Opening Statement
Research Ranking Member Daniel Lipinski
Subcommittee on Research
Committee on Science, Space, and Technology
Next Generation Computing and Big Data Analytics

April 24, 2013

Thank you, Chairmen Bucshon and Massie for holding this hearing on examining the next generation of computing and big data analytics. I want to welcome and thank the witnesses for being here today.

Today's hearing gives us an opportunity to talk about the new tools and analytics that are being developed for big data. Big data can be thought of as large volumes of complex and diverse types of data that are also high velocity – meaning they change rapidly with time.

As a member of the Research Subcommittee for several years now, I have watched as the amount and complexity of data has grown by leaps and bounds. The field of astronomy is a great example. When the Sloan Digital Sky Survey started work in 2000, its telescope in New Mexico collected more data in a few weeks than had been collected in the history of astronomy. And that telescope will be surpassed when the Large Synoptic Survey Telescope goes online in about 2020. LSST will photograph the entire sky every few days. That's difficult for any of us to wrap our heads around.

The types of data are changing as well. Data has gone from being mostly numbers entered in excel spreadsheets to data coming from sensors, cellphone cameras, and millions of email messages. In fact, it is estimated that over 85 percent of data generated today are these kinds of unstructured data—data like videos or emails.

The change in the volume and variety of data as well as how fast data is being produced and changed creates almost limitless opportunities. For example, since cybersecurity data is massive, varied, and changing quickly, big data technologies have the potential to detect and prevent cyber attacks before they even happen. I know that organizations like IBM are developing technologies to do just that. Additionally, big data could be used to establish new business models, create transparency, improve decision-making, and reduce inefficiencies within businesses and government.

But along with the opportunities, there are a number of challenges. We need new tools and software packages to manage, organize, and analyze all these different kinds of data. Additionally, we will need an analytic workforce to ensure the gains of big data. These challenges necessitate involvement from government, academia, and the private sector. That is why I am happy to see all those sectors represented today.

The government has and will continue to play an instrumental role in this area. For instance, the Networking and Information Technology Research and Development -- or NITRD -- program created an interagency big data group that is coordinating federal efforts in technologies, research, competitions, and workforce development for big data.

In some cases, agencies have teamed up to issue joint solicitations. For example, NSF and NIH have a joint big data grant program that awarded nearly $15 million of grants to eight teams of researchers last year. These first awarded grants went to projects focused on designing new tools for big data and new data analytic approaches. We will hear more about these and other interagency activities from Dr. Jahanian in his testimony. We will also learn more about specific programs at NSF, one of the leading agencies in federal big data efforts on both the analytics side and the computational resources side.

As I mentioned before, one of the areas being coordinated through NITRD is the workforce development needs for big data. Several agencies, including NSF, have education activities to support a new generation of big data researchers. As you will likely hear from all of the witnesses, we face a looming shortage of workers with the skills needed to analyze and manage large, complex, and high-velocity data sets. There is some overlap with the broader STEM skills we often speak of in this committee. But there are also some unique skills required to address the challenges of big data. We need to consider how to build those skills into STEM curricula, especially at the undergraduate and graduate levels. I look forward to hearing from our witnesses about the current educational efforts and what additional initiatives may be necessary.

Finally, since big data involves different types of data that can be produced and transferred quickly, there are concerns over privacy. We need to ensure that we strike the right balance between exploring and implementing all of the potential benefits of big data while also protecting individuals' personal information.

I look forward to hearing the witnesses' testimonies and to our discussion today.