding. I wonder if either of you would like to go first and jump in and maybe talk a little bit about – I know both of your work really relates to identifying data gaps in global data and also using this kind of data for AI in some similar ways. But Priya, would you like to begin?

Dr. Priya Donti: I think Jed unmuted, so go ahead Jed.

Joel Gurin: Oh, OK. That's fine. Jed, you go first.

Jed Sundwall: Yeah, no. I do. I just wanted to respond enthusiastically to what both Bob and then just what Cecile just added that like it is uncomfortable. These conversations need to happen earlier. I have a few thoughts on this. I mean one is that like there's – we've had a great

example actually from the European community so that the Copernicus Program, which people in Europe will yell at me if I get this wrong. The European Space Agency, I think, they fly the satellites but I think it's an European Commission program.

I may have not gotten that right but they fly a satellite constellation, produces phenomenal data, all openly licensed. You know, so it's free to use. But the program was explicitly set up to support European policy goals around agriculture, commerce, sustainability and environmental conservation and things like that.

So it was set out at the beginning that this was going to be a tool, an instrument that they can use to achieve policy goals. That doesn't always happen and I think – I know. I mean Joel and I have talked about this for years. There are a lot of open data initiatives that assume – so we're going to make some data and make it open and then like something good will happen.

That doesn't always happen. I think it's very good to be deliberate about it upfront and I think speaking to the discomfort that Cecile pointed out, this might be somewhat provocative but I will just say that there are many political organizations that don't really make decisions based on data. They have completely different structures they use to come to decisions and the idea that if we merely show these people better data, they will make better decisions.

I think we might be deluding ourselves if we think that sometimes. We need to be very deliberate about how we take data and turn it into metrics that people can use to make decisions and to inform programs that are driven by metrics to make decisions.

The notion that like – I think there's a myth that's very seductive. It's the idea, it's like if only I had more information or if I had the right information, then I would know what to do and I would make the right decision. I can tell you from my own personal experience that I know that I shouldn't eat as many doughnuts as I do and I have all the data I need about that decision but that's just not how I work. I think most humans don't work that way and certain political systems don't work that way.

So that's where I think a lot of this discomfort comes from but I think we should acknowledge it head on. So I will stop there. I can keep going on this but …

Dr. Theresa Pardo: Right, right, because I love this focus in my – when you think about ultimately the purpose is to inform decision making. But it requires us to look at how are decisions actually being made. What is the decision culture in different organizations? Priya, you wanted to join in on this conversation as well.

Dr. Priya Donti: Yes, absolutely and I think that the focus of today is global data and satellite data but for that reason, I think that it's also important to really think about how does satellite and more global data complement the kinds of on-the-ground data we have, the kinds of

engineering-based data, performance of our buildings, performance of our power grids, the kinds of control or optimization kinds of workloads we may want to enable.

I think often when we talk about this conversation on data and climate, the first thing that one's mind goes to is how do we monitor and I think a lot of what everybody is saying is it's not just about monitoring and understanding. It's also about acting and that usually requires some combination of not just kind of earth observation data but also many types of on-the-ground data that are pertinent to actual like mitigation and adaptation goals in the climate sphere and there too we've talked about kind of co-design of data collection and kind of launching data initiatives but then also the discussion of co-design with effective stakeholders comes up a lot in – when you're implementing algorithms based on data in order to solve some kind of problem. You also want to understand how that's going to affect various stakeholders, what problems actually need to be solved and so there becomes this whole like data collection, like data analysis and stakeholder co-design process that I think it becomes very important to consider.

Joel Gurin: Fantastic. Well, this is a great conversation and a great start to this conversation and I think you're all saying really very similar kinds of things that are – I think for those of us who may not be in the climate research community are encouraging because you're talking about a way of really kind of co-creating research and knowledge with people who are going to be impacted.

One question I would like to pose to any of you or all of you is so we – it seems like we all kind of agree that the problems we're facing are going to require a combination of this kind of global data that we've been working with for years with perhaps new kinds of data on local conditions and how do those come together to determine or to support strategies for mitigation adaptation or resilience.

One place to start is by asking what are the limits of what we can now get from global data in terms of local granularity. So for example after the last COP meeting, Climate TRACE got a lot of – the program Climate TRACE got a lot of attention because they are claiming to be able to really zero in and pinpoint different kinds of activities that can impact the climate on quite a local level and we're certainly seeing increases in resolution over time.

Are we getting some global data now that really can give us insights that a local city or a state government official might be able to use? And what are the limits of what we can get from earth observation or ocean observation? We will come back to as well.

Dr. Robert Chen: I can take a quick stab at that. I mean I think the long-standing differences in what you could do globally versus locally are really shrinking, given the advances in computing and data science abilities.

So we've had a partnership with the Facebook Meta data for good activity for some years and they're able to process for the globe submeter earth observation data and they were able to

bring to bear probably 100 times more computing power than your typical academic center can get to.

That was not the barrier for that kind of group to be able to handle such massive amounts of data and produce something that's very – it's available on the Humanitarian OpenStreetMap website. But it's very useful and it's very detailed.

Now the limit is how do you validate for localities and how do you make the data more customized to what is needed and that's still a big challenge but it's I think becoming less of a challenge given technology and access. So that particular distinction I think is changing but they're still – not everybody has access to that kind of computing capacity and ability to handle such large data sets.

Joel Gurin: Thank you. I was just going to say Bob, because you just also mentioned Humanitarian OpenStreetMap which is I think really kind of a massive sort of citizen science sort of crowdsourcing approach. So is part of the solution to sort of begin with satellite data at a very, very local level but then have groups like Humanitarian OpenStreetMap or others do I guess what's called ground truthing now? Sort of going out and actually sort of testing those observations more locally?

Dr. Robert Chen: Yeah. I think you are seeing a lot of efforts like that in all different fields where distributed people can really help make data more useful and reduce uncertainties and all that sort of thing. So that's still a growing area.

Dr. Theresa Pardo: That's a perfect segue to the question that we have for Priya, which is about utility, right? So Priya, you run a new nonprofit. I mean yeah, a new nonprofit that's focusing on impactful work at the intersection of climate change and machine learning.

As we talk about all the data and the data as we understand it being available to us or potentially available in new ways, what are you hoping that AI can do in terms of transforming the way we use data to understand the climate itself?

Dr. Priya Donti: Yeah, thanks Theresa. So yeah, I mean I would say that I think there is a lot AI can do to transform not just how we use data to understand the climate but how we use it to act on climate change and we see this happening in a lot of ways. So I think there has already been some discussion in this panel of how AI and machine learning are being used to analyze satellite imagery in order to understand information or let's say proxies to information that we might get via ground truthing, so things like the locations of solar panels or granular emissions estimates or pinpointing deforestation or various applications like that.

We also see use cases where machine learning is used to forecast quantities that might be helpful to provide some foresight. So these are things like forecasting solar and wind power on your power grid in order to help us better balance the power grid or in combination with

physical and weather models being used to forecast crop yields or things that will allow us to adapt to how the climate will be affecting these things.

It's AI and machine learning in addition to providing kind of information in these ways. It can also be used to help us actually optimize and control systems of pertinence to climate change. So things like controlling heating and cooling systems in buildings more effectively or in refrigeration systems or in data centers in order to reduce the amount of energy that's being used to still provide some kind of level of comfort or keeping your data center at a certain temperature.

There are also applications like AI being used to accelerate scientific experimentation by analyzing the outcomes of your past experiments to try to synthesize a battery and then suggesting which battery you should synthesize next and as a result just speeding up the number of design cycles that you need to use and then also places where AI machine learning are helping us to provide approximations to all our parts of the kind of engineering and scientific simulations that we use.

So for example in the power sector, we run these kind of computation models that are to understand how we balance a power grid and machine learning is being used to try to approximate parts of these simulations in order to allow us to run them more quickly and we see this also come up in for example climate science related areas.

So I would say there are lots of ways that AI can play a role and I guess just quickly on this point of ground truthing when it comes to kind of analyzing earth observation data. This is super important because in some sense when we are using AI and machine learning to analyze earth observation data, unless we have a certain level of granularity in the raw data, we are ultimately learning proxies to our ground truth. We're not actually learning the ground truth.

Even though we can do some ground truthing but collect some granular data in order to calibrate our models, it's also worth noting that the on-the-ground situation doesn't look the same everywhere as if you're giving your machine learning algorithm data that represents certain geographies more comprehensively. Then your algorithm is extrapolating the assumptions associated with those geographies to everywhere else in the world where ground truth data doesn't exist.

So it's really important to understand with your earth observation data and your ground truth data and any physical and local and domain knowledge you may be bringing in where you still might see gaps in terms of how well you can expect these methods to scale and extrapolate and kind of ensure that you're taking that into account accordingly either by collecting ground truth data in more places or by doing things like incorporating your physical knowledge or other things that you know do kind of generalize across contexts.

Dr. Theresa Pardo: Really exciting and it sounds like the big part of that, the strategy here is taking advantage of the data that Bob is talking about where we have increasingly local level data that gives us kind of the ground truth, if you will, in more places than we might otherwise have had in previous years.

So Cecile, Jed, Bob, would you like to weigh in on this question as well? What we're asking is how do you hope that AI will transform the way we use data to understand the climate or to respond to a comment specifically that Priya made?

Dr. Cecile Rousseaux: I'm just going to note that it's funny. We're talking about ground truthing and it's really quite a different situation in the oceans than on land because for us the ground truth, it's very expensive. We can't just say, hey, if you by any chance have your sailboat at back, can you take this instrument and collect some data? But I don't have that many friends that have that.

So that's definitely something that the ocean community is struggling with is the ground truthing that we have of quality control and also there are a variety of methods out there to measure different things and I'm sure that's true for land as well atmosphere. But there is a standard that we need to have on how we measure things and I think a big challenge about those big data sets that either go into AI or models or whatever is that there's always a question – at least I'm talking about the ocean here, about the quality and like how was it collected. Who analyzed it? Is it representative or not?

So I think it's – that's part of what I was talking about when we try to integrate those different levels of data whether it's spatial, like local data into global data sets or different approaches. We want to make sure that the data that goes into that is of quality and that requires an expertise on its own to know what is it. What is the data of quality? So that was my two cents or five cents.

Dr. Theresa Pardo: Data governance comes back. Joel, did you want to jump in there?

Joel Gurin: Yeah, to Cecile while we're on the topic of ocean data and then Jed, I would love your thoughts on the AI question as well. But while we're on the topic of ocean data, I think people may be actually somewhat less familiar with some of the ocean data than they are with satellite data which has gotten so much attention. We also know that just in general research on the oceans until recently has not been maybe funded and prioritized as much as it needs to be.

So it seems like there's a lot of new approaches happening. I wonder if you could just talk a little bit about some of the most important kinds of data that you're seeing coming from the ocean particularly in the context of decision making as we were talking about at the beginning. Ocean data is going to be critical both for the global view of climate change but also especially important for coastal cities, people living in coastal communities, communities that make their living through fisheries and so on.

What are some of the most important data types that you're seeing and how can they be put together with local data for climate adaptation, resilience, for coastal areas especially that are impacted?

Dr. Cecile Rousseaux: Yeah. So the ocean community here is you have the physics of the ocean and you have the ocean biogeochemistry. So I'm probably going to be biased a little bit towards the ocean biogeochemistry because that's where I'm from. But we have had satellites measuring properties of the – physics and biology. We have had global coverage since 1997, 1998.

So over the last 25 years, we have learned a lot from those data but the measurements that they do is really just a total of what's in the water in terms of phytoplankton which are the equivalent of trees but in the water. So we really are still behind a little bit in terms of understanding exactly what's in the water and in terms of coastal area and they deal with things for example like harmful algal blooms, that specific species and we have – I'm going to do a shameless plug here but we have an upcoming vision that was cited in the intro, PACE launching in January 2024 that is going to help with that.

So I think in terms of coastal communities, that that's going to be very beneficial. Now in parallel to that, there is more and more efforts being done in integrating all those data sets. So I was mentioning sea surface temperature but we have sea level pressure and we have ocean color. Instead of just looking at one variable, we are really moving towards finding ways to integrate and AI is one of those that we are looking now into.

So how can we combine those different variables to get us to have a better understanding of what's going on in the environment because right now we are just looking at chlorophyll for example by an instrument. But if we were to combine chlorophyll and temperature or algorithm, we would be better. We would know that.

So I think the integration of the various data sets that are out there are going to be really a critical step into us being able to make better decisions and that includes satellite data sets and in situ data sets that are collected in coastal areas. But there is a – it's a very careful process on how you combine those data. So I think we have to be very careful at not just taking every data and dumping it in and let's see what happened, but I do think it's the future.

Joel Gurin: So expose new sources of data and new methods of analysis and Cecile, if you could just say. I know PACE is an acronym. If you could just tell us what the acronym is for because I don't think I mentioned that in the introduction.

Dr. Cecile Rousseaux: Thank you. I will have that … So PACE stands for Plankton, Aerosol, Cloud and ocean Ecosystem and it's going to be a satellite that will have ocean color measurement at a hyperspectral. So think of a box of crayons that instead of having five crayons has hundreds. So

a lot more information and we're also going to have two polarimeters that are going to give us information about aerosols and clouds and help us understand that and also correct the atmosphere so that we can see the ocean better.

Joel Gurin: Terrific. Thank you and that's very exciting. Jed, let's go back to you and really I guess a little open-ended but maybe tell us a little bit about how Radiant Earth is seeing the prospects for AI applied to earth observation and climate data and just how you personally see the opportunities and the challenges in using these new methods of analysis.

Jed Sundwall: Sure. Yeah. So Radiant, my predecessor at Radiant, I've been in this role for about a year now but my predecessor is Hamed Alemohammad who's at Clark University now but he was really a luminary pioneer in applying machine learning to earth observation. I don't have much to say other than to corroborate with what Priya was saying. It's just that like you need good ground truth data. You need good label data to make this effective.

If you want to produce a model that scales globally, you need what we term – I mean I don't know if Hamed came up with this term but like what we call geo-diverse data, right?

For example we help produce a very large data set in collaboration with the group called DevGlobal and we're working with the UN on this of building footprints worldwide. Buildings look different in different parts of the world and they look different from space. They have different shapes. They're clustered differently.

So if you were to train a model on how to recognize buildings in Brooklyn, it's not going to work in Lagos or name another city, Jakarta. So you need geo-diverse data if you want to produce models like this that will scale.

Now of course there are many instances where like you don't need to scale globally. You might just need a model that's like I need to be able to detect building footprints in a certain country and that's fine. You can have smaller models.

But one thing that we contend – we contend with a few issues. One is that a lot of people assume that you can just take computer vision techniques that people can see – like right now with generative AI and these things that can produce really neat images and artwork and things like that.

Those are all based on imagery that has been available on the internet. A lot of photography that humans have produced and we've trained computers to see and to recognize or to generate content based on that imagery. Satellite imagery, satellite data is fundamentally different. It's not imagery the way it – like a photograph is.

The satellite that is orbiting the planet and reading off – reading just data about what's reflecting off the planet, it's not framing pictures and taking pictures of things. It's just gathering

data and so you run a risk of – like if you just use normal computer vision techniques on satellite imagery, it doesn't really work because a lot of computer vision models assume that a picture that it's fed is of something.

There are lots of satellite images that aren't really of anything. So much sand maybe, you know. It's very hard to detect like a thing in it. So there's still a lot of education that needs to be done for funders to help them understand like how big of a challenge it is to – or how much work remains to be done to transfer a lot of like what we've learned about machine learning and computer vision to satellite imagery.

Joel Gurin: Very interesting, yeah.

Jed Sundwall: Yeah.

Dr. Priya Donti: If I could jump in there too, yeah, I second all of that and also as I mentioned, that yeah, I think on the machine learning side, there has been like in general the development of machine learning tools for very specific types of data and workflows largely as Jed was mentioning. You know, images on the internet, text on the internet, data that has a certain volume, that has a certain cleanliness to it and this is not necessarily the case for a lot of data that comes up in climate workflow.

So we've talked about earth observation data but also physical data, like data about your building or your power grid. That's not images or text. That's physics data and it also doesn't necessarily come in abundant and huge volume which a lot of machine learning methods these days assume that you have that volume of data.

So I think that kind of – for funders, it's kind of getting a sense of not just what data is necessary to push forward certain areas of application in the climate change space but also where does that – where are some of those – where is it that we're assuming certain necessity for volume or availability of data because of how machine learning has looked versus how it could look and where does it make sense to also push the methodologies to be responsive to the kinds of data that we're more likely to get.

So I think there's a bit of a codesign in terms of how we push forward methods and how we push forward data.

Joel Gurin: That's fascinating, yeah.

Dr. Theresa Pardo: Yeah, really fascinating. Bob, did you want to jump in before we move on to the next question for Jed?

Dr. Robert Chen: No, I'm fine.

Dr. Theresa Pardo: I thought I saw you, you know. So this really again dovetails – you guys, as if you practiced, right? Because it's really – you know, Jed, your work at Radiant Earth about expanding access to, expanding use of, of climate data and what are the challenges. How do you overcome those challenges? Certainly you talked about increasing the understanding of the different kinds of data for different purposes, right? We've talked a lot about that already and I love Priya's comment about not only assuming that it's the data but the tools themselves in terms of understanding how they might be changed over time.

So if you think about challenges to making data – climate-related data more widely available, but more importantly more widely used, what do you see the biggest challenge is?

Jed Sundwall: There's a pretty fundamental challenge which is how we think about – well, how we think about data products and this is like my soapbox these days is that we have to think about data products. We actually don't – like many conversations we have are about data kind of in the abstract. It's like we will produce more data. We will have the data and then we will know what to do.

I have this one weird trick that I use to really ground the conversation which is to talk about data products. Many programs – I see the way a lot of things are funded. You say we're going to fly a satellite, which produces a ton of data. But if you think in terms of like what is the data product that we need, just using the word "product" forces you think in terms of like who's going to use this thing. How is it going to be delivered? What is the cost of delivering it? What's the cost of maintaining it? What can we expect from our users? That really sharpens the conversation a little bit more I think for funders and to make things a bit more practical.

Along these lines, I also think that we are in a bit of a bind because of the way that agencies fund things, the way that universities fund things and that every agency and lab at a university and many departments within agencies, et cetera, have their own data portals. The notion is that like the unit of consumption or the way to consume data is to go to whoever has it and define their data portal and access to the data from there.

That's I think beyond broken, this approach. It doesn't work. It makes – all the discussions we've been having today about the need to combine local, let's say municipal level data or other local government data and combine that with global climate models, to do that, you've got to run this gauntlet of like finding lots of different data portals that are going to provide data in different ways and consolidate them.

But that has been true from like the early days of data portals is that like no agency was ever really going to have all the data that somebody would need for their research. It would be distributed around a lot of other places.

So this is also a bit of a shameless plug but like that's something that we're building at Radiant is trying to build common infrastructure that people can use to share data. So we're in beta right

now but we're building I think called source cooperative which is basically a cloud store that's very generic but that allows people to share data products in a pretty generic way.

Our mentality there is that really what we need is lots of organizations sharing different products but kind of all in the same place and using some of the more common interfaces and I would also say the common interface that we like is HTTP. It's just the web.

When I was at AWS, like I mean like Joel said, we host over 100 petabytes of data in that program. We moved a lot of data and it's just done over the web. It's done over HTTP and in generic like commodity cloud. I think more people should understand this.

So anyway, so yeah, so we tend to focus on this issue that we've invested a ton in producing data but we're not investing enough in making it accessible and available to people in generic ways. So that's what we're trying to tackle.

Joel Gurin: Fantastic.

Dr. Theresa Pardo: Thank you.

Joel Gurin: Yeah, Bob, yeah.

Dr. Theresa Pardo: Bob, yeah.

Dr. Robert Chen: Yeah. Now I do want to say something. I think I would go a little further than Jed. I mean definitely it's important to have data products but users really often want information or knowledge products or the outcomes. I don't think the science community certainly and others are all that good at translating their data into the decision frameworks or the kind of information that decision makers need.

The audience on this webinar is different from my usual audience. You know, decision makers that I'm familiar with are concerned about cost, money, and those sorts of issues. They're concerned about impacts and trustworthiness and liability and all sorts of things that aren't immediately available from data. But if we can translate the data into the appropriate products, that makes it much more useful.

So we did a study over 15 years ago for the World Bank and it had lots of maps. I had lots of good raw data but buried at the end of the report was a summary table of the potential dollar impact by country and the economists, obviously at the bank, loved that table. The maps are great but in the end, they responded to translation into dollars and another example, we do a lot of work with indicators, policy-relevant indicators and the key is to – that you meet a threshold as the risk got up.

The key to the decision, the value the indicator has, how does your country or whatever compare with other countries in the same development level? Decision makers will take action if they rank low relative to their peers and it almost doesn't matter what the absolutes are.

The Scandinavian country – I mean developing countries don't really care if they don't compare well with developed countries. What they care about is how are they doing relative to other countries sort of in their same category.

So thinking about how to translate our data into much more meaningful units of analysis and context I think is one of the keys that we all have to get better at.

Joel Gurin: That's great. So yeah, thank you. So I mean again it's a fantastic discussion and I think Jed and Bob, in some ways this conversation is coming back to Bob where you started us off at the very beginning, which is really the need to begin to see this kind of data gathering and data production and data analysis not as just a scientific research enterprise but as an enterprise that's incredibly and powerfully connected to issues that people are facing on the ground and that decision makers at a city, state, local, national level have to deal with increasingly more and more every year.

So I would like to open a kind of question. We're really talking about – it sounds like what we're really talking about is the need for a kind of continual cycle of dialogue. I mean Bob, you were just – and Jed, you were just talking about the need to kind of provide information in a way that could be – ways that can be used better by decision makers. Bob, you started off by talking about the need for decision makers to actually have input into what data is gathered in the first place because they know or have some inkling of what kinds of data they may need to solve the problems they have.

So it sounds like ideally what we would all like to see is an ongoing dialogue between scientists, decision makers, policy makers, communities that are impacted by climate change, et cetera. This is my turn for a very brief shameless plug which is that the kind of thing CODE has done since we started has been to try to foster those kinds of dialogues in a really structured, facilitated way on all kinds of data issues.

But I wonder in the climate space whether any of you are seeing examples of that kind of dialogue happening or if not, if you have ideas about how it should happen or how we might begin to get that kind of conversation going.

Dr. Theresa Pardo: Go ahead, Cecile.

Dr. Cecile Rousseaux: So again I'm just going to give an example of something that I'm very close to and I know that PACE mission that we have, we have an early adopter program which meant that from the very beginning, we partnered with people that would eventually use those

data and they participate in meeting and we had ongoing data of doing the entire – I think it's for the last five years or something like that prior to the mission.

So that was very beneficial and I think that's a very good example of a success story and that workshop where they put together the scientist and the application people together and it was extremely well-done and I think that has really helped us prepare the path so that the product that we're going to produce were the ones that were needed but also it highlighted some gaps and so we were prepared before launch to tackle those gaps or to see, to even give feedback back to the potential future funding opportunities, to say this is something that isn't done and it's needed.

So yeah, that's my point of view and if I can just add a little bit on the data that we're talking about. I think one thing that we haven't talked about, we've talked a lot about having the people using the data in connection with the data but data accessibility, the reality is a lot of people especially in developing countries cannot download that amount of data, and that's something that we really need to focus on. Everyone needs to be able to have access to those data and know how to look at them because not everyone knows how to open a satellite file. And the more people we have looking at them and the more people we have exploring them, comparing them to their data on the field, the more we are going to understand what's going on. So that was something I wanted to add on the previous conversation. Sorry to have derailed.

Joel Gurin: No, no, absolutely. So it's not just about providing data or even providing insights from data but providing – helping with capacity-building, training, and so on. Yeah, for sure.

Dr. Theresa Pardo: And that also goes all the way back to our beginning, which is, who is living where and under what conditions? And we think about that as health and well-being but we also need to think about it in terms of access to and readiness to use the data, how does accessing this kind of data mean in different parts of the world?

Jed Sundwall: Yeah.

Joel Gurin: Different parts of the US as well. Yeah.

Dr. Theresa Pardo: Oh yes, absolutely. For sure. For sure. Go ahead, Jed.

Jed Sundwall: I hope you don't mind if I chime in on this.

Dr. Theresa Pardo: No, not at all. That's why we are here.

Jed Sundwall: Yeah. So just to piggyback on Cecile's point and also to build I think a little bit on what Bob was saying earlier, like making data accessible to people and visible to people is extremely important. And that's just another, sorry, another plug. But like that's sort of what we do at Radiant also is try to find ways that you can stream data from cloud environments so that

people can like in places of low connectivity, can access just the data that they need or can visualize it in a browser without having to download a massive file. And we've done some pretty interesting work on this, and there's still much more to be done.

But to go back also to Bob's point like why this is important, I agree with Bob that people need answers like they need information products. I like to say like oftentimes the best infographic is a sentence. But who gets to make sort of that decision to interpret the data? We need as diverse an audience as possible that's able to do that. So I'll maintain that we still need to make sort of underlying data products accessible to people so that – I always use the example, I don't know where I got this from, like the Hungarian farmer or somebody who knows the needs of Hungarian farmers can look at the data and make a decision on behalf of that community because a researcher in Chicago would not know the answer that's relevant to them.

So making the underlying data accessible is absolutely paramount to this. And so anyway, just agreeing with Cecile on this.

Dr. Theresa Pardo: Right. So I'm going to go – we've got our panels with hands-up. So thanks for doing that. So Priya then Bob and then we are going to move to our last question after those comments and give Cecile a chance to take a first pass at the question and then after that discussion, we will open up the floor. So Priya then Bob then we will start on our fourth and final question to the panel.

Dr. Priya Donti: Awesome. Yeah. I mean I just wanted to maybe share an example of how we are approaching this at Climate Change AI. So we, at Climate Change AI, run an innovation grants program that gives seed grants to teams of researchers kind of in consortium with "deployment" partners so this maybe industry, this maybe government, these maybe NGOs. And we provide seed funding for these teams to pursue a project that is leveraging machine learning to address some climate change related problem.

And alongside that, we actually require that the teams produce a publicly-available dataset that can then go on to catalyze further work in their area of application. And the reason we did this approach is we wanted to help foster the creation of datasets where they didn't exist, but felt that if you just ask for the creation of datasets, sometimes this means that they're not necessarily as deeply tied to the used cases as they could be. And so by having the researchers who are themselves executing on a project be the ones to create the data that is useful to them, we hope that there is that ability to really have the decision-making in the loop and in the creation of data. So this is one way that we are approaching it at Climate Change AI.

Dr. Theresa Pardo: Very cool.

Joel Gurin: Yeah, Bob, please. Yeah.

Dr. Robert Chen: I just want to go back to actually the point that I was making at the very beginning that there's a difference between how the research community thinks about things and then particularly about accessibility to data versus a decision-making or operation community. So researchers being able to download data at odd times is maybe the norm, but they're not dependent on having data and services available 24 by 7 by 365. And often, scientists or the graduate students finished up in this data that they made available isn't updated in the way that operational dataset would be from a government agency.

So users I think do want trustworthy data sources that are available at a much – with greater reliability and availability than your average science user. And they need someone to talk to, to ask for questions and importantly, they are maybe interested in making sure that they have rights to use the data, so legal interoperability and open access is important as well. And again, researchers tend to kind of skim over some of these things.

So I just wanted to put that out as one of the things that we've certainly been working on. We've been partnering with some groups that work with industry and such and trying to address these broader questions of accessibility and usability and quality and transparency and so forth. So I think that's an important area to think about in moving to the decision-making world.

Dr. Theresa Pardo: Absolutely. And maybe we can ask for a little bit more information for these shameless plug. Priya, maybe you can tell folks where they can go to find it, the seed funding program? And Bob, if there are places where folks can go to, to follow up on your comments. Jed certainly as well.

But let me turn to Cecile because I think again, our conversation is naturally flowing back to a focus on what are the problems? What is the biggest risk that we want to be looking at and how might we identify the data that's necessary? What data and have we been able to determine the data? So what's the biggest risk with the oceans? Every day we read the news, we can see the temperature changes. Certainly, that's one of the major indicators of change. But from your perspective, what do you see is the biggest risk that the oceans are facing and what data do we need to really understand and then manage those risks?

Dr. Cecile Rousseaux: I know we only have 20 minutes left.

Dr. Theresa Pardo: I know.

Dr. Cecile Rousseaux: And I'm guessing you don't want me to talk for 20 minutes. No. So, I think what I'm saying is that's a very complicated problem and I think we are clearly seeing changes now and so I feel like we are finally waking up and realizing what's going to happen.

But yeah, the sea surface temperature increased. It's a bit of a cascade effect, because as the temperature increased, you have different effects. The first one is that you have ice melt and

that means that the sea level rises. Now, that's a whole other cascade of effect where you have coastal communities that are at great risk. And unfortunately, a lot of farmers, well, at least the first one might be very vulnerable communities. But it also affects the weather pattern, right? We are going to have stronger storms, storm surges, and things like that that are going to be a bit more extreme than we are used to.

In the ocean themselves, you're going to have species that cannot really migrate, right? They can only tolerate that much temperature change and then that's as much as they can handle. So we are going to see a change in what species we have. We are going to see a change in the food chain. Corals are a good example. They don't move and they are very, very sensitive to temperature change. So unfortunately, coral bleaching is a thing. And if anyone has had the chance to go to the Great Barrier Reef in Australia before, you will know that things have changed a lot in the last just 10, 15 years.

In terms of going back to the ocean themselves, it's not only the food chain but it's also the type of species. You heard me talk about harmful toxic bloom. Those can happen very often when you have temperature change and condition, the nutrients increase. So the effect is very wide spread and diversified and that links back to the type of approach that we are going to have to take to look and understand not only what's happening but what method we have to mitigate what's happening. And I think having tools that we can rely on that have been validated and that we know are working can help us look into the future as a focus but also do sensitivity analysis.

So one thing that is being talked a lot about is the CDR, which is carbon dioxide removal approach. There is a bunch of different approaches to dealing with that but all of them, the idea is that you do something in the ocean to sink the carbon dioxide from the atmosphere and that would help.

Now to make sure that those methods that we are talking about would be a) efficient, and b) would not have something long-term where if you – one of the things that people have done for a long time, it's not new, but they dump iron in the ocean. Not – as an experiment, not on a routine basis. But they dump like a whole container of iron and you have phytoplankton that grows, takes up carbon dioxide, sink to the deep – very deep ocean. And that's great when you think about it except that we don't know how long they are going to stay very deep in the ocean. And eventually, there's some deep circulation that it comes back and we can't start doing any of those methods without knowing what the long-term repercussion is going to be. And also, which ones are going to be more efficient, right?

So we need to have those tools that are available to us to test, forecast, and see what we are dealing with right now. So – and yeah, again, working with local communities and trying to integrate all the data that we possibly have, but also all the expertise because you have a lot of knowledge in local communities that I think we are still not relying on. People know how the area works like some stuff that you can't measure from space.

And so I think we have a lot of work to do in reaching out with local communities, policy makers, integrating dataset, making sure that the data that we use are good quality like I said earlier. So I think that – I keep saying the word integration but that to me is like that's my focus in terms of where I think we are going to make progress is integrating dataset, integrating people. I want to integrate it all.

Dr. Theresa Pardo: Novel research and collaboration networks that support research translation to practice. That's one of the things I talk about a lot, right?

Dr. Cecile Rousseaux: Yeah.

Dr. Theresa Pardo: Great. So thank you very much for that, Cecile. Did anyone want to jump in on that question or on Cecile's response where it sounds like this – the local knowledge and how do we codify that knowledge. I think one of our speakers in our next panel, Ann Marie Chischilly, is part of a community of Native Americans looking at how to codify traditional ecological knowledge, how do we bring that knowledge to the table in this kind of conversation – in these conversations about climate action. How does that get brought? How does that get used? And so I think that's a critically important aspect of the work that's being done is how do we a) get the global data, the local data, but understanding that sometimes the local data isn't just tabular data, that it has all kinds of nuances and facets that we need to create mechanisms for integration.

Comments from anybody? We have about 14 minutes. So we will take a – we will keep this conversation going but Joel, you want to – if you've looked to the Q&A, if there's a question from one of our participants that we want to bring to the table.

Joel Gurin: Yes, for sure.

Dr. Theresa Pardo: But anybody first who wants to comment?

Joel Gurin: Sorry. Did one of the panelists want to comment? Sorry, I just missed something. We are getting some great questions and let me – and thanks to everyone who is participating for posting them. I'll start with the question from Esther Dyson who is in our audience, "How do you engage – how do you all engage with schools and youth so they can learn how to understand the data and ask good questions?"

Dr. Theresa Pardo: Great question. Great question. Who would like to jump in?

Jed Sundwall: I would just say we haven't done it yet, but I want to more. But that's an explicit goal of ours is to reach younger people. We are constantly learning that younger people are capable of more than we ever thought. And so, a big part of my interest in making large-scale climate data and earth science data accessible to people in browsers is that I want classrooms –

I want classroom experiences where kids can open up a browser and start interacting with data right away without having to install software, field licenses and stuff like that. So, it's in our radar.

Dr. Theresa Pardo: Great. Bob?

Dr. Robert Chen: Yeah. I was just going to say, I mean at the Climate School, there are a lot of educational activities at all levels. Mind Center has a NASA grant involving high school students getting involved in citizen science and collecting data with themselves and processing and contributing to the community databases.

We have a relationship with a Minority Serving Institution here in New York where the students are using machine learning and other techniques to extract relevant data on flood vulnerability and that sort of thing.

So, there are a lot of opportunities. And I think scaling up, it's always an issue. It's relatively straightforward to work with a school or a small group, but how do you reach the bigger audience? We worked with the NASA apply remote-sensing training project, which gets people from around the world and leaves a recording that other people can use afterwards. So there are a lot of different mechanisms, but I think there's no universal solution.

Dr. Theresa Pardo: So maybe – did someone want to jump in on that question? I'd like to make a comment but I'll take back the moderator's prerogative. So that's a critical question and while I can't like others maybe answer it directly in terms of K through 12, one of the new initiatives that's occurring here at the University at Albany is what we are calling the AI Plus through an allocation from our governor. We have a new $200 million AI super computer cluster that is back to the early comment, really seeking to provide the kind of compute capability in the university environment. There's a couple around the country that will have this level of capability.

But as important, we are also looking at – we have created 27 new faculty positions that are being – that 17 of which have been hired and we will bring on those next 10 next year, and all of those folks are being brought in from across the curriculum. So we got folks in public policy and the school of business and social welfare, and they're all focusing on AI Plus and that discipline. And the goal being exactly this, how do we increase exposure to the ideas or the concepts or principles. And we believe that of course includes their attention to the practical realities of data and how do we think about data within the context of AI Plus that they are primarily focused on.

So I think those kinds of initiatives and I think someone said this, no single initiative is going to bring the – is going to take us all to the finish line. It's going to have to be this kind of portfolio of ideally complementary programs that engage people at all levels, and certainly, the K through 12 is key to that.

Joel Gurin: Perfect. Great. Thank you all. We are getting a lot of questions in now so let's just go through a few of these. One is – another one really is kind of about communication. This is from Gavin Starks at Ice Breaker One. Gavin, good to see you. The question here is referring to a recent report on the changes in Arctic ice coverage since 2020. He says, "We are seeing six standard deviations difference. So to a statistician, this is obviously huge. But the question is, how can we communicate this to non-statistics people?"

So in other words, when we are seeing these huge kinds of statistical variations resulting from climate change, how do we get that message across in ways that a non-statistician can understand?

Dr. Theresa Pardo: Cecile has her hand raised. Was that question to one person, Joel?

Joel Gurin: No, to anyone.

Dr. Theresa Pardo: All right. Go ahead, Cecile.

Dr. Cecile Rousseaux: I can try to answer that one. So I think the approach that – and I think Bob was trying to say that at the beginning as well, it's like what the scientific community wants to see in terms of data and what someone who does not have much science background wants to see are two different things. So I think we need to – that's what I was saying at the beginning is we need to be able to translate data and data products at different scale, right? I might want microgram per liter because that's what I work on. But for some end-users, they want red, blue, green. Red means bad. Green means good. Blue means something else, whatever.

But we have to be able to translate a product at different level of complexity so that we can reach out to those different levels of users. And I think how do we explain to a non-scientific person what six standard deviation is, is by switching the words to something that everyone can understand. And we have to understand that there's going to be different levels of complexity in how we present the data and bring that back with the previous question about how do we involve school? A lot of media outreach and having people writing and explaining the science you are doing but in no more general term is I think key to getting everyone involved in understanding because no one wants to read, in the general public, no one wants to read a peer review publication. I mean our parents say they do but they don't. They fall asleep.

Dr. Theresa Pardo: Right. Anyone else want to weigh in on this question?

Dr. Robert Chen: Yeah, just one point. The story I mentioned earlier was looking across lots of different hazards and therefore, lots of different disciplines. And one of the problems we have is language, say, the hydrologists, speak about the 100-year flood. It's a different way of portraying risks than a volcanologist or an earthquake specialist or a crowd manager and so forth. So that gets confusing because everybody has different way of expressing statistics in the different

disciplines. That's very hard to do what Cecile just said, which is translating that into something that is kind of understandable across risks because a manager is trying to figure out, "Well, should I spend money to deal with the crowd or should I spend money to deal with the flood or earthquake proofing or something?"

And if you can't compare the risks because everybody is talking a totally different language, you don't know which one to pay attention to, maybe the loudest or the most recent hazard. So that's always a problem and I don't think – it's hard to bridge the whatever, statistical dialects into something that is really useful for decision-makers.

Dr. Theresa Pardo: Joel? Joel, you're on mute.

Joel Gurin: That was bound to happen once. Thank you. Given that we are a little short on time, we are getting some great questions. I'd like to take the liberty of synthesizing them into two questions. One is the question about some existing kinds of government or other kinds of knowledge-sharing networks, I'm wondering if they're helpful at this point. There was a question earlier about data.gov, which Jed, I think you had some thoughts about.

Also, somebody asking if the open knowledge network being promoted by the National Science Foundation is helpful in the kinds of things we are talking about. Jed, maybe – I know you had some thoughts about the availability of current tools and where we might need to go from there. And then if anyone else wants to comment on that.

Jed Sundwall: Yeah. I mean I'll just say really quickly. Yeah, I just think the concept of data portals is kind of broken. And at the risk of just being too self-promoting, I mean that's precisely what we are trying to build right now with this thing called Source Cooperative. My thesis is that data portals don't really work very well. They don't provide a great customer experience, are expensive for people to maintain. And data.gov hasn't been a viable solution for that.

And so, I would encourage people working in government to think more about what high quality data products they could produce and I would love to talk to anybody about how to get those out to the world. So, I'll start plugging myself now.

Joel Gurin: Thank you. That's good. And let me ask one other question because I know we need to wrap up. A couple of questions just on Cecile, someone pointing out that the pace process you described of how that was developed with a lot of engagement and input was really sounds like a terrific best practice. Just wondering if anyone has other examples of really good examples of engagement or this kind of collaboration with end-users and with people impacted that you'd like to call out as examples to learn from.

Anyone?

Dr. Cecile Rousseaux: I was actually writing an answer that I'm just going to type for the sake of time because I've spoken enough on this one, but …

Joel Gurin: Well, I mean Cecile, go ahead. Yeah.

Dr. Cecile Rousseaux: It's the – so we also have something called the carbon monitoring system program that's a congressionally-mandated program that basically partners every science person that wants funding needs to have a stakeholder group of several stakeholders to get funded. And if they get funded, they work for three years together and the stakeholders have to be from the very beginning, at every meeting. And so it's a total partnership and you cannot get funded unless you have one or more stakeholders.

So that's just a good example I thought because it gives you another approach to solving their problem is if you don't give people money unless they are talking. It also works.

Joel Gurin: Fair enough. Good. Thank you. Good. So I think we are about ready to wrap up. Before we do, maybe if we could just – something we always like to do, I mean I could just go around and ask each of you in turn. If you have one thought, one sentence that you'd like to sort of leave everybody here with and not only as a thought from this but also as a kind of maybe preview of the next session that we are going to have. Jed, do you – I'll put you on the spot. Would you like to start?

Jed Sundwall: Sure. Consider producing better data products. That's it.

Joel Gurin: Great. Very concise. Thank you.

Dr. Theresa Pardo: Perfect.

Joel Gurin: Priya?

Dr. Priya Donti: Don't view machine learning as a solution in and of itself to data gaps but rather, one where we have to co-design how we think about data and algorithms jointly in order to address climate problems.

Joel Gurin: Okay. Great. Cecile?

Dr. Cecile Rousseaux: Talk to a diversified amount of people, not just in your own discipline.

Joel Gurin: Excellent. And Bob?

Dr. Robert Chen: I think the point I'd like to make is investing in data is worthwhile. That is, data – private sector, you would make sure that some of your investment is to get you the data you

need for decision-making. And for some reason, that doesn't always translate to the public sector.

Joel Gurin: Fair enough. Well, listen, thank you all. Fantastic insights, incredible discussion. And Theresa, if you'd like to close out and say a few words about the next one in the series?

Dr. Theresa Pardo: I would. Thank you very much, Joel. Jillian, we have some slides but I don't think it's necessary to pull them up. So we just want to say again thank you to everybody on the panel and thank you to our audience. I just want to foreshadow and remind you if you will of the event, the second panel in the series on September 14th where we will have a group of folks from the local government talking about the local data perspective. And then hopefully, Joel and I will capstone this with some sort of integrated product where we highlight the thoughts that were shared both today and by the panel and by the participants and by our panelists in September and the participants there as well.

So, thank you very much on behalf of Joel, myself, the technology leadership panel, NAPA and frankly, all of the people around the world who are addressing as well as worrying about and experiencing the climate crisis. So thank you very much. Have a good afternoon.

Joel Gurin: Thank you.

Dr. Theresa Pardo: Thank you. Bye.

[End of transcript]

Panelist Biographies

- **Dr. Robert S. Chen, Ph.D, Director, Center for International Earth Science Information Network, Columbia Climate School, Columbia University:** Robert S. Chen is Director of the Center for International Earth Science Information Network at Columbia University's Climate School. He managed the NASA Socioeconomic Data and Applications Center (SEDAC) for more than 25 years until July 2023 and co-leads the Intergovernmental Panel on Climate Change (IPCC) Data Distribution Center. He has worked on climate impact assessment and related data issues since the late 1970s. He received his Ph.D. in geography from the University of North Carolina at Chapel Hill and M.S. and B.S. degrees from the Massachusetts Institute of Technology.
- **Dr. Priya L. Donti, Ph.D, Co-founder and Executive Director, Climate Change AI:** Priya L. Donti is the co-founder and executive director of Climate Change AI, a global nonprofit initiative to catalyze impactful work at the intersection of climate change and machine

learning, and is an incoming assistant professor at MIT EECS. She was previously a U.S. Department of Energy Computational Science Graduate Fellow, a Siebel Scholar, an NSF Graduate Research Fellow, and a Thomas J. Watson Fellow. She received her Ph.D. from the Computer Science Department and the Department of Engineering & Public Policy at Carnegie Mellon University and attended Harvey Mudd College for her undergraduate degree, with a major in computer science and math as well as an emphasis in environmental analysis.

- **Jed Sundwall, Executive Director, Radiant Earth:** Jed Sundwall is the Executive Director of Radiant Earth. Before joining Radiant Earth, he created data sharing and sustainability initiatives at Amazon, including the AWS Open Data Program which hosts over 100 petabytes of open access data. He has helped create data sharing best practices that have been adopted worldwide by NASA, USGS, Google, Microsoft and other institutions around the world. He serves on NatureServe's Board of Directors and has a master's degree in foreign policy from UCSD.
- **Dr. Cecile S. Rousseaux, Ph.D, Research Scientist, NASA:** Cécile S. Rousseaux has been a Research Scientist at NASA Goddard Space Flight Center since 2011. Her research focuses on the role of oceans in the carbon cycle, the effects of climate variability and trends in the ocean biogeochemical cycle, and the support of upcoming field and satellite missions such as the Plankton, Aerosol, cloud and ocean Ecosystems (PACE) mission set to launch in January 2024. She received her PhD in Environmental Engineering from the University of Western Australia, a M.Sc. and B.Sc. degrees from the University of Namur in Belgium and a Masters Degree in Oceanography from the University of Liege (Belgium).

Links Shared in Chat

- https://www.opendataenterprise.org/
- https://napawash.org/working-groups/standing-panels/technology-leadership
- https://radiant.earth
- https://source.coop
- jed@radiant.earth
- https://cloudnativegeo.org
- https://www.esipfed.org/orl#
- https://www.climatechange.ai/blog/2022-04-13-innovation-grants
- https://www.climatechange.ai/newsletter