

ABOUT THE SPEAKERS

Welcome & Opening Remarks

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 became Chair of the National Information Standards Organization on 1 July 2018.

J. Michael McQuade is Carnegie Mellon University's Vice President for Research, providing leadership for the University's research enterprise and advocating for the role that science, technology and innovation play nationally and globally. From 2006 to 2018 he served as Senior Vice President for Science & Technology at United Technologies Corporation. At UTC, McQuade's responsibilities included providing strategic oversight and guidance for research, engineering and development activities throughout the business units of the corporation and at the United Technologies Research Center, focused on a broad range of high-technology products and services for the global aerospace and building systems industries. Dr. McQuade held senior positions with technology development and business oversight at 3M, Imation and Eastman Kodak. He served as Vice President of 3M's Medical Division and President of Eastman Kodak's Health Imaging Business. His early career at 3M was focused on research and development of high-end acquisition, processing and display systems for health care, industrial imaging and remote sensing. He has broad experience managing basic technology development and the conversion of early stage research into business growth. Dr. McQuade holds Ph.D., M.S. and B.S. degrees in physics from Carnegie Mellon University. He received his Ph.D. in experimental high-energy physics for research performed at the Fermi National Accelerator Laboratory on charm quark production. Dr. McQuade served as a member of the President's Council of Advisors on Science and Technology and of the Secretary of Energy Advisory Board and is a member of the Defense Innovation Board.

Beth Plale: Dr. Plale is a Science Advisor at the National Science Foundation in the Office of Advanced Cyberinfrastructure (OAC) of the Computer Information Science and Engineering Directorate (CISE). Dr. Plale is working on open science and serving as program officer for the Big Data Regional Innovation Hubs (BD Hubs) program. She is a “rotator” from Indiana University Bloomington where she is Full Professor in the Dept. of Intelligent Systems Engineering. Her research interests are data management, cloud computing, Big Data, and open science.

Keynote Speakers

Tom M. Mitchell is the E. Fredkin University Professor and Interim Dean of the School of Computer Science at Carnegie Mellon University. His research involves machine learning approaches to natural language understanding, and uses brain imaging to study how the human brain understands language. Mitchell is a member of the U.S. National Academy of Engineering, a member of the American Academy of Arts and Sciences, and a Fellow and Past President of the Association for the Advancement of Artificial Intelligence (AAAI). He is also interested in societal impacts of computer science, recently co-chaired a U.S. National Academies study on “Information Technology, Automation, and the Workforce,” and has recently testified to several Congressional forums on the impacts of AI on society.

Glen de Vries: Glen is the President and Co-founder of Medidata Solutions, the leading cloud platform for life sciences research. Glen has been driving Medidata's mission since the company's inception in 1999: Powering smarter treatments and healthier people. His publications have appeared in Applied Clinical Trials, Cancer, The Journal of Urology, Molecular Diagnostics, STAT, Urologic Clinics of North America and TechCrunch. He is a trustee of Carnegie Mellon University, a Columbia HITLAB Fellow, and a member of the Healthcare Businesswomen's Association European Advisory Board. Glen received his undergraduate degree in molecular biology and genetics from Carnegie Mellon University, worked as a research scientist at the Columbia Presbyterian Medical Center and studied computer science at New York University's Courant Institute of Mathematics.

Invited Speakers & Panelists

Natasha Noy is a scientist at Google AI where she works on making structured data accessible and useful. Currently she leads the team building Google Dataset Search. Prior to joining Google, Natasha worked at Stanford Center for Biomedical Informatics Research where she made major contributions in the areas of ontology development and alignment, and collaborative ontology engineering.

Clifford Lynch has led the Coalition for Networked Information (CNI) since 1997. CNI, jointly sponsored by the Association of Research Libraries and EDUCAUSE, includes about 200 member organizations concerned with the intelligent uses of information technology and networked information to enhance scholarship and intellectual life. CNI's wide-ranging agenda includes work in digital preservation, data intensive scholarship, teaching, learning and technology, and infrastructure and standards development. Prior to joining CNI, Lynch

spent 18 years at the University of California Office of the President, the last 10 as Director of Library Automation. Lynch, who holds a Ph.D. in Computer Science from the University of California, Berkeley, is an adjunct professor at Berkeley's School of Information.

He is both a past president and recipient of the Award of Merit of the American Society for Information Science, and a fellow of the American Association for the Advancement of Science, the Association for Computing Machinery, and the National Information Standards Organization. He served as co-chair of the National Academies Board on Research Data and Information from 2011-2016; he is active on numerous advisory boards and visiting committees, notably the National Academies Committee on Forecasting Costs for Preserving, Archiving and Promoting Access to Biomedical Data. His work has been recognized by the American Library Association's Lippincott Award, the EDUCAUSE Leadership Award in Public Policy and Practice, and the American Society for Engineering Education's Homer Bernhardt Award.

Casey Greene: Casey's lab at the University of Pennsylvania is dedicated to developing computational tools that biologists use to gain insights from other labs' data as easily as from their own. More than 2 billion dollars' worth of publicly funded genomics data are freely downloadable. These data represent a rich and underused resource, but they are hard to use because the data are comprised of many different experiments. Standard algorithms struggle with these "messy" datasets. Casey's lab develops computational techniques that are robust enough to analyze and interpret this public resource. These algorithms have applications across many disease areas and, perhaps most importantly, to questions of basic biology. A list of publications is available on his lab's website (<http://greenelab.com/publications>). Casey's devotion to the analysis of publicly available data doesn't stop in the lab. In 2016, Casey established the "Research Parasite Awards" after an editorial in the New England Journal of Medicine deemed scientists who analyze other scientists' data "research parasites." These honors, accompanied by a cash prize, are awarded to scientists who rigorously reanalyze other people's data to learn something new.

Lisa S. Parker: Dr. Parker, a philosopher, is Professor of Human Genetics and Director of the University's Center for Bioethics & Health Law. She also directs the University's interdisciplinary Master of Arts in Bioethics program. Her research focuses on ethical concerns related to the design and conduct of research, particularly genetic research and mental health research, ethical issues in genomics and precision medicine, confidentiality, and informed consent. She chaired the Genomics and Society Working Group of the National Advisory Council for Human Genome Research for the NHGRI, where she has served on two national working groups examining the management of incidental findings and research results, as well as on the Expert Scientific Panel of the Electronic Medical Records and Genomics (eMERGE) Network.

Sean Davis: Sean is a pediatric oncologist and mechanical engineer by training and is currently at the Center for Cancer Research at the National Cancer Institute. He has developed a number of tools and software products that facilitate genomic data reuse (GEOquery, GEOmetadb, SRadb, and GenomicDataCommons) in the context of his involvement with the Bioconductor project. He advisory roles with several large data and compute-related efforts at the National Institutes of Health. Sean is a community advocate for open data and data science literacy and education.

Irene Kaplow develops computational methods to study the mechanisms involved in transcriptional regulation, the process that is partially responsible for defining the differences between tissues and cell types within an organism, and how these mechanisms evolved. Irene does this by viewing different species as natural perturbations and analyzing multi-species high-throughput sequencing datasets. Because these

perturbations occur at DNA sequence level, comparing regulatory genomics data across species can provide insights into how DNA sequence affects different components of transcriptional regulation. In addition to providing insights into transcriptional regulatory mechanisms, these studies can help us predict how the transcriptional regulatory mechanisms function in a species for which the relevant high-throughput sequencing data has not been generated. Since many differences between species are regulated through transcription, such studies can improve our understanding of evolution, enabling us to understand traits that make humans unique and helping us to determine how to preserve species that are on the verge of becoming extinct. Irene received her B.S. in Mathematics from the Massachusetts Institute of Technology in 2010 and her Ph.D. in Computer Science from Stanford University in 2017.

Fiona Nielsen: Fiona is a bioinformatics scientist-turned entrepreneur. Having worked at Illumina developing tools for interpretation of next-generation sequencing data and analysing cancer genomes, Fiona realised the main bottleneck for genome interpretation and precision medicine is accessing the right data. She decided to disrupt outdated practices with the aim of accelerating drug discovery.

In 2013, Fiona founded DNAdigest as a charity to promote best practices for efficient and ethical data sharing, and in 2014, she co-founded Repositive to develop an online community platform and global exchange for genomic data.

Fiona's familiarity with the industry ensured that Repositive's platforms would provide data access solutions while also respecting the industry's need for privacy, data governance and IP protection. This awareness was fundamental in securing venture capital investment and gaining support from worldwide partners including, among others, AstraZeneca, Boehringer Ingelheim, XenTech and Shanghai LIDE Biotech.

Fiona features on the [2018 WISE100 list](#), an index of the UK's 100 most inspiring and influential women in social enterprise, and she regularly speaks on invited panels on the subjects of genomics, the future of medicine, and ethical data sharing.

Alex John London, Ph.D., is the Clara L. West Professor of Ethics and Philosophy and Director of the Center for Ethics and Policy at Carnegie Mellon University. He is an elected Fellow of the Hastings Center whose work focuses on ethical and policy issues surrounding the development and deployment of novel technologies in medicine, biotechnology and artificial intelligence. He has published over 85 papers in leading philosophy journals (such as *Mind* and the *Philosopher's Imprint*), high impact science and medical journals (such as *Science*, *eLife*, *JAMA*, *The Lancet*, and *PLoS Medicine*), as well as numerous other journals and collections and is co-editor of *Ethical Issues in Modern Medicine*, one of the most widely used textbooks in medical ethics. Professor London's work on ethics and AI examines the nature of algorithmic bias, the relative value of explainability and verifiability, how to encode alternative models of moral decision making in formal systems, social trust and the nature and source of uncertainty in AI systems. He is currently a judge for the IBM Watson AI X-Prize. From 2016-2017 he was part of the U.S. National Academy of Medicine Committee on Clinical Trials. During the 2014-15 Ebola Outbreak he has served as an ethics expert in consultations with numerous national and international organizations including the U.S. National Institutes of Health, the World Health Organization, the World Medical Association, and the World Bank.

Nick Nystrom is Chief Scientist at the Pittsburgh Supercomputing Center (PSC), which is a joint effort of Carnegie Mellon University and the University of Pittsburgh. Nick is architect and PI for Bridges, PSC's flagship system that successfully pioneered the convergence HPC, AI, and Big Data. He is also PI for the NIH Human Biomolecular Atlas Program (HuBMAP) Infrastructure and Engagement Component and co-PI for projects that

bring emerging AI technologies to research (Open Compass), apply machine learning to biomedical data for breast and lung cancer (Big Data for Better Health), and identify causal relationships in biomedical big data (the Center for Causal Discovery, an NIH Big Data to Knowledge Center of Excellence). His current research interests include hardware and software architecture, applications of machine learning to research data (particularly in the life sciences) and to enhance simulation, frameworks for improving the sharing and accessibility of data, and graph analytics.

Robert F. Murphy is the Ray and Stephanie Lane Professor of Computational Biology and Head of the Computational Biology Department in the School of Computer Science at Carnegie Mellon University. He also is Professor of Biological Sciences, Biomedical Engineering, and Machine Learning. He is Honorary Professor of Biology at the University of Freiburg, Germany, and was the recipient of an Alexander von Humboldt Foundation Senior Research Award, and is also a Senior Fellow of the Allen Institute for Cell Science. He is a Fellow of the American Institute for Medical and Biological Engineering and a Senior Member of the IEEE and the International Society for Computational Biology. He received the Distinguished Service Award from the International Society for Advancement of Cytometry in 2016 and served as President of the Society from 2008-2010. He was a member of the National Advisory General Medical Sciences Council and the National Institutes of Health Council of Councils. He is an Associate Editor for Bioinformatics, and was Associate Editor for BMC Bioinformatics and Software Editor for PLoS Computational Biology.

Dr. Murphy's career has centered on combining fluorescence-based cell measurements with quantitative and computational methods. His group at Carnegie Mellon pioneered the application of machine learning methods to fluorescence microscope images depicting subcellular location patterns, and was the first to demonstrate superior machine performance in interpreting diverse patterns in biological images compared to human interpretation. His current research interests include machine learning of image-derived models of cell organization, automated detection of protein location changes during oncogenesis, and active machine learning approaches to experimental biology.

Matt Fredrikson is an Assistant Professor in the Computer Science Department at Carnegie Mellon University. His research focuses on security, privacy, and fairness in systems that handle personal data, leveraging formal methods to find and repair problematic system behaviors that lie at the root of these issues. His work has identified new classes of vulnerabilities and sources of bias in machine learning systems, and explored the feasibility of provable mitigations to applications in medicine, social science, and cybersecurity. Prior to joining Carnegie Mellon, he completed his PhD at the University of Wisconsin in 2015.