Good afternoon and thank you for being with us. I'm Lynne Weil, the Director of External Affairs for CSET, the Center for Security and Emerging Technology at Georgetown University. Today, we will walk you through the new features of an exciting CSET project - the Emerging Technology Observatory – which creates high-quality data resources to inform critical decisions on emerging technology issues. But first, a brief bit of housekeeping. All attendees' microphones are muted. If you are on a computer and you experience any technical issues, use the chat function at the bottom of your screen and a CSET team member will try to help you out. And once the presentations get going, please use the chat to pose your questions to the panelists who will try to answer them either in real time or during the Q&A portion. Now, it is time to turn the mic over to my colleague, Zach Arnold, the Analytic Lead for the Emerging Technology Observatory. Zach previously served as a CSET fellow publishing widely cited analysis, including AI safety, global technology, investment flows, and trends in high-skilled immigrations. His writing has been featured in the Wall Street Journal, MIT Technology Review, Foreign Affairs, and leading law reviews. Before joining CSET, Zach was an associate at Latham & Watkins, at judicial clerk for the United States Court of Appeals on the fifth circuit and researcher and producer of documentary films. Zach, over to you.

>> Zachary Arnold: Great, thanks, Lynne and thanks everyone for joining us today. It is exciting to be able share this project, this recently launched project that in some ways has been building the whole time CSET has been around. We're going to talk today about the Emerging Tech Observatory, this is CSET's new platform for data-driven tools, resources, visualizations, datasets on the global emerging tech landscape. As the analytic lead for the project, I'm going to lead our discussion and we are really fortunate to have my co-panelists, Melissa Flagg and John VerWey here to talk about their own, sort of, areas of expertise and how they relate to a few of the tools that we will be describing.

I should say, at the outset, we're going to go fairly quickly, give a high-level overview of the tools that we're able to offer and save hopefully about half of the entire session for Q&A, so please, whenever it occurs to you, please drop questions into the chat. We are collecting them as we go and we're looking forward to get to as many of them as we can in the back half of the presentation.

So, in a nutshell, the Emerging Technology Observatory, we have been publicly launched for a couple of months. We're a new initiative within CSET's fabulous data team and we're building data resources not just for CSET analysts, but for others outside CSET to help inform critical decisions on emerging technology issues. Our thinking, and this has been part of the thinking behind CSET since its inception, is that open-source data, you know data, whether it is from commercial sources, from the public domain, all these data resources are really essential for understanding the global emerging tech landscape. But at the same time making them useful, making them, you know, actually relevant for decision-makers, for policy implementors, it takes a lot of work, it takes a lot of work, a lot of resources. Part of CSET's value proposition from the beginning has been making that investment and using what we can get out of it for our own analysis. We're now hoping to put that investment to work for others. We're inquiring, maintaining, enriching, distributing data visualization, data tools, other resources, infrastructure that empowers others. You know, like all of CSET's work, the Emerging Technology Observatory resources are free to access. We're building them on a nonprofit basis, philanthropically funded like the rest of CSET.

So, we're live right now, I know eto.tech is our website. I know no one would think of multi-tasking during this presentation, but in the event you do have a browser window open, please feel free to go ahead and follow along with the demos or just browse. You can also see after the

presentation, we have support and contact resources on the website. You can contact our team directly. We have live support appointments if you're interested in using this in your day-to-day work, mailing list, a great blog, all sorts of other interesting resources up there.

I'm going to dive right in. So when we launched back in October, we launched with a set of three initial tools, and we're going to go through a few of them in-depth today, again, just to show at a high level what they are capable of and, again, we would really welcoming any feedback or questions whether it is in the Q&A or following up directly with me or another member of our team afterward. I'm going to start with the Map of Science, which is, in some ways, sort of, the largest, most open-ended tool we have. It is one that CSET's data team has been working on for several years at this point, and its roots go back much further than that and we are really proud of it and what we think it can do.

Map of Science is ETO's tool, again based in CSET's data team work, for exploring global research across topics, sources and languages, and we're talking a very broad scope. So, we're covering the entire world. We've got detailed metadata on well over 100 million academic articles. What the Map of Science does is this overwhelming data legible is to apply an algorithmically driven, citation-based clustering structure to provide a legible way to interpret the data. So, what we do is we group all of these articles into clusters, into aggregations based solely on the patterns of citation among papers. We see that that often correlates, not always, but very frequently correlates with things that are more intuitive. Things like shared topic, shared language, institutional similarities between these authors writing these articles. It gives you a way to sort in a rigorous and scalable way, so you can begin drawing insight from this honestly huge and overwhelming and constantly evolving global literature.

We like to use this tool to pinpoint the fastest growing subfields, or sort of, sub-areas within these much broader topics to understand the context around topics that are especially active. You know, see what other clusters are similar or share some sort of citation relationships and other key features. In general, to, you know, use this as a first pass to try to find entry points for further analysis, for further action. There are all sorts of methodologies that people use for trying to interpret the global literature, you know, subject-matter expertise, keyword searching, browsing propriety datasets and so on and so forth, we're hoping to provide a good springboard for that further research with a very efficient, but very powerful user interface. I'm actually going to share a live demo of how the Map of Science works and again, I'm going to go through pretty quickly, but on our website, we have got blog posts demonstrating all of these capabilities and other resources, and again our team is available if anything piques your interest.

The Map of Science, this view we like to call the docs, it is a 2D visualization of all of the 100, 000+ clusters in our dataset. What we do is make these, this sort of, very long list of groupings, filterable and sortable in a number of different ways. So, let's say, for example, I'm interested in exploring the next big trends in computing hardware. So, I can start by fooling around, we have a model-driven approach to defining high-level topics out of these clusters. So, I'm going to look for computer hardware. We can see that the dots are already becoming fewer. I want to look for groupings of research that are relatively larger, maybe a little more active. Let's say a minimum of 50 articles in each cluster. I want to look for clusters that are growing pretty fast. Let's say the top quartile of growth. What else, let's stop there, actually.

We started at about 100,000 clusters, sort of, groupings of research that are densely connected by citations. With a handful of filters, we have narrowed down to about 70. This is one way to Zoom in on research that is of interest to whatever particular question you might be bringing to it. You can look, you know, if you like the dots, you can look at the dots. We have this handy list

view that lets you sort and filter by a variety of different columns. From this view, you can click on any particular cluster to find out a little more about the research that is going on in it. So again, from the bottom up, just from citations, we've built a way to very quickly parse the literature.

Before I get too much into this detailed view, I want to introduce a co-panelist who I think will have probably much more interesting things to say about what is going on in here. We're really glad to have Dr. Melissa Flagg here to discuss the map of science with us. Melissa is a Senior Adviser at CSET, she is a PhD chemist. She has had a diverse and distinguished career in and around D.C. serving as Deputy Assistant Secretary of Defense for Research and other keys roles at DOD, at State, and other organizations. She is also a leading advocate of open-source data as a tool to unlock and accelerate innovation throughout the United States. Needless to say, we're very glad to have her thoughts on how the Map of Science might be able to help with those sorts of efforts. So, Melissa, over to you for a while: How do you see the map being useful in practice?

>> Dr. Flagg: Thanks, Zach. You know that I absolutely love this. I think this is really an amazing opportunity for folks to have access to something like this without having to go behind a pay wall, without having the restrictions that most of us have to just whatever data is at our fingertips. I'm really excited to see CSET doing this. For most of my career, I sat on the decision maker side of this, where I was trying to get folks to use data to give me answers on things like, what's the next thing we should be funding? When has an area plateaued? And maybe we need to be backing off on our funding from basic research, when is something translating to something more applied? When should I b shifting my funding along with that, perhaps to stimulate the next level in a translational pathway for the research? So, I think there are a lot of ways for folks who are policymakers and decision makers who are thinking about programmatics, prioritization of research topics, what types of expertise they should be hiring in, and what level of maturity the research is at can use tools like this to get a perspective on that.

I think maybe with the particular Map of Science tool that folks have at their fingertips right now, there are two ways that I think about using it most frequently. One is where I started, which is what is the next thing? So, perhaps Zach, I think we have a cluster that we talked about, if we want to pull that up, we can look at the -- on the side, the actual filters or statistics that got us there. So, I started searching based on age of the cluster, of the papers that were in the cluster, so I was looking at clusters that had very young papers, very new papers, because I'm really thinking about what's the edge of the science and technology, and what's being done right now. I also looked at things that were growing extremely rapidly, I did not just go to the 75 percentile. I went to 99 percentile.

This is just what is exploding right now that we can begin to understand this is going to be coming at us very quickly. We may not have been as prepared for it. We were not necessarily the creators of it, so I think another thing to look at here is the nations that are the primary drivers of this, so you can see that the growth over time is since 2020. You see India is a big player. China is a big player then you see the U.S. and a smattering of other countries, so a lot of this is not being driven internally. It is being driven externally, so we may not have the kind of visibility into it that we would have if we were, perhaps the primary driver of that cluster.

Again, this is an opportunity to open your aperture to make sure that you understand the things that are coming at you most quickly in the landscape of science that are brand new, and help you make decisions on, okay, how much does this matter? Do I need to be a leader in this? If I

do, I have a very short window of getting engaged. If being a fast follower fine? This might be a way to start to look at partnerships to be able to dig into the actual institutions and actual researchers that are, perhaps making up the largest -- the largest participants in say the Indian contingent here, the 84 articles. We could go down and actually look at the authors' institutions and funders and think about, are there places here that we might want to develop partnerships.

You can also see, interestingly, the University of Arkansas-Little Rock is a participant here at a relatively high level. That is probably not on the radar, as much as a place like MIT or Harvard might be, so it is an opportunity to lean in and make sure we not only do we understand the international landscape, but also do we understand the domestic landscape and are we being inclusive enough that we can leverage our capabilities and we are not being surprised.

I would use this tool, the approach to looking at extreme growth, age of the cluster, and really the makeup of the cluster itself as a way to think about what I'm funding right now, and whether I'm leaning into funding, leaning into partnerships, or perhaps leaning into intelligence, right, depending on what type of organization you sit at.

I think another thing that is really important to look at here, and perhaps Zach you can look at the top articles. So, you can look at key concepts and it will tell you things like IOT, maybe it will say something like "vehicles," but when you look at the top articles and you just read the titles, you can start to get a much richer sense of what is going on in this research cluster. Very quickly we can see it is at the intersection of the internet of things, and 5G. Where did we actually see the vehicles come in? It is not self-driving cars or something like that. It is actually this is ad-hoc vehicular networking architecture and you start to see this 5G VANET that pops up along the way. Some of that has to do with transportation and vehicle-to-vehicle communication, but some of it has to do with a range of other applications as well. What is interesting about citation-based clusters as Zach mentioned, they are not topic-based. They are based on commonalities and relationships between people citing each other.

What we often see in these, they have a similarity of the challenge they are trying to solve, rather than just a really tight science domain or scientific topic. The nice things about citations versus keywords is, I may only think of keywords that are relatively specific to a domain where that approach is applied. When you start to look at citations and references, what it allows you to do is go outside of the keywords that you know, and get into slightly perhaps different jargon that is being used for similar approaches, but in a different domain. This can help inform broadening your keyword search as well, which is very helpful. I think I really enjoy using the citation-based cluster approach that we have in the Map of Science here in combination with other techniques. I think they are very complementary, as long as you can understand the value that you take from each of them.

There is one other approach and one other real value can get from this that I want to highlight quickly, if that is okay, Zach, and that is they just put in a beta section. If you scroll down to the connections section, and you can see that this can begin to allow you to see, not just the growth of the cluster itself, but the actual relationship that it has to other research clusters. What's nice about this, is if I step away from something that is really new, if I look at a cluster that perhaps is even plateauing, something like, deep learning or in this case, sort of reinforcement learning, imitation learning, what is nice about this is, okay it may be plateauing itself, but what that might say to a lot of people is - that is not interesting anymore. We should be putting our money somewhere else. But my question is, just because the growth of the cluster itself is plateauing, what does that tell us about how it is being used in other domains? Growth can happen in a lot of ways.

So, as we look at a cluster that has gotten quite large, that may be plateauing a bit, we want to look at is not only where has it been derived from, so how did it grow from other communities, which is like, who has that cluster been citing? But also, who are the other people in the universe that are using that research to further their own problems, to tackle their own challenges. And that is who is citing back to, who is referencing that primary cluster of interest.

We look at who is actually using that, now you can look at the impact of that research beyond its own little expert network. In fact, often when you talk to scientists, they're not nearly as aware of the domains that are using their work as they are of the work they are using. Right? It is not necessarily the literature they are reading; they are not digging into it, they are simply being cited by it. So, when you're actually pulling together a workshop or thinking about translational research and the pathways of that, or you're thinking about where that science may come back to surprise you through an application, you really need to look at the people using the research, not the researchers themselves or not the cluster itself. Does that make sense?

>> Zachary Arnold: Yeah, no, I am really glad we have got this sort of connections filter. We have a number of other filters at the beta level, but this one particular, I have been fooling a lot over the last few weeks. It is really cool, you have a topic, you know a very high-level topic – reinforcement learning - and you can start seeing in these panes, people are using it for trajectory projection, maybe. Again, you would want to check out these clusters and what see what is going under the hood, a little bit with their own detailed views, but you can see applications emerging out of this citation network. That is really exciting. I think it is a perk of the approach we have taken here.

>> Dr. Flagg: Absolutely. One other thing you all have done is adding in the patents is another aspect of this at the more applied level, as the research is being used, you can start to see who else is using it, as well as what patents are actually citing back to it, and what patents are using it. And I think this can be really powerful. It also helps you to understand how it is being translated and how it is being applied internationally, and this is a great example where you see Google and Alphabet, but you also see a lot of Korean, perhaps Japanese, and a few British and other companies that are leveraging that work as well.

From a policymaker perspective, this is really interesting to begin to understand the effects of your scientific funding. Who is actually benefiting from the funding and who is actually leveraging it.

>> Zachary Arnold: For sure. I know this is a lightning round on the Map of Science. I see questions already coming into the chat about different methodological aspects, we will get to as many of those as we can. I do want, in the meantime, to take a look at another tool that we're very proud of, that is a little newer to CSET, but has launched recently with the ETO platform and it is our Supply Chain Explorer. So the Supply Chain Explorer in general is a visualization and orienting framework around the supply chain for complex technologies, you know, complex supply chains in-depth, covering many stages, covering countries. We're trying to make that more accessible and the first, sort of industry, we have applied that to is advanced computer chips.

As some of you know, CSET has been doing research, producing white papers, policy recommendations and so forth, for quite some time now on advanced semiconductors and our Supply Chain Explorer is a way to turn that into an accessible orienting resource that is, again, providing data-driven information, in some cases directly providing data that we hope can be

useful in actual decisions. We know that semiconductors in particular are, you know, a space that for many of us, myself included has become much more prominent in our work recently. Relatively few true specialists in this field. A lot of people trying to get up to speed very fast and we're hoping this tool will be a resource for them. Again, I will flip through this quickly, and again, this is live on our website like the rest of the resources – chipexplorer.eto.tech in this case. So the Supply Chain Explorer, basically, you can think of it as a flow-through visualization. We've broken down, and again based on published CSET research in every aspect, we have broken down the supply chain for advanced logic chips into a series of stages. These stages involve different inputs.

When we've got it, we have plain English summaries, what is a crystal growing furnace, not something I had heard much about it before I worked on this tool, but now I have a better sense from having spent time with the summaries we have put together. We have information on supplier countries, in this case German is the monopolist, supplier companies involved and we repeat this, again, to the extent data is available at every stage of this chain. Starting from design, going all the way down through assembly, packaging and testing. Compared to the Map of Science, it is a high-level tool with the orienting use case in mind. We have built in the ability to slice and dice the supply chain by a few different characteristics of the stages. You can filter it down and quickly visualize all of the stages of the supply chain for which companies in a particular country they have significant market shares. You can do similar sort of operations for supply companies, you can also glance at relatively less or more concentrated stages in the supply chain. Stages and processes within those stages, I should say. Again, all of this is oriented helping you make sense of what is we know is an immensely supply chain, one that we are oversimplifying for the purposes of tool, but helping people get a footing in this complicated area pretty quick.

I'm glad -- actually, John, why don't we speak with you. I'm glad to have John VerWey to talk through some of the applications and uses of the Supply Chain Explorer. John has been involved with this tool since the early days. John VerWey is a consultant to CSET and a East Asian National Security Advisor at Pacific Northwest National Labs. There he works on issues related to export controls, supply chain security, nonproliferation. He previously served with the U.S. trade representative, the International Trade Commission, the Department of Commerce, in other words, he is a power user for the Supply Chain Explorer tool. We're glad to have him with us today. John, how would you recommend approaching this tool if you're coming to it for the first time?

>> John VerWey: Thank you for the introduction, Zach. I think this tool offers something unique that I am not aware of other resources presenting, which is the ability to understand choke points for particular technology supply chains within the chip ecosystem. I think that it is particularly relevant right now, because a lot of the interest people are expressing in semiconductors is in the context of supply chain choke points or unanticipated vulnerabilities that people were just not aware of before. And some of the filters that you showed earlier, especially related to market share by country, or market share by type of tool or piece of equipment are particularly valuable.

Something I wanted to highlight is something that could be a scenario that is timely and relevant to policymakers right now, which is some controls instituted two months ago by the Bureau of Industry and Security, which called out 10 types of semiconductor manufacturing equipment, and for a policymaker that is not familiar with the semiconductor industry, they may lack the context to understand why these particular tools were chosen, where they are made, what choke points exist, what foreign availability exists, and this tool can answer all of those

questions for lay users who don't need to understand what the tools themselves do and their significance at a technical level.

The example that I thought we could go through is chemical vapor deposition equipment, which, as the name implies deposits chemicals on wafers, to keep it very simple and high-level. But the controls that were announced a few months ago were on a couple of sub segments of deposition tools, and what is nice about them, this Supply Chain Explorer, is it shows not only where the tools fit within the semiconductor manufacturing process and what their value proposition is, but also who are the leading supplier countries, and who are the notable companies that are supplying and you can go through and look at basically what the foreign availability worldwide looks like for a particular product, which is important when understanding potential export control. Interestingly, Zach while you were doing this, I did have another window open and I looked at the Map of Science to see how the things compare. There are 42 research clusters associated with chemical vapor deposition equipment. If you use the Supply Chain tool and correlate it with the Map of Science, you can see interesting trends related to market share and patent filings and which companies own patents and citation intensity, which is a nice example of how the two tools complement and reinforce each other. It gets back to what Melissa was saying earlier, standard tools are useful, but by combining multiple tools, you end up with robust new source of information, and that is the big value proposition of this tool.

>> Zachary Arnold: Yeah, and I think certainly, my own personal experience, my background is nowhere near semiconductors, I'm sure many can I empathize with and this tool, even just working on it as the constant stream of news and new controls and policies around semiconductor have come through. I feel a little better oriented than I used to. And certainly, something like chemical vapor deposition, and I know physical vapor deposition was also a subject of the recent controls. You can get a footing pretty quickly and have an idea of where to look at. We're pretty pleased with these tools and we want to save plenty of time for questions, so I won't get into the third of our initial tool suite in detail, but it is really good, so I want to take a moment to talk about it here.

The Country Activity Tracker is a high level country-level dashboard for different indicators in activity in emerging tech and for this one, we're looking at artificial intelligence. It is a cool and flexible tool. I won't demo it here in the interest of time, but if you go to catETO.com/tech, you can play with it as much as you want. What this tool dies is is to allow country-level comparison of AI research metrics, patenting metrics, and investment activity. You can see collaborations and links between countries in all of these areas, and one of the really nice things about it, compared to a lot of, somewhat similar stuff out there on the market is the scope is worldwide, right. So, I think there is the fair criticism that we often focus a little too much on the U.S. and China and the rest of the world as a distant memory out there. If you're interested in a particular region, a particular set of countries, you know, we have worldwide coverage in our datasets and it comes through in this tool, so you can build your own cohort. You can choose from premixed ones that we include in there and go browsing. See what you can find. It is powerful, especially the cross-border collaboration tools I think a lot of people have been interested in and we have gotten good traction with them to be frank.

I'm going to stop there. I could talk about any of these tools quite a lot longer, but I am going to leave it off there and just remind everyone that all of these tools are live, completely free to access and online right now at ETO.tech. We have demos especially on our blogs. We have walk-through and show off different things you can try doing on your own, integrated with the tools themselves, so you can go back and forth. We have very extensive documentation and a host of support options. We consider our success with this stuff to be whether people are using

it. We are a policy and action-oriented think tank. We're not looking for peer review or any other accolades. We want this tool to be in the hands of users improving decisions and making an impact. Your feedback is more than welcome. That is true both for these resources, which we love to hear positive or negative experience, as people dig into them. It is also true for tools that we have in the works.

We have a variety of topics that we're hoping to release new resources on in the next several months, basically. I included a few on the screen, but we are very, very open to input about where we should be prioritizing and what sort of the resources would need to look like to become part of your work. Several of the existing tools and several of the ones we have under development really come directly from sort of input from users and sort of friends of CSET and others, and we consider that one of our superpowers, hopefully in making useful stuff.

With that, I will close. My contact information is there. It is also on the ETO website, and thanks so much for joining us. We're excited to share this resource with everyone and we have now carved out a good bit of time for Q&A. I have seen the questions pouring in as I am talking, so I am going to catch up for a moment and maybe start, let's see, so I see a bunch of questions in the Q&A about sort of the methodology behind different aspects of the Map of Science. Updating, whether machine learning is involved. I would say Melissa, John, feel free to jump in if I say anything that piques your interest or if I get anything wrong. We will move on from there and more topical questions branching out and have a good conversation.

I will try to show my work as I go to the extent I am able to multitask, so you may have to bear with me a little bit. I see several questions, questions we get about frequency of updating of the Map of Science, which are good questions, because this is a fire hose. Global research is happening quite a lot, and one of the key considerations for the Map of Science, as we developed it is to build a tool that can kind of keep up with that fire hose.

I do want to show off, I will give a quick summary of our general updating methodology. I want to point everyone interested in this question to the documentation of the Map of Science, as well as the datasets that feed it that are on our website. This goes into the nitty gritty details about everything from the algorithms and libraries we're using to shortcomings. We try to be frank about the limitations of these tools, how they are being maintained. We do this for datasets that are not made public, which is true of the Map of Science. The underlying datasets are not made public because they involve a lot of commercial data that we're not licensed to share publicly. We describe what is going on with these datasets and the choices we're making, because we want to build tools that are trustworthy, so you will actually use them. In the event something is not covered in this documentation, this is nerd stuff for us, we are more than happy to talk about this one-on-one.

So, updating, Map of Science is fed by a set of six or seven different very large datasets. Some of these are commercial, so we're drawing in good data, for example, from digital science dimensions and Clarivate, some are public domain, so the archive preprint server, papers with code for A.I. papers and there are a few others in there. Our engineering team has built, sort of infrastructure that has incorporated all of the datasets into the map as an ‘as available’ on basis. In the sense of the underlying data, the is papers going in, those are updated as frequently as the datasets are, in some cases that is continuous, daily, weekly, something like that. There are other datasets, such as Chinese-language only articles are updated less frequently, I believe ever few months. When available, we're pulling all of that data into the tool. The different enrichments and structuring that we do on that are updated on a periodic basis again depending on what they are. For example, the overall clustering structure of the Map of Science, this sort of

groupings of clusters into dots like this, we update periodically. Again, because it is a fairly computationally-intensive change and changes some of the metadata related to the what is displayed through the tool. I will stop there, so I don't make more mistakes than I probably just did. Please look at the documentation for the tools, because we go into great detail into the frequency of updating. It is based on technical infrastructure that we have put in place and ticking away and we, thankfully, don't have to do a lot of manual pulling with most aspects of the map.

One more technical question I see here and there in the Q&A that I want to address. Is this tool driven by machine learning, if not, how do you search and cluster? The overall clustering approach, we use sort of a fairly standard modularity maximizing clustering algorithm, I don't think it would be considered machine learning. We get into that in the documentation, but the clustering is sort of determined by that type of algorithm, which we've, you know, made some sort of decisions in the different parameters that are described in the documentation. Once we have those clusters, we generate a lot of metadata that you saw in the detail view, which I will bring up really quick, so people can see what I'm talking about.

The details about each cluster are generated once we have the clusters in a variety of ways. Some of them are relatively simple operations over the underlying articles that have gone into the cluster, you know, adding up different citation counts and so forth. We do use machine learning methods in a couple of different places to generate this metadata. We have, for example, part of CSET's legacy as an AI specialist organization, we have classifications for how many of the articles in a given cluster have to do with A.I. or A.I. subfields. Those are developed using classifiers, which again, I will refer you to the documentation for the gory details for all of that. Key concepts, this is a very handy field for figuring out when there is a topic in these clusters, what topic might it be. That's, again, a machine-learning approach. For the most part, we're using sort of a fairly standard sort of graph algorithm to figure out the overall structure, and one nice thing about that going back to the updating topic is, we do not need a lot of very manually intensive human interpretation of the structure of the map every time we bring in new data and that is one way we feel this framework is pretty scalable and rigorous.

I'm going to switch to the Supply Chain Explorer. Actually, John, Antonio asks, regarding the Supply Chain Explorer, can choke points be parsed by selecting imposed trade sanctions, such as those levied against Russia? Can you turn on and turn off the sanctions to see impact to supply chains? We don't directly build a sanction layer into the tool, but I think the tool is useful getting at these sorts of questions. Can you talk a little bit about how you use, not necessarily the explorer, but resources like this to figure out the impact of whether it is sanctions, export controls and so on?

>> John VerWey: Yeah, I think that is a very interesting question and relevant potential use case. The way I would frame it is looking at technology-based controls and entity-based controls in the export control world that relate to semiconductors. So, what the tool can show you in terms of technology-based controls is not the export classificational control network of the tool itself. There would be a five-digital alpha numeric combination number associated with each of those pieces of equipment that is shown in Supply Chain Explorer. But the tool does not necessarily show that, that classification is done, in the case of U.S. government, U.S. Department of Commerce Bureau of Industry and Security. Those lists are public, so what you could do is you can find the export control classification number of advanced photolithography equipment and look at the Supply Chain Explorer to see which firms lead in thesupply of that particular type of tool. You can also look at where research is being done in that field on the Map of Science and use the findings of the Map of Science as a way to generate leads and

identify other firms that are "leading" in the development of that technology with publication intensity, science intensity, patent intensity, being used as proxies for leadership.

>> Zachary Arnold: Fair enough. I'm watching the questions continue to flow in. A lot of technical questions related to different aspects of the Map of Science, I'm going to save some of those for if we have spare time, or point people to the publication or to the documentation rather. We have an interest question. I hope I'm pronouncing that right. Can the tools like the Map of Science help with the identification of weak signals? Emerging technologies that are characterized by low volume or low growth rates? Melissa, I will take a pass at it, but it is your territory, so I will turn it over to you fairly quick. The Map of Science is filterable by not just volume, growth rates, although those are ones that I happen to rely on. This filtering pane over to the left has a variety of different filters that you can kind of compose to, you know, in some cases get directly at the subject of your question, in other cases to approximate it. We have other ways besides volume and growth rate to parse. We have subjects, which one thing that is worth stressing, these subjects are not used to generate the map. They are used to interpret it after we generate it based on the citation patterns. It is an inference into what might be going on, and again this is another aspect of the tool that is machine learning based. It is based on methodologies developed by Microsoft that we have adapted for this purpose. But this is one way to parse what is going on in there. Average paper age, country involvement, particularly hot topics, we're looking at A.I. and related fields for a lot of those. You can filter by different aspects of industry and patent involvement.

Again, when you go into the list view, you can use any or all of the filters to parse the results you're getting, including, I mean for example, relevant to your question, you can sort in reverse by a lot of these. These are the slowest growing clusters in the Map of Science. The poor things, slowest in the 110,000 or so. So, it is fairly customizable in a bunch of different directions. Melissa, maybe at a higher level, weak signals, emerging technologies by different metrics, not just raw growth. What do you see as directions to go in?
>> Dr. Flagg: First, hi, I think we know each other if you're my Canadian friend. I think you can absolutely do this. What is interesting is what Zach was getting at, you might use some of the filters we were using, but reverse others. For instance, you may do young clusters, so seeing them when they are first emerging, so do a young paper age, but keep your growth -- nice to see you -- keep your growth rate broad. Don't filter by growth at all, but look at the clusters that are actually the ones that are really coming up now, where the papers are mostly very new. That can tell you a lot about something that is sort of emerging into a white space. The cluster itself may not be brand new, but if it was small and the predominance of papers when it grew suddenly came recently, you would see that pop up in that way.

I think if you're trying to decide where to spend money that is one thing, right, because then you want to know where do I put the money, if you're trying to convene a group of experts to figure out how nervous should I by about this, how excited should I be about this, this is the best way to find the clusters that are most relevant. But if you're trying to find out what topic is out there that is new without filtering or direction based on domain or topic, I think you're going to be focusing on growth rate and the age of the actual papers. The age of most of the papers in the cluster itself, because that will tell you what is new, and how fast it is moving and how much do I need to care about it. Hopefully that answers the question, if not, always happy to chat about that. Also I will just throw out that there was an article published in Plos, the title is: A Novel Approach to Predicting Exceptional Growth in Research. It was co-authored by Dick Clavans, Kevin Wheck and Dewey Murdick, the Director of CSET, a lot of these approaches on the predictions have been tested and published, so that is available if you want to Google that.
>> Zachary Arnold: And that is linked through the documentation in the tool and the Map of

Science, there are a couple of different publications. We don't focus on publications in that sense, but we have a few describing various aspects of them, the clustering methodology prediction, as well as how we use models to subject-classify different aggregations of literature, so definitely recommend those.

I'm seeing questions about how ETO tools are financially supported. ETO is part of CSET's data team. We are funded of CSET's operating budget. CSET is 100% philanthropically funded. Our current donors are on CSET's website. I am not going to list them for fear I will mess it up you, but look there. Basically, CSET as a policy doesn't sort of, we don't take task specific grants from government, for example, we don't take funding from corporations and that sort of thing. So it is the same scenario for CSET in general.

I will say, we've been able to create and put out these tools as ETO. ETO is a smaller organization within CSET, but we are standing on the shoulders of giants here. We are taking advantage, at this point has been a multiyear investment by CSET with the philanthropic support into things like data infrastructure, data licensing and contracting, often fairly constly, very specialized sort of work. Our hope is that we are making all of that work useful beyond the bounds of CSET, and we plan to fund all that work the same way that CSET's analysis is in general, that is the long and short of it.

>> Dr. Flagg: If you wouldn't mind, I see a lot of questions coming in, why one approach rather than another. I want to maybe build on something that John was saying as well. Really kind of emphasize to you that CSET doesn't look at this as a fully automated capital T. truth, you slap some filters up there, you get answers on it and you don't think about it, you just accept automation as the answer. All of these tools should be used with as many kind of tools and as much expertise as you have to begin to guide you down a path to get better knowledge and understanding of an incredibly complex and dynamic space. So, using keyword approaches have huge biases and problems. Citations have biases and problems. Every approach that you would use has biases and problems. The key is to know your biases and to use these approaches in a complementary way, along with expertise.

So, when I use this, I look at the clusters and I dig into those titles of those primary articles, the core articles. I look at who are the folks are and I will sometimes dig into what are the other things those people are publishing, right. I will look at the difference between what the Chinese institutions are doing, what the U.S. institutions are doing, what the German institutions are doing, is there similarity. What you often see is that the similarity is around a problem, not necessarily a topic. If I care about the topic rather than what they are trying to use that thing to solve it, I will go back and supplement this with keywords searching across different clusters. I may use the clusters to find keywords and that I will go back and use other approaches that are more keyword-based. I might use the clusters as a launch pad for that or use a a few keywords as a launchpad to find a few clusters that expand my keyword list.

I think what I would encourage folks to really embrace is there is not a single way to use automation and data that you have available to you as a capital T., truth of the entire world in emerging technology. This is all about beginning to give you more ways to understand this environment.

>> Zachary Arnold: I think that is true of the Map of Science and it is certainly true of the Supply Chain Explorer too. Our highest goal is giving people a footing, like an orientation and good entry points, entry points that you can get at in a rigorous way that put you in a productive

direction for whatever other resources you might have at your fingertips. If you have access to subject matter experts, use this tool to have a productive conversation with them. If you go into a room negotiating something related to one of these technologies, use it to get smart. Use it to tell when someone is coming at you with an agenda and maybe a little more straightforward. If you have access to other data resources, we have, for example, integration directly into the tool, I should have pointed this out earlier if I can find my tab. We have integration directly into the tool into the datasets such as digital science dimensions, into web of science from Clarivate, again, we're pulling some metadata from those vendors into the tool and displaying it here, but you can get more by using it as a portal into those resources. We're always happy to talk with folks about how can I -- these are my resources. This is my situation. How can I plug this in most effectively and that is definitely how we think of it.

We have time for one or two more questions. I might combine a few on the Supply Chain Explorer and ask for John's advice. We have a few questions about the sources of data for the supply chain explorer, how to maintain the sort of understanding of the semiconductor supply chain ecosystem. I think it is interesting to get at this, because in some ways it is the very opposite -- I won't say the very opposite, it is different from bibliometric data where we have excellent public domain sources and excellent vendors and we're able to build infrastructure on top of that. The data that feeds the Supply Chain Explorer has been developed with the blood, sweat and tears of CSET analysts over the past couple of years. Whether it is the textual descriptions, or the listings of notable supply companies, every data point in this tool has been published in a CSET paper with very, very few exceptions. We have done a little bit of enrichment on the company side and a little bit sort of checking to make sure everything is up to date since the papers were published. It is developed from a wide variety of sources, some of those are commercial datasets that, CSET has licensed, again, for our own analytic purpose, but that will have to be, what you see in the tool takes a lot of interpretation and processing from the form that vendors are providing it.

In a lot of other cases, it is open news. Every type of source that you can think of and it does take an analyst to turn it into a format that makes sense. So, this tool, one of the reasons it is at a high level, we are not confident if we could keep it updated and it will be updated if it was that much more granular, and it will be updated on a periodic cycle than the Map of Science, for example, but John, I think it is an interesting question. A lot of people in CSET's orbit are now looking really hard for good, actionable information on semiconductors supply chain. I know there is a lot of interesting policy decisions around that, too. What do you think is going to be needed to or what do you hope to emerge in this information space over the next few years whether it is from ETO or someone else that can better orient U.S. policy on this topic?

>> John VerWey: I think it would be a great outcome if you're able to use the Map of Science to predict some sort of innovation in the semiconductor industry in the same way that certain things has happened in the chip industry that has had some indications beforehand. Currently, to make leading-edge logic chips in particular, we use thin fet architecture that was a result of some innovation funded by DARPA about 15 years ago, initially. And that started being manifested in publications, in patent filings and the signals that we talked about today. So to the extent that the information being collected in the Map of Science and the Supply Chain Explorer could be used to predict something that we will see some out commercially in the next five, 10 years that would be a powerful value.

>> Zachary Arnold: Goals for us in the long run.

>> John VerWey: Yeah.

>> Zachary Arnold: So, we are at time, unfortunately. I know there is a bunch of other questions. A lot of them fairly technically focused. If you asked a question that you're still -- that we didn't get to and you still want to know the answer, please e-mail it to CSET underscore ETO at Georgetown or to me directly. I believe my contact information is available on CSET. We respond to those e-mails, and we will definitely get back to you, and to be honest, questions and feedback about, sort of the -- whether it is the technical aspects or the broader orientation of the tools goes directly into the development process, because it shows this is what people want to know to be able to use this tool in their work. So, please follow up if we were not able to get to your questions. Really appreciate all of the detailed questions about what we're up to. It says to me, we might have an interested user pool for this stuff, which is the dream. That's about the end of our time, so for now, I will turn it back over to Lynne with a big thanks to everyone for coming out and for your interest in what we're doing here.

>> Lynne Weil: Yes, thank you, Zach for being such a capable moderator, but also as the lead analyst of ETO. I think you and your team have a lot to be proud of. I want to thank Melissa Flagg and John VerWey for being panelists and consulting with CSET on a variety of projects. Thanks to everyone else for joining us today. Your comments and questions were really engaging. We're sorry we couldn't get to them all. If you're interested in following up directly, send an e-mail to CSET underscore ETO at Georgetown.edu. If you would like to learn more about CSET, please go to cset.georgetown.edu, and sign up for our newsletter and research updates. And keep an eye out about news on our next Tech and Security event taking place in January. Until then, here is wishing you a happy and healthy holiday season. Thanks all.