

Big Data, Life Sciences, and National Security

April 1, 2014

Renaissance West Ballroom

Renaissance Washington, DC Downtown Hotel
999 Ninth St, NW
Washington, DC

8:15am – 8:45am

Registration

8:45am – 9:00am

Introduction to Event

Welcome Remarks: Alan Leshner, American Association for the
Advancement of Science

9:00am – 10:30am

Big Data: Definition, Sources and Data Sharing

- How is Big Data defined?
- What are the sources of data input?
- In the life and clinical sciences, is data added voluntarily to a database (i.e., genomic databases) or involuntarily?
- How, through what means, and with whom (e.g., broader scientific community, public, and other stakeholders) are big data sets and analytic results shared?
- What are the current and near future tools for analyzing big data?

Moderator: Simon Mercer, Microsoft Research

Speakers: Scott Spangler, IBM Almaden Research Center
Peter Speyer, Institute for Health Metrics and Evaluation
Joel Dudley, Mount Sinai School of Medicine

10:30am – 10:45am

Break

10:45am – 12:15pm

Applications of Big Data and Analytics to International and National Biological Security

- What capabilities could big data and analytics contribute to addressing national and international biological security?
- What lessons have been learned by existing programs/companies/activities in applying big data and analysis to address global health security goals?

Moderator: Tanya Berger-Wolf, University of Illinois, Chicago

Speakers: David McIver, Harvard Medical School and Boston Children's Hospital
Jason Matheny, IARPA
Toby Richardson, Synthetic Genomics

12:15pm – 1:15pm **Lunch** in the Renaissance East Ballroom

1:15pm – 2:30pm **Security Risks of Big Data: Privacy, Openness, Data Management**

- What security risks are associated with big data?
- What security concerns are associated with public availability of data and results?
- What controls, if any, should be in place on the type of data collected and the people or organizations that have access to that data?
- Should controls be stronger for organizations or individuals with greater access to the data and analytic tools?
- How do the security concerns differ between data collected from research activities, in clinical settings, and from public sources?
- What precautions must be taken to prevent cyber crime and release of sensitive data?
- How might repositories of data, which because compiled could be valuable, be vulnerable to hacking?
- How vulnerable are data repositories to theft, changes in content or functionality, or other type of attack?

Moderator: Piers Millet, Implementation Support Unit of the Biological Weapons and Toxins Convention

Speakers: Mark Greaves, Pacific Northwest National Laboratory
Robert Sloan, University of Illinois at Chicago
Herb Lin, National Academy of Sciences

2:30pm – 2:45pm **Break**

2:45pm – 4:00pm **Increasing Access to Big Data and Analytics: Implications for Biological Security?**

- Will the data collection, sharing, and analytic technologies tools become available and accessible to a broader set of individuals, including amateur scientists and the public?
- Who constitutes the “amateur” scientific community and broader public collecting and analyzing large biological data sets? What is their skill level and access to specialized knowledge, equipment, facilities, experts, and/or online analytic tools?
- What degree of specialized knowledge and skills are required for collecting and analyzing large data sets?

- Is there a difference in capability between users of purchased analytic tools compared to freely available tools?
- How significant is the gap - in capability and access - between amateurs, creative but inexperienced scientists-in-training, and expert scientists, now and in the near future?
- Can collection and analysis of large data sets be carried out by “other” scientists, health practitioners, and/or amateur scientists?
- What capabilities are associated with technology diffusion of big data and analytics?
- What security risks might arise from greater access to big data and analytic tools?

Moderator: Nathan Price, Institute for Systems Biology

Speakers: Folker Meyer, Argonne National Laboratory
Ivo Dinov, University of Michigan
Lorelei Kelly, New America Foundation
Daniel Grushkin, GenSpace

4:00pm – 5:15pm

Implications of Big Data and Analytics to National and International Biological Security

- What risks might be associated with big data and analytics?
- What challenges and opportunities exist in adequately preventing harmful uses and promoting peaceful uses of big data in the life science and analytics?
- How do these efforts change or promulgate as non-US companies and organizations begin to develop and use big data and analytic tools?
- What capabilities could big data and analytics offer to national and international biological security? Capabilities would *not* include unwitting surveillance of people.
- What additional considerations are associated with big data and analytics?

Moderator: Margaret Kosal, Georgia Institute of Technology

Speakers: Susan Collar-Monarez, National Security Council (*invited*)
James Lawler, Navy Medical Research Center
Edward You, Federal Bureau of Investigation
Gaymon Bennett, Fred Hutchinson Cancer Research Center

5:15pm – 6:00pm

General Discussion and Comments

6:00pm

Adjourn