

Testimony of

Paul Misener Vice President for Global Public Policy, Amazon.com

Before the

United States House of Representatives Committee on Energy and Commerce Subcommittee on Communications and Technology Subcommittee on Health

Hearing on

21st Century Technology for 21st Century Cures

July 17, 2014

Good morning, Chairman Walden and Chairman Pitts, Ranking Members Eshoo and Pallone, and Subcommittee Members, my name is Paul Misener, and I am Amazon.com's Vice President for Global Public Policy. On behalf of Amazon and our customers, thank you for inviting me to testify today on how 21st Century technology enables 21st Century cures. We are grateful to you, and to Chairman Upton and Representative DeGette, for exploring this important topic.

After briefly describing cloud computing technology, my testimony will illustrate how innovative healthcare organizations – large and small, established and start-up, public and private – already use cloud computing to foster the innovation cycle of discovery, development, and delivery of new biomedical treatments and cures. I will conclude by suggesting three ways that Congress could help accelerate this cycle by adopting policies to facilitate use of cloud computing for healthcare.

Amazon.com opened for business on the World Wide Web in July 1995 and seeks to be Earth's most customer-centric company. After over a decade of building and running web applications and

databases for our retail business, we realized that we had developed a core competency in operating massive scale computing technology infrastructure and datacenters. So, we set out to serve a new set of customers – including medical research institutions and healthcare providers – with a cloud computing business, which we named "Amazon Web Services."

With cloud computing, information technology (IT) users, including healthcare providers, research institutions, start-ups, and other enterprises, now can acquire technology resources such as compute power and storage on an as-needed basis, instead of buying, owning, and maintaining their own datacenters or servers. This is known as the "utility" model of obtaining and using IT capability, analogous with how enterprises obtain water, gas, or electrical power. Users only pay for what they use – by the compute-hour or storage-gigabyte – and they are not locked into long-term contracts. If a program is funded one year and then unfunded the next, or a pilot project or initial testing phase does not achieve the expected results, organizations no longer need to be tied to large, capital IT expenditures that can cost hundreds of thousands or even millions of dollars. As a result, healthcare start-ups, medical research institutions, hospitals, and clinical providers of all sizes have more agility, enabling biomedical innovations for 21st Century cures. In sum, cloud computing offers IT users, including healthcare enterprises, many benefits:

- First, with cloud, IT users can trade capital expenditures for variable expenses. That is, users can pay only for what IT they actually consume, and only when they consume it.
- Second, with cloud, those variable expenses are lower than they would be if the user selfprovided IT services. With inherent economies of scale, the large-scale commercial cloud is simply more efficient than anything a particular user could build and operate for itself.
- Third, users don't need to guess their capacity needs. Before cloud, users risked the waste of buying too much IT capacity if demand were lower than guessed, or they risked dissatisfaction of their customers or citizens with shortages, if the users bought insufficient IT capacity to meet demand.

- Fourth, the speed and agility of user innovation is dramatically increased with cloud. Instead of waiting many weeks to obtain IT infrastructure, virtually unlimited capacity is available to users within minutes.
- Fifth, cloud computing allows a user's scarce technical talent to focus on its core mission, not on maintaining basic compute and storage infrastructure to support it. With the budget challenges that organizations face today, that focus is valuable now more than ever.

In 2006, Amazon's cloud computing business, Amazon Web Services, or "AWS," began offering

developer customers access to in-the-cloud infrastructure services. AWS now has hundreds of

thousands of customers, including over 3,000 academic institutions and 800 government agencies.

In the healthcare sector, enterprises of all sizes and types are beginning to use cloud computing

technology for the *discovery* of new biomedical treatments and cures. Here are a few examples from

our AWS experience:

- Spiral Genetics, a Seattle, Washington-based bioinformatics company formed in 2009, makes high-performance software that helps researchers analyze DNA in the cloud. Before Spiral Genetics was founded, the cost of sequencing a human genome was \$100,000, the chemical process of sequencing took 30 days, and computational processing in traditional hardware-based infrastructure took several weeks. Since then, various sequencers have been developed that can analyze a human genome in one day for a few thousand dollars. Spiral Genetics can now process thousands of datasets simultaneously while complying with the strictest security requirements for its customers.
- Seven Bridges Genomics, a Cambridge, Massachusetts-based bioinformatics firm, offers researchers and labs a cloud platform for analyzing genetic data generated through nextgeneration sequencing (NGS) technologies. Through its "IGOR" platform, Seven Bridges provides a one-stop solution for managing NGS projects and enables customers to create and run complex data analysis pipelines easily using a drag-and-drop interface.
- In 2013, Novartis scientists discovered a large molecule involved in the disease mechanism of a particular type of cancer. The scientists wanted to virtually screen 10 million compounds against the large molecule. Such a large number of screenings is extremely computationally intensive. Novartis did not have capacity in their existing datacenter to do this type of test, and building new infrastructure would have cost an estimated \$40 million dollars. Instead, they built a virtual high performance computing center in the cloud, and were able to perform the equivalent of 39 years of science in less than nine hours for a cost of \$4,200. In the process, they identified three compounds that may be able to interact with the large molecule they were targeting.

- The number of genetic tests available to doctors and hospitals is constantly increasing, and they
 can be very expensive. Professor Peter Tonellato at Harvard Medical School's Laboratory for
 Personalized Medicine is interested in determining which tests will result in better patient care
 and better results. Using cloud computing technology to cost-effectively and securely analyze
 large amounts of human genomic data, Dr. Tonellato's laboratory is able to identify the tests,
 protocols, and trials that are worth pursuing aggressively for both FDA approval and clinical use.
- Funded by the National Institutes of Health (NIH), the Human Microbiome Project (HMP) is a
 collaborative effort, involving over 300 scientists from more than 80 organizations, to
 comprehensively characterize the microbial communities inhabiting the human body and
 elucidate their role in human health and disease. Large amounts of genetic information on the
 microbes that naturally colonize our bodies—enough information to fill more than 1,000
 standard DVDs—are now available as a free dataset, enabling users to access and analyze the
 data online.
- The 1000 Genomes Project is an international research effort coordinated by a consortium of 75 companies and organizations to establish the most detailed catalogue of human genetic variation. The project dataset has grown to 200 terabytes of genomic information, including DNA sequenced from more than 1,700 people, that researchers can now access via cloud computing for use in disease research.
- Baylor University manages the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) project. CHARGE is a collaboration of hundreds of scientists from around the world who are using data to research the causes and prevention of disease. Together with DNAnexus, Baylor scientists were able to sequence the DNA of more than 14,000 individuals (3,751 whole genomes and 10,771 whole exomes) and analyze over a petabyte of human genomic data in the largest-ever cloud-based analysis of genomic data.
- The High Performance Computing Facility of the New York University (NYU) Center for Health Informatics and Bioinformatics delivers forefront-computing capabilities to researchers at the NYU Langone Medical Center. The facility allows medical informatics and bioinformatics researchers to accelerate discovery and innovation through access to computational power, data storage, supercomputing resources and data sharing with collaborators throughout the world. By using the cloud, the HPC facility expanded the set of services it can offer to NYU researchers, who can now access the resources they need, when they need them. The cloud also helps researchers collaborate so that they easily share their findings and datasets with researchers around the world.
- The Global Alliance for Genomics and Health is an organization formed to help accelerate the
 potential of genomic medicine to advance human health. It brings together over 190 leading
 institutions working in healthcare, research, disease advocacy, life science, and information
 technology. The partners in the Global Alliance are working together to create a common
 framework of harmonized approaches to enable the responsible, voluntary, and secure sharing
 of genomic and clinical data.

- The Icahn School of Medicine at Mount Sinai uses StationX GenePool to mine The Cancer Genome Atlas (TCGA) in order to better understand the genetic component of breast and ovarian cancers. TCGA is a comprehensive and coordinated effort to accelerate our understanding of the molecular basis of cancer through the application of genome analysis technologies, including large-scale genome sequencing. TCGA is a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), two of the 27 Institutes and Centers of the National Institutes of Health.
- The National Database for Autism Research (NDAR) is an extensible, scalable informatics platform for Autism Spectrum Disorder (ASD) relevant data at all levels of biological and behavioral organization (molecules, genes, neural tissue, behavioral, social and environmental interactions) and for all data types (text, numeric, image, time series, etc.). NDAR was developed to share data across the entire ASD field and to facilitate collaboration across laboratories. Cloud-based hosting of this research data is enabling autism researchers to increase collaboration and the sharing of data across federally-funded studies.

Enterprises of all sizes and types also are beginning to use cloud computing for the *development*

of new biomedical treatments and cures. Here are a few examples from our AWS experience:

- Illumina is a San Diego, California-based company that manufactures lab instruments for genetic sequencing. Genetic data produced in the lab requires significant analysis before meaningful information can be produced, such as comparing the raw data produced by the instruments to libraries of genetic anomalies. As the amount of genetic information being produced began to grow, Illumina realized that data storage and processing was a major bottleneck for biologists using their products. Illumina created the IT platform BaseSpace to help address this problem. Previously, researchers who did not have the on-site computational expertise to analyze these datasets would often have to store them on disks and send them in the mail for analysis. Using BaseSpace, they have the analysis performed in the cloud using as much compute power as necessary. This also helps facilitate collaboration between different sites by allowing remote collaborators to log in to BaseSpace and contribute data to a trial or receive data.
- In the summer of 2012, Merck was noticing higher-than-usual discard rates for certain vaccines. Excessive discard rates can lead to disruptions in supply of the vaccines for patients and healthcare providers. However, given the sensitive nature of the manufacturing and storage processes, the high discard rates could result from many factors, including process errors, plant equipment errors, or building environmental controls. Evaluating these factors for every vaccine produced was extremely challenging using a traditional "spreadsheet" approach, so instead Merck worked with a partner to implement a cloud-based solution. Over a three month period, they were able to combine all of their data sources and perform over 15 billion calculations and more than 5.5 million vaccine batch-to-batch comparisons. They were able to identify, quantitatively, that certain characteristics in a fermentation step had a direct impact on

the vaccine discard rate. They were then able to use this information to propose changes in their manufacturing process which could then be tested in the lab.

• Siemens Healthcare Diagnostics is a leading global manufacturer of diagnostic instruments and tests for clinical labs. Over the past several years, Siemens has been partnering with pharmaceutical companies to develop a type of personalized medicine known as Companion Diagnostics, where a diagnostic test is used to help determine whether a particular pharmaceutical treatment will be effective for a particular patient. Siemens has recently developed a new cloud-based Companion Diagnostics analytical platform with applications that typically involve merging genomic analysis with other patient data to determine if a particular drug would be effective. This would be extremely challenging to deliver to a typical clinical lab without drawing on cloud computing to perform the data analysis.

Lastly, enterprises of all sizes and types are beginning to use cloud computing for the *delivery* of

new biomedical treatments and cures. Here are a few examples from our AWS experience:

- Health Guru is a New York City company that provides online health information videos, such as
 instructional videos on how to look for signs of a heart attack and how to manage diabetes
 through diet. Founded in 2006, Health Guru now has a cloud-based library of more than 3,500
 videos and more than one billion cumulative streams on Healthguru.com. Health Guru also
 developed a video syndication technology so that its partners can access the video library for
 their own websites and be able to use Health Guru products and services to manage content.
- The Schumacher Group is a physician practice management company based in Lafayette, Louisiana that supports healthcare providers in over two dozen states. They adopted cloud computing in a number of areas of their business in order to be more nimble and provide better services to their customers and their patients. Some of the services they are now migrating to the cloud include aspects of their business intelligence system and data archiving. Adoption of the cloud has already led to quicker response times and decreased costs: in one example, they were able to save three months of time and \$75,000 on a critical project.
- The U.S. Food and Drug Administration (FDA), which receives 900,000 handwritten reports of adverse drug effects each year, needed a way to make the data entry process more efficient and reduce costs. Using cloud computing, the FDA and Captricity quickly turn manual reports into machine-readable information with 99.7 percent accuracy, reducing costs from 29 dollars per page to 25 cents per page.
- Ideomed has developed innovative mobile health applications focused on managing patients with severe chronic diseases. Ideomed was asked by Spectrum Health, a non-profit Healthcare network in West Michigan, to create a custom healthcare portal. The portal allows patients to track relevant health-related activities, such as whether the patient took their medication or adhered to a recommended exercise regime. Patients are able to provide information via the

web or mobile devices. The information is then transmitted to the patient's physician so he or she can monitor the patient's progress. Ideomed's platform continues to evolve as they add new capabilities, support for new diseases and, by using cloud services, they were able to react quickly to the demands of patients and physicians, while controlling the total cost to Spectrum.

The U.S. Center for Disease Control and Prevention (CDC) is responsible for providing awareness
for all health-related threats and to support responses to these threats at the national, state,
and local level. With CDC's cloud-based BioSense 2.0 program, health officials can exchange
information faster, improve their common awareness of health threats over time and across
regional and national boundaries, and better coordinate investigations and community actions
to protect health.

Chairman Walden and Chairman Pitts, please allow me to suggest three ways that Congress could help accelerate the innovation cycle of discovery, development, and delivery of new biomedical treatments and cures by facilitating the use of cloud computing for healthcare.

First, to help accelerate the *discovery* of new biomedical treatments and cures, Congress could work with NIH to establish and operate cloud-based data management platforms, which federally-funded researchers could use to share their data. Such federally-funded researchers currently are required to share their data openly or make it accessible to others in their field of research. However, researchers are left on their own to fund, develop, and operate data sharing platforms and, as a result, research projects that cost hundreds of thousands (or even millions) of dollars to collect experimental data are forced to discard the data after publication. The only lasting remnant of the data may be a visual graph or written description of the data in a published paper. In some cases, data are forever lost or can only be recreated with investments similar to the cost of the original experiment.

If federal funding agencies, such as NIH, established and operated cloud-based data management platforms, federally-funded researchers would simply upload their research data along with any relevant software resources required to reproduce their analysis of the data. Other researchers in the field could then access the data and software in order to reproduce results, reanalyze previously collected data in novel ways, or even automate the analysis of new data using the same approach as the original experiment. This would result in the elimination of costly and unnecessary duplicative research and thereby accelerate the pace of biomedical discovery.

Second, to accelerate the *discovery* and *development* of new biomedical treatments and cures, Congress could enact both H.R. 967, the Advancing America's Networking and Information Technology Research and Development Act (which would require an assessment of how federal science agencies can facilitate the use of cloud computing for federally-funded science and engineering research), and H.R. 1232, the Federal IT Acquisition and Reform Act (which would assist federal agencies with healthcare, research, and scientific missions to improve their technology capabilities and efficiency).

Third, to help accelerate the *delivery* of new biomedical treatments and cures, Congress could work with the Department of Health and Human Services (HHS) to modernize implementation of the Health Insurance Portability and Accountability Act (HIPAA) so that healthcare providers can readily employ the benefits of cloud computing without any compromise of the strong privacy protections HIPAA now affords health information. To date, HHS has indicated that a cloud computing services provider that stores information for a healthcare provider also is subject to HIPAA as a "business associate," even when the stored information is encrypted and the cloud services provider lacks the decryption key, and regardless of whether the cloud services provider agreed to do so or had otherwise received notice of the health information. This interpretation impedes healthcare delivery entities from leveraging cloud services by causing the parties to negotiate a "business associate agreement" in which virtually all of the terms are inapplicable because the cloud services provider does not have access to health information. By narrowing the application of HIPAA to situations where a cloud provider has access to, and knowledge of, the information, parties can avoid wasting money on contracts that are mostly inapplicable, and cloud services providers can more reasonably comply with HIPAA by focusing on areas where they have access and knowledge of health information.

Chairman Walden and Chairman Pitts, thank you again for holding today's hearing and inviting me to testify. I look forward to your questions.

* * * * * *