Institute for ADVANCED ANALYTICS

Dr. Michael Rappa, Founding Director

Testimony of

Michael Rappa, Ph.D.
Executive Director and Distinguished University Professor
Institute for Advanced Analytics
North Carolina State University

Before the

Committee on Science, Space, and Technology
Subcommittee on Technology
And the
Subcommittee on Research
U.S. House of Representatives

April 24, 2013

Next Generation Computing and Big Data Analytics

Good morning, Chairman Bucshon, Chairman Massie, Ranking Member Lipinski, Ranking Member Wilson, and members of the subcommittees. Thank you for the opportunity to speak with you about big data, analytics, and the important role institutions of higher education play in advancing the field.

My name is Michael Rappa and I am the Executive Director of the Institute for Advanced Analytics and Distinguished University Professor in the Department of Computer Science at North Carolina State University. I am responsible for overseeing the nation's first graduate degree in Analytics, which was founded in 2007 and currently enrolls 80 students each year. I also co-direct a government sponsored research project that seeks to advance the science of security in cyberspace.

My testimony today draws on my 25 years of experience in graduate education and research and, in particular, my experience over the past seven years leading the Master of Science in Analytics degree program at NC State. I will focus on the role different academic disciplines play in the field of Big Data Analytics, the success of the Institute for Advanced Analytics in partnering with industry to produce analytics professionals and lessons for other universities, and conclude with recommendations for policymakers in their support of Big Data and developing the analytics capabilities of the workforce.

What is the role of different academic disciplines in Big Data and Analytics?

Many of the things we do each day as individuals and organizations generate data. How much data? No sooner would I give you an estimate than that number would be surpassed. The global use of connected digital devices—computers, smart phones, tablets, and the like, that propagate numerical, voice, image and text data—is growing just that fast. With all of this data come both necessities and opportunities, which have given rise to what today is called "Big Data" and the concomitant need for Analytics; namely, the tools, methods and applications for drawing insights from large quantities and varieties of data. Big Data is a relative term that suggests a realm of data that pushes the upper limits of existing capabilities in computation, storage and analysis. In this sense there always has been and always will be Big Data.

What's interesting about Big Data today is less about its inherent bigness then its ubiquity. A case can be made that Big Data is both a systemic and transformative technological phenomenon—the latest phase in a decades long process that began with the rise of modern digital computing in the 1940s. The computerization of business processes and later personal computing had the effect of creating first large and then highly dispersed stores of data. Subsequently, the rise of the Internet and World Wide Web facilitated the movement of data, as well as created huge amounts of data through its use—dramatically increasing the amount, immediacy, and the interconnectivity of data flowing from an increasing number of users and digital devices that now extends far beyond computing in the conventional sense. Thus, the attention to Big Data and Analytics should come as no surprise. It is part of a natural progression that has grown out of the digital world we live in today. The ability to analyze and draw insights from data is without question important. In the opinion of some observers, it has the potential to become a defining factor in how well organizations perform their missions in the future.

Looking at the academic landscape, certain disciplines play a critical role in pushing forward the research frontier: work in computer science, statistics, operations research, and applied mathematics is particularly relevant. But it doesn't stop there. Just as Big Data and Analytics will touch almost every corner of the economy, so too will it draw in researchers from a wide spectrum of disciplines and sub-disciplines as the Internet did a generation earlier. I encourage the Committee to embrace an expansive perspective in its support of Big Data research in the academic community. When the hype around Big Data eventually dies down, we will nonetheless find ourselves in a world saturated with data. Our ability to draw value from this data will be as important as it ever was to every sector of the economy. Only through sustained investments in the academic community will we ensure the technology and workforce necessary to leverage data to its fullest extent.

Research is needed to continue to advance our understanding of how to collect, store and process extremely large amounts of data, where the definition of "large" is a rapidly expanding quantity that knows no bounds. Many areas of computation and data analytics are relevant, including high performance computing, databases, networking, machine learning, data mining, algorithms, data visualization, natural language processing, optimization, geospatial analytics, remote sensors, data privacy and security, among others. Again, I suggest taking a broad view. It's important to recognize that the academic landscape is fluid with new disciplines and sub-disciplines emerging over time. Under the right conditions these new frontiers of knowledge will become woven into the institutional framework of universities through the establishment of centers and institutes.

How academic disciplines combine and split would matter little were it not for the organizational boundaries that are created or dissolved in the process. The domain of Big Data and Analytics is not unusual in how it spans disciplines, but we need not be so concerned about crossing disciplinary

boundaries with respect to research. Many incentives are in place to overcome the barriers, and the current generation of faculty has grown up in an environment where multidisciplinary work is not uncommon, if not the norm. At NC State, for example, we have a large multidisciplinary project on the analytics needed to advance a science of security. There are natural hurdles researchers will encounter working across disciplines, certainly, but it is something that can be overcome and is overcome on a regular basis.

More attention should be paid to how the interplay of disciplines can facilitate or inhibit the development of educational programs that fulfill the needs of employers and an evolving workforce. Academic disciplines and the related degree majors don't always align well with occupational roles. There is a need to pull together disparate disciplines in graduate education to produce the kind of technical professionals employers want, such as data scientists. This can be difficult because it goes beyond faculty collaboration to the larger challenge of reorganizing how universities design and deliver degree programs. To overcome the projected scarcity of data scientists, universities need to reduce or eliminate organizational boundaries so that students can acquire the bundle of skills they will need to succeed in the workplace.

To address this challenge, North Carolina State University partnered with SAS in 2006 to create the Institute for Advanced Analytics and embark on a grand experiment in designing a new graduate degree in Analytics, the first of its kind. By every measure, the experiment has been an unambiguous success. The program's enrollment has grown to 80 masters students annually, with over 95-percent self-financed—allowing the program to be self-sustainable based on tuition revenue. As a group they are among the most sought after and highest paid graduates of the university. Students come from a variety of disciplines and range in age from their early twenties to their late fifties. The majority of students return to school from the workplace, and 80- to 90-percent are U.S. citizens or permanent residents. This year the Institute will record its sixth consecutive year in which 90-percent or more of the graduates are placed by graduation in positions as analytics professionals—despite the worst economy since the Great Depression. Key to this success is close collaboration with industry combined with an innovative way of organizing graduate education.

The Institute for Advanced Analytics

When the Institute for Advanced Analytics was proposed in April 2006, it was conceived as an educational initiative intended to address the growing demand for professionals who can draw insights from (what we now call) Big Data. At the cornerstone of the proposed Institute was the idea of creating the nation's first Master of Science in Analytics (MSA) degree. The proposal was unusual in that it was adopted as a greenfield project that allowed us to start with first principles when creating the degree program. Furthermore, the Institute was launched as a university-wide collaboration—organizationally independent of the colleges—to give it maximum flexibility to blend together whatever disciplines were needed in the education without being anchored to a single department or college by default.

The MSA curriculum was designed by a large group of faculty from more than a dozen academic departments across six colleges, in close consultation with colleagues at SAS and a large industry council. We shed conventional thinking about degree programs as much as we possibly could. Instead of starting with a menu of core courses and electives—the basic inputs—the design process began with a clear focus on desired students outcomes, positioning employers as the core customer and seeking to understand what they look for when hiring analytics talent. Analyzing job descriptions and talking with employers led to a balanced perspective embodying five core objectives for student outcomes that

guided the MSA:

- Technical skills
- Teamwork skills
- Communication skills
- Tool skills
- Domain knowledge

When it comes to analytics professionals, employers want to hire individuals with the requisite technical skills to clean, analyze and interpret data, but it doesn't end there. Employers also want people who can perform well in multi-functional teams, who have strong verbal and written communication skills, who have experience programming with industry standard software tools and programming languages, and who have prior knowledge of the business domain whenever possible.

The MSA curriculum was laid out with all of these objectives in mind, not as a series of courses, but as a single integrated learning experience. Faculty members are engaged not by the course, but instead in customized lecture streams that align closely with their expertise. This allows us to use a larger number of faculty and to integrate subject matter across different disciplines with greater flexibility. The result is quite different from conventional programs. The MSA is optimized to achieve the five core objectives by leveraging both the content and programmatic structure of the learning experience. While the focus on employers and the inclusion of professional skills development aligns the MSA with the larger Professional Science Masters (PSM) degree movement, it takes the idea a step further than most, given the Institute's flexibility in crafting the curriculum.

The MSA is a cohort-based 30-credit hour curriculum spanning ten months. It uses the intensity of a condensed format to immerse students in the learning process. MSA students pursue the degree full-time—literally. The program calendar runs 9:00 to 5:00, five days a week for most of the ten months (vacation time is more limited than the normal student calendar). When students aren't in class, they are typically working on a long-term team project (the practicum) or in study groups that rotate membership every five weeks. A premium is placed on learning by doing. The result is a highly structured team-based learning experience that has more of the look and feel of a normal workplace.

The MSA places heavy emphasis on "soft" skills. Professional skill training is woven into the learning experience throughout the curriculum. Students receive training in public speaking, technical writing, teamwork, leadership, project management, and conflict resolution, among other areas. The soft skills are not treated as non-credit add-ons or extra-curricula activities. Instead, these skills are taught as an integral component of the curriculum and students are measured and tested against them.

Teamwork is extremely important in most organizations today. Teams form the basis for most of the work students perform in the MSA program. The cohort driven format enables the students to work in teams continually across the entire learning experience. The team structure of the curriculum and diverse student body maximizes peer-to-peer learning, which is known to be very effective and perhaps as important as direct instruction. Students are given guidance on teamwork and undergo frequent iterations of peer and self-evaluation and feedback on their performance. The peer evaluation is both structured and open-ended and forms the basis for personalized coaching to help students improve their ability to function effectively in teams.

At the core of the MSA curriculum is a practicum—a hands-on learning experience that gives students

the opportunity to conduct real-world analytics projects using data from sponsoring organizations. Students work in teams of 4-5 members each to understand an actual business problem and then clean and analyze the data. The Practicum spans seven months and culminates with a report and presentation to the sponsor. The teams perform their work under confidentiality agreements and the results are the sole property of the respective sponsors. Past projects span virtually every industry segment—advertising, banking and financial services, consulting, e-commerce, energy, entertainment, healthcare, insurance, retail, and transportation—and a number of state and federal agencies. The Institute currently conducts 17 practicum projects each year. This year's sponsors included:

- Allscripts (Chicago, IL)
- Caterpillar (Peoria, IL)
- Central Intelligence Agency (McLean, VA)
- Federal Communications Commission (Washington, DC)
- GE Energy (Atlanta, GA)
- Global Knowledge (Cary, NC)
- GlaxoSmithKline (Research Triangle Park, NC)
- Hanesbrands (Winston-Salem, NC)
- Houston Astros (Houston, TX)
- Inmar (Winston-Salem, NC)
- Lowe's Home Improvement (Charlotte, NC)
- M&T Bank (Buffalo, NY)
- Monsanto (St. Louis, MO)
- North Carolina Department of State Treasurer (Raleigh, NC)
- Procter & Gamble (Cincinnati, OH)
- U.S. Postal Service (Washington, DC)

The Institute's open solicitation of practicum proposals keeps the curriculum grounded in the kind of problems facing industry, which continue to evolve in character and complexity. Dozens of proposals are received each year from an equally large number of organizations representing almost every industry segment. The proposals inform decisions made about the kinds of methods and tools we choose to emphasize in the curriculum, given the limits of time in a 10-month program cycle.

In both the practicum and class instruction, students learn analytics by using industry standard software tools. The Institute's collaboration with SAS is particularly important in this regard. Enterprise class analytics tools are a challenge to deploy and maintain in university settings, given the lack of standardized platforms and operating systems. Through the generous ongoing technical support of SAS, not only do students have the opportunity to use an industry leading tool set, they can (and frequently do) complete numerous SAS certifications in route to their degree. The significant market value of such industry certifications is well documented and further underscores the readiness of MSA graduates when they land on the job.

The MSA program was designed from top to bottom with the intention of producing the kind of analytics talent employers seek to hire. Since the beginning, the focus remained on how to achieve successful student outcomes by understanding employer needs. It should be no surprise that no sooner did we start to produce graduates from the program than those graduates attracted the strong interest of employers. The Institute has achieved over 90-percent job placement by graduation each year since the first MSA cohort entered in 2007. The current class of 81 students, which will graduate in a few weeks, will enter new positions with over 30 different employers in 12 states and the District of

Columbia—at record high salaries. Nearly three-quarters of the class had two or more job offers, and over half had three or more offers of employment. Five-year benchmark studies put MSA student outcomes, in terms of placement and salary, on par with some of the country's most prestigious universities.

Six years later there are now at least two-dozen graduate degree programs in the U.S. that focus explicitly on Big Data and Analytics, and new ones are announced with regular frequency. Some programs, like one at Louisiana State University, have sought to closely replicate our model, while other programs, like one at Northwestern, have adopted similar components like the practicum and industry standard tools. It's common for schools to work with one or more supportive companies like SAS, IBM and others, to help them deploy industry standard tools or share business problems and data with their students. While many programs are situated in business schools, it remains to be seen whether organizationally this is the best approach in balancing the disciplinary knowledge needed to produce the kind of data scientists required by industry.

Conclusion

The value of big data lies in our ability to extract insights and make better decisions. The acute shortage of analytics professionals and data-savvy managers will be addressed most successfully through creative partnerships between industry, government and universities. There are several efforts underway, but we must intensify and accelerate the national investment in proven models.

I had the pleasure of serving as academic co-chair on the TechAmerica Foundation's Big Data Commission, along with many representatives of the Foundation's member companies. Among the Commission's findings was a recommendation to continue to invest in research and development of advanced computing technologies that can effectively process not only the vast amounts of data being continually generated but also the various types. Those investments should focus on key government priorities such as education, fraud and abuse, cybersecurity, healthcare and public safety.

The Commission also recommended a strong focus on skill development in the workforce. Public-private partnerships should be strengthened and expanded to invest in skills-building initiatives for the workforce in the area of Big Data. While many of the Commission's recommendations were directed specifically to the federal government's own needs, it also encouraged the development of data-intensive degree programs and scholarships to prepare a new generation of data scientists.

The Institute for Advanced Analytics has a proven track record over the past six years with the Master of Science in Analytics degree program that shows we can succeed in educating a new generation of data savvy professionals to satisfy the needs of employers. We can do it quickly with an intensive and highly targeted educational format that yields consistent student outcomes. The program attracts a diverse, high quality, domestic student population and yet runs with a sustainable, self-financed business model. Professional Science Masters programs like ours show that graduate education can be designed effectively in collaboration with employers to fill the skills gap in critically important areas like data science and cybersecurity.

Thank you for the opportunity to appear before you today to provide this testimony. I welcome your questions.

Biographical Sketch

MICHAEL RAPPA

Michael Rappa is Executive Director of the Institute for Advanced Analytics and a member of the faculty in the Department of Computer Science at North Carolina State University. As head of the Institute, he leads the nation's first and preeminent Master of Science in Analytics as its founder and principal architect. Before joining NC State as Distinguished University Professor in 1998, for nine years he was a professor at the Massachusetts Institute of Technology.

Dr. Rappa has 25 years of experience as a professor working across academic disciplines at the intersection of management and computing. An accomplished researcher and instructor, his passion is to bring an entrepreneurial and forward-thinking mindset to innovation in higher learning. His current role is to prepare a new generation of data savvy professionals for leadership in a digital world.

In addition to his duties as director, Rappa is co-principal investigator of the Science of Security Lablet, a large multidisciplinary research project sponsored by the U.S. National Security Agency. The project is run in parallel with sister labs at Carnegie Mellon University and the University of Illinois at Urbana-Champaign.

A study published in 2006 in the British journal R&D Management identified Rappa as a leading scholar in the field of technology management, ranking him in the 99th percentile among over 9,000 authors in terms of research productivity in top journals over the past 50 years. His research has been cited on three occasions as an outstanding contribution to the field, and his early work on business models is one of the most often cited and widely read publications on the subject.

Rappa is perhaps best known as the creator of *Managing the Digital Enterprise*, an innovative and award-winning educational Web site devoted to the study of management in the digital world. Launched in 1999, originally as the foundation for a course he taught, the site is a valued resource used by several million learners and hundreds of university instructors from around the world. In 2010, Rappa presided as general co-chair at the *19th International World Wide Web Conference*, the premiere annual gathering of the Web research community.

NC State has recognized Rappa on several occasions for his contributions to teaching and service. He is the recipient of the Outstanding Extension Service Award, the Award for Graduate Teaching Excellence, and the Gertrude Cox Award for Innovative Excellence in Teaching and Learning with Technology. He is also winner of the MERLOT Award for Exemplary Online Learning; a four-time recipient of the IBM Faculty Award; and twice a finalist for the Epton Prize.

Rappa began his teaching career at the University of Minnesota, where he earned his doctorate in 1987.

May 23, 2013

Hon. Derek Kilmer (D-WA)
Congress of the United States
House of Representatives
Committee on Science, Space, and Technology
2321 Rayburn House Office Building
Washington, DC 20515-6301

Dear Rep. Kilmer:

I am writing in reply to your letter following the Committee's April 24th hearing on Next Generation Computing and Big Data Analytics. Thank you for the kind invitation to testify at the hearing and this opportunity to address your questions.

1. *How is your organization specifically addressing the need for more data scientists and employees with STEM backgrounds?*

The Institute for Advanced Analytics at North Carolina State University has been at the forefront of educating a new generation of data savvy professionals to address precisely this need for data scientists. Launched in 2007, our Master of Science in Analytics (MSA) degree program is the first of its kind in the nation and has become a template for programs at other universities. The MSA blends together statistics, mathematics, computer science and business topics into an interdisciplinary curriculum and is classified as a STEM graduate degree. The Institute works closely with employers, professional societies, government agencies, and academic institutions to enhance the pipeline of data science professionals.

As I mentioned in my testimony, the Institute has a proven six-year track record that shows we can succeed in producing the kind of talent employers need. We can do it quickly with an intensive and highly targeted educational format that yields consistent student outcomes. The program attracts a diverse, high quality, domestic student population and yet runs with a sustainable, cost-effective, and self-financed tuition business model. We have demonstrated the ability to produce 80 graduates annually, and I am confident we could produce tenfold with the necessary upfront investment in facilities and personnel.

2. *In your testimony, you discuss how our nation is facing a data science shortage. What policies would you recommend Congress consider to address that shortage?*

The Institute has shown this is a problem we can solve through close collaboration with the private sector. The task now is to scale quickly to meet demand. I believe markets will eventually adjust, but we could accelerate the process with a better flow of data.

<u>Recommendation 1</u>: Provide guidance and incentives to degree granting institutions to make public standardized reports of student employment outcomes, such as job placement and salary data, by degree majors.

Accurate and up-to-date information is key to achieving market equilibrium. Prospective students who are contemplating the time-consuming and financially burdensome decision to invest in their education must have the data needed to calculate the expected return on investment.

There is nothing particularly novel about this recommendation. Already many of the nation's top business schools make public each year comprehensive employment reports for their MBA programs (as does UW's Foster School of Business). This data is absolutely essential to prospective students in making their decision to pursue the MBA degree (typically two years of study, and an average tuition cost of $80,000 at the best schools). Before making a decision about whether or where to attend, individuals can calculate an expected ROI based on recent placement rates and salaries.

Similar employment reports should be produced for STEM graduate programs. Once we have timely and accurate employment data, the marketplace will adjust quickly to the opportunity in data science. We have seen exactly this kind of reaction in the marketplace for our Master of Science in Analytics degree. The Institute has published detailed employment reports since the inception of the program. Our applicants clearly make decisions based on data driven ROI calculations of expected employment outcomes.

We know from employers the demand for data science talent is there and growing. We also know there are large numbers of students coming out of school or already in the workforce who are underemployed, if not unemployed. There is simply poor alignment between skill training and skill needs. The key to achieving a closer alignment is up-to-date and accurate degree outcomes data that would allow students to make sound educational investments. STEM employment reports need to become as commonplace in universities as they are today with MBA programs.

Recommendation 2: Provide incentives for college and university students to pursue fields of study in data science contingent on institutional performance metrics for student outcomes that meet threshold levels.

Congress may wish to create or target existing incentives for students to pursue STEM degrees. I recommend tying incentives to institutional performance metrics such as graduation rates and employment outcomes. This will ensure students are making sound investment decisions in furthering their education based on expected outcomes.

The current system of subsidized student loans would be on a more solid footing financially if students taking loans knew in advance the likelihood of their ability to repay loans. Perhaps such data could be included in the loan application process. Again, it comes down to more timely and accurate public data about educational outcomes so the marketplace can make better decisions.

Recommendation 3: Provide guidance to Federal agencies to seek-out and take advantage of educational opportunities for existing employees to pursue fields of study in data science. Furthermore, encourage the creation of job categories within the Federal workforce to employ and promote data scientists.

Existing employee training programs within government agencies should be encouraged to take advantage of data science related educational programs and to use institutional performance metrics to determine what programs qualify. Sometimes existing program classifications within agencies can be slow to change and fail to include in a timely manner new areas of studies like Analytics.

I have worked with one Federal agency to help define a new job category of "data scientist." Congress may wish to encourage other agencies to move in a similar direction. By providing leadership, the Federal government will also help the private sector as it, too, seeks to define job categories for data science.

Clearly, the educational community needs to accelerate its efforts to increase the supply of talent if there is any hope of meeting the demand over the next few years. Working with employers and government, I am confident that universities can and will step-up to the challenge.

Once again, thank you for this opportunity. Please do not hesitate to contact me as needed. If I can be of any further assistance to the Committee or to the citizens and employers of Washington State, I will be glad to help.

Sincerely,

Michael Rappa, Ph.D.
Executive Director of the Institute for Advanced Analytics
and Distinguished University Professor
North Carolina State University
920 Main Campus Drive, Suite 530
Raleigh, NC 27606