

# AIDR 2020 Speaker Bios

## Session 1

### **Melissa Haendel**

Melissa Haendel is the Director of the Center for Data to Health (CD2H) at Oregon Health Science University, and the Director of Translational Data Science at Oregon State University. Her background is molecular genetics and developmental biology as well as translational informatics, with a focus over the past decade on open science and semantic engineering. Dr. Haendel's vision is to weave together healthcare systems, basic science research, and patient-generated data through development of data integration technologies and innovative data capture strategies. Dr. Haendel's research has focused on integration of genotype-phenotype data to improve rare disease diagnosis and mechanism discovery. She also leads and participates in international standards organizations to support improved data sharing and utility worldwide.

### **Rayid Ghani**

Rayid Ghani is a Distinguished Career Professor in the Machine Learning Department and the Heinz College of Information Systems and Public Policy at Carnegie Mellon University. Rayid is a reformed computer scientist and wanna-be social scientist, focused on the use of large-scale AI/Machine Learning/Data Science in solving critical public policy and social challenges in a fair and equitable manner. Among other areas, Rayid works with governments and non-profits in policy areas such as health, criminal justice, education, public safety, economic development, and urban infrastructure. Rayid is also passionate about teaching practical data science and started the Data Science for Social Good Fellowship that trains computer scientists, statisticians, and social scientists from around the world to work on data science problems with social impact.

Before joining Carnegie Mellon University, Rayid was the Founding Director of the Center for Data Science Public Policy, Research Associate Professor in Computer Science, and a Senior Fellow at the Harris School of Public Policy at the University of Chicago. Previously, Rayid was the Chief Scientist of the Obama 2012 Election Campaign where he focused on data, analytics, and technology to target and influence voters, donors, and volunteers. In his ample free time, Rayid obsesses over everything related to coffee and works with non-profits to help them with their data, analytics and digital efforts and strategy.

### **Bo Alroe**

Bo has worked in the private sector with higher education institutions since 2004. He has collaborated with universities, publishers, and government institutions to develop tools, data, skills, and processes for research management and administration.

As co-owner and Product Director at Atira, the Danish software company behind the PURE system, Bo worked over a decade with universities to build and manage CRIS or RIMS systems; research information management systems and grants management. After working for Elsevier from 2013 to 2016 following their acquisition of Atira, he has been at Digital Science in London as Director of Strategy, where his primary task has been the launch and commercial operations for Dimensions, Digital Science' linked scholarly information database. He leads a small team of specialists in research administration, management and analyses.

### **Lily Troia**

Lily Troia is an academic librarian and archivist with a background in research services and metrics, scholarly communication, open access, and digital and data preservation + diverse experience in project management and digital media outreach. In her current role with Digital Science she helps guide organizational leadership

teams strategize around digital tools and data that align with mission, actionable use, increased efficiency, and decision-making. Lily currently sits on the Board of Directors of the Society for Science at User Research Facilities, where she serves as Communications Chair. She held previous positions at the College of William & Mary (Virginia Institute of Marine Science) and Harvard's Office of Scholarly Communication. She earned her undergraduate degree at Northwestern University in ethnomusicology and her Master's in Library and Information Science at Simmons University.

## Session 2

### **Mark A Musen**

Dr. Musen is Professor of Biomedical Informatics and of Biomedical Data Science at Stanford University, where he is Director of the Stanford Center for Biomedical Informatics Research. Dr. Musen conducts research related to open science, intelligent systems, computational ontologies, and biomedical decision support. His group developed Protégé, the world's most widely used technology for building and managing terminologies and ontologies. He served as principal investigator of the National Center for Biomedical Ontology, one of the original National Centers for Biomedical Computing created by the U.S. National Institutes of Health (NIH). He directs the Center for Expanded Data Annotation and Retrieval (CEDAR), founded under the NIH Big Data to Knowledge Initiative. CEDAR develops semantic technology to ease the authoring and management of biomedical experimental metadata.

### **Ken Koedinger**

Kenneth R. Koedinger is a Professor of Human-Computer Interaction and Psychology at Carnegie Mellon University. Dr. Koedinger has an M.S. in Computer Science, a Ph.D. in Cognitive Psychology, and experience teaching in an urban high school. His multidisciplinary background supports his research goals of understanding human learning and creating educational technologies that increase student achievement. His research has contributed new principles and techniques for the design of educational software and has produced basic cognitive science research results on the nature of student thinking and learning. Koedinger directs LearnLab (learnlab.org), which started with 10 years of National Science Foundation funding and is now the scientific arm of CMU's Simon Initiative. LearnLab builds on the past success of Cognitive Tutors, an approach to online personalized tutoring that is in use in thousands of schools and has been repeatedly demonstrated to increase student achievement, for example, doubling what algebra students learn in a school year. He was a co-founder of Carnegie Learning, Inc. that has brought Cognitive Tutor based courses to millions of students since it was formed in 1998, and leads LearnLab, now the scientific arm of CMU's Simon Initiative. Dr. Koedinger has authored over 250 peer-reviewed publications and has been a project investigator on over 45 grants. In 2017, he received the Hillman Professorship of Computer Science and in 2018, he was recognized as a fellow of Cognitive Science.

### **Adriana Kovashka**

Adriana Kovashka is an Assistant Professor in Computer Science at the University of Pittsburgh. Her research interests are in computer vision and machine learning. She has authored eighteen publications in top-tier computer vision and artificial intelligence conferences and journals (CVPR, ICCV, ECCV, NeurIPS, AAAI, ACL, TPAMI, IJCV) and ten second-tier conference publications (BMVC, ACCV, WACV). She has served as an Area Chair for CVPR in 2018-2021, NeurIPS 2020, ICLR 2021, AAAI 2021, and will serve as co-Program Chair of ICCV 2025. She has been on program committees for over twenty conferences and journals, and has co-organized seven workshops. Her research is funded by the National Science Foundation, Google, Amazon and Adobe.

## Fireside Chat

### **Martial Hebert**

Martial Hebert is a Professor of Robotics and Dean of the School of Computer Science at Carnegie Mellon University. His research interests include computer vision and robotics, especially recognition in images and video data, model building and object recognition from 3D data, and perception for mobile robots and for intelligent vehicles. His group has developed approaches for object recognition and scene analysis in images, 3D point clouds, and video sequences. In the area of machine perception for robotics, his group has developed techniques for people detection, tracking, and prediction, and for understanding the environment of ground vehicles from sensor data. He currently serves as Editor-in-Chief of the International Journal of Computer Vision.

## Session 3

### **Lucy Lu Wang**

Lucy Lu Wang (@lucyluwang) is a Postdoctoral Investigator at the Allen Institute for AI, where she is part of the Semantic Scholar Research team. She works on research related to scientific document representation and understanding, specifically in the areas of biomedical natural language processing, ontology, open access, and meta-science. Her work on supplement interaction detection and gender trends in publishing has been featured in Geekwire, Gizmodo, Axios, VentureBeat, and the New York Times. She completed her PhD at the University of Washington in Biomedical and Health Informatics.

### **Kyle Lo**

Kyle Lo (@kylelost) is an applied research scientist at the Allen Institute for AI on the Semantic Scholar Research team, where he works on NLP for scientific literature with emphasis on knowledge representation, discourse, and science-of-science. His work on domain adaptation of language models was nominated for best paper at ACL, and his work on sex bias in clinical trials and scientific fact checking has been featured in articles in MIT Tech Review and Quartz. He has an MS in Statistics from the University of Washington.

### **Imran Haque**

Imran is the VP, Data Science at Recursion Pharmaceuticals, where he leads a team at the intersection of machine learning, statistics, biology, and chemistry to develop Recursion's unique AI-driven approach to decoding biology. He was the first chief scientific officer at Freenome, where he built and led an R&D team spanning molecular biology, computational biology, and machine learning to develop non-invasive blood-based screening protocols for early cancer detection. He also served for over five years in roles of increasing responsibility at Counsyl (now Myriad Women's Health). He earned a Bachelor's degree in electrical engineering and computer science from UC Berkeley and a Ph.D. in computer science from Stanford University, where his research on large-scale machine learning systems for drug discovery was co-advised by Vijay Pande and Daphne Koller. In his spare time, Imran builds fast computers, drives fast cars and skis and bikes at a much more reasonable pace than he did in his blithe youth.

Follow Imran on Twitter @ImranSHaque

### **Yubin Kim**

Yubin Kim is Director of Technology at UPMC Enterprises where she works with stakeholders both in the provider and payer side of the healthcare system to identify and solve problems through the application of

NLP and ML. She received her PhD from the Language Technologies Institute of Carnegie Mellon University. Her current interests lie in real-world applications of search, NLP and ML technology to the healthcare domain.

### **Jeremy Chen Weiss**

Jeremy C. Weiss, M.D. Ph.D., is Assistant Professor of Health Informatics in the Heinz College of Information Systems and Public Policy at Carnegie Mellon University and affiliate/adjunct in the Machine Learning Department and the Pittsburgh Department of Biomedical Informatics. His research involves the development and application of machine learning techniques for the improvement of clinical prediction in internal medicine and critical care settings, with applications in opioid abuse, diabetes, and sepsis. He directs the Care, Health, and Reasoning Machines (CHARM) lab that develops longitudinal forecasting models from electronic health records, where the data therein are a result of distinct collection mechanisms and necessitate advanced analytics for performance, appropriate use and interpretation. He holds awards from the Center for Machine Learning and Health, the CMU Portugal Program, Amazon Web Services and Microsoft Azure, publishes across venues from NeurIPS to JAMA, and has won top reviewing awards from NeurIPS to the Annals of Internal Medicine.

## **Panel Discussion**

### **Sean Davis**

Sean Davis is a Senior Associate Scientist at the Center for Cancer Research at the National Cancer Institute, part of the Intramural Research Program at the National Institutes of Health. He received his B.S. in Mechanical and Aerospace Engineering from Princeton University, M.D. and Ph.D. from the University of Pittsburgh in the Schools of Medicine and Public Health (thesis advisor, Dr. Daniel Weeks, Human Genetics), respectively. Children's Hospital and Regional Medical Center was where he completed his residency in Pediatrics, followed by pediatric hematology/oncology training in the Johns Hopkins School of Medicine/National Cancer Institute joint fellowship. Sean works broadly in the areas of genomics and bioinformatics of high-throughput data analysis. In addition to data analysis and biologically-driven projects, he has a strong interest in developing software and tools for analyzing and interpreting genomic data and helps to lead the Bioconductor project.

### **Ross Epstein**

Ross is the Chief of Staff at SafeGraph. Prior to SafeGraph, he was the CEO and co-founder of sales engagement platform Sendbloom, later acquired by LinkedIn. Ross holds a B.S. in Computer Science from Northwestern University.

### **Lucy Lu Wang**

Lucy Lu Wang (@lucyluwang) is a Postdoctoral Investigator at the Allen Institute for AI, where she is part of the Semantic Scholar Research team. She works on research related to scientific document representation and understanding, specifically in the areas of biomedical natural language processing, ontology, open access, and meta-science. Her work on supplement interaction detection and gender trends in publishing has been featured in Geekwire, Gizmodo, Axios, VentureBeat, and the New York Times. She completed her PhD at the University of Washington in Biomedical and Health Informatics.

### **Alison Specht**

Alison Specht is an environmental scientist with broad expertise in research, teaching, and community engagement. In the past 10 years she has focussed on facilitating interdisciplinary groups to tackle complex

environmental problems using existing data. Her major interests, apart from her domain activities, are to enhance collaboration between scientists, policy-makers and managers to improve environmental outcomes, and improve data management and preservation, data sharing and re-use.

In 2009, Alison established the first synthesis centre in the southern hemisphere, the Australian Centre for Ecological Analysis and Synthesis, a facility of the Terrestrial Ecosystem Research Network. Between 2015 and 2018 she was Director of CESAB, the CEntre for the Synthesis and Analysis of Biodiversity in France. In 2011 she initiated the formation of the International Synthesis Consortium. Alison has been a member of the DataONE Usability and Assessment Working Group in the USA since its inception in 2010. She is a member of several RDA interest groups. She is currently an honorary associate professor at the University of Queensland, Australia and an Ecosystem Research Analyst for TERN. She is a co-lead of PARSEC, a five-country Belmont-funded project entitled 'Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas'.

### **Keith Webster**

Keith Webster was appointed Dean of University Libraries at Carnegie Mellon University in July 2013 and was additionally appointed as Director of Emerging and Integrative Media Initiatives in July 2015. He also has a courtesy academic appointment at the University's H. John Heinz III College. Previously, Keith was Vice President and Director of Academic Relations and Strategy for the global publishing company John Wiley and Sons. He was formerly Dean of Libraries and University Librarian at the University of Queensland in Australia, leading one of the largest university and hospital library services in the southern hemisphere. Earlier positions include University Librarian at Victoria University in New Zealand, Head of Information Policy at HM Treasury, London, and Director of Information Services at the School of Oriental & African Studies, University of London.

Keith has held professorships in information science at Victoria University of Wellington and City University, London. He is a Chartered Fellow and an Honorary Fellow of the Chartered Institute of Library and Information Professionals (UK), and has served on government advisory boards, journal editorial boards, and as an officer in professional and learned societies around the world. He was Chair of the National Information Standards Organization 2018-2019.

Keith's professional interests include research evaluation, learning space design and trends in scholarly communication. He is a regular speaker on topics such as the future of research libraries and the impact of open science on publishing and libraries.