

DATA SCIENCE: OPPORTUNITIES TO TRANSFORM CHEMICAL SCIENCES & ENGINEERING

A Chemical Sciences Roundtable Workshop

FEBRUARY 27-28, 2018 • 2101 Constitution Ave NW

DAY 1 • NAS 125

February 27, 2018

8:30 AM	Opening Remarks Bruce Garrett US Department of Energy	1:30 PM	Creating Data Science Programs in Academia Thomas Ewing Virginia Tech
8:40 AM	What is Data Science? Andrew Ferguson University of Illinois at Urbana-Champaign	2:15 PM	ADJOURN TO BREAKOUT SESSIONS
9:20 AM	What Can Data Science Do for Me? Patrick Riley Google	Breakout 1: Data Analytics • Members' Room Co-chairs: Sharon Glotzer University of Michigan Bryce Meredig Citrine Informatics	
10:00 AM	Data Science in Industry Salvador Garcia-Munoz Eli Lilly and Company	Breakout 2: Data Quality Assessment • NAS 114 Co-chairs: David Higdon Virginia Tech Carlos Gonzalez National Institutes of Standards and Technology	
10:40 AM	Break	Breakout 3: Data Management and Translation • NAS 118 Co-chairs: Jens Hummelshøj Toyota Research Institute John Gregoire Caltech University	
11:00 AM	Untapped Data Resources Thomas Bligaard SUNCAT Center for Interface Science and Catalysis SLAC National Accelerator Laboratory		
11:40 AM	Panel Discussion (Moderator, Bruce Garrett; Panelists, Morning Speakers): How can data science be used to gain insights or new knowledge from data resources? What are the barriers that are preventing broader adoption of data sciences in the chemical sciences and engineering?		
12:30 PM	LUNCH BREAK	4:45 PM	Wrap-up and Plan for Day 2
		5:00 PM	Adjourn Day 1

DAY 2 • NAS 125

February 28, 2018

8:30 AM	Reporting from Co-Chairs	Melissa Cragin University of Illinois at Urbana-Champaign
10:00 AM	Panel Discussion: Training a Future Generation of Chemists/Chemical Engineers	Al Hero University of Michigan
Moderator:	Leo Chiang The Dow Chemical Company	
Panelists:	Richard Braatz Massachusetts Institute of Technology	Rebecca Nugent Carnegie Mellon University
	Lloyd Colegrove The Dow Chemical Company	11:30 AM Closing Remarks Bruce Garrett
		12:00 PM Workshop Concludes

Speaker Biographies

Thomas Bligaard is a Senior Staff Scientist at the SLAC National Acceleratory Laboratory and the Deputy Director for Theory at the SUNCAT Center for Interface Science and Catalysis. His research goals are to design materials directly from atomic-scale simulations and to establish the tools and the concepts to carry out this task systematically. He currently focuses primarily on the physical and chemical properties of surfaces, heterogeneous catalytic processes, the properties of transition metal alloys, the development of improved methods for sampling entropic contributions, and establishing methods for improving accuracy and error estimation in exchange and correlation functionals for electron density functional theory. Dr. Bligaard received his PhD in Theoretical Solid-State Physics from the Technical University of Denmark.

Richard Braatz is the Edwin R. Gilliland Professor of Chemical Engineering at the Massachusetts Institute of Technology. His research interests include systems and control theory, materials and pharmaceuticals manufacturing, systems nanotechnology, and applied mathematics. A few of his many honors include the Automatica Paper Prize, the AIChE Computing in Chemical Engineering Award, and the Technical Innovation Award from the Internal Society of Automation. Dr. Braatz received his PhD in Chemical Engineering from the California Institute of Technology.

Lloyd Colegrove is a Director of Data Sciences and Fundamental Problem Solving at the DOW Chemical Company. His research focuses on applications of novel numerical analysis approaches to chemical operations in real time over thousands of data streams. A few of his many honors include the Frost and Sullivan Kinetic Process Award in 2014 and the Frost and Sullivan Manufacturing Leadership Award in 2015 and 2016. Dr. Colegrove received his PhD in Chemical Physics from Texas A&M University.

Melissa Cragin is the Executive Director of the Midwest Big Data Hub (MBDH) based at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign. The MBDH is one of four Big Data Hubs launched by the National Science Foundation in 2015, which are intended to strengthen the data ecosystem and develop effective academic-industry-government-NGO networks to address scientific and social issues of regional and national interest. Dr. Cragin received her PhD in Information Sciences from the University of Illinois at Urbana-Champaign.

Thomas Ewing is a Professor in the Department of History and the Associate Dean of the College of Liberal Arts and Human Sciences at Virginia Tech. His research includes using tools and methods from medical history, data analytics, and digital humanities to explore the spread of information, the spread of disease, and the dissemination of knowledge during a public health crisis. Some of his many honors include Virginia Tech Research Awards from Phi Beta Kappa and the College of Liberal Arts and Human Sciences, and the Land Grant Scholar Award. Dr. Ewing received his PhD in Russian history from the University of Michigan.

Andrew Ferguson is an Assistant Professor of Materials Science and Engineering in the Department of Materials Science and Engineering at the University of Illinois at Urbana-Champaign. His research lies at the intersection of materials science, molecular simulation, and machine learning, with particular interests in the rational design of antiviral vaccines, computational discovery and engineering of self-assembling colloids and peptides, and development of nonlinear dimensionality reduction and enhanced sampling algorithms in molecular simulation. Some of his many honors include the UIUC College of Engineering Dean's Award for Excellence in Research, the ACS OpenEye Outstanding Junior Faculty Award, and the AIChE CoMSEF Young Investigator Award for Modeling and Simulation. Dr. Ferguson received his PhD in Chemical and Biological Engineering from Princeton University.

Salvador Garcia-Munoz is a Senior Engineering Advisor at Eli Lilly and Company and is the senior technical leader for digital design in small molecule development. His current research spans the drug substance and drug product areas with particular focus on the use of systems engineering tools in regulatory documents and the transfer of modeling technology to manufacturing. Some of his many honors include the Lilly Research Laboratories President's Scientific Award, the AIChE Pharmaceutical Development, Design, and Manufacturing Forum Award, and the Pfizer Global Research and Development Achievement Award. Dr. Garcia-Munoz received his PhD in Chemical Engineering from McMaster University.

Sharon Glotzer (NAS) is the John Werner Cahn Distinguished University Professor of Engineering, the Stuart W. Churchill Collegiate Professor of Chemical Engineering, and a Professor of Materials Science and Engineering, Physics, Applied Physics, and Macromolecular Science and Engineering at the University of Michigan in Ann Arbor. She is member of the National Academy of Sciences and the American Academy of Arts and Sciences, a fellow of the American Physical Society, and the American Association for the Advancement of Science. Her research on computational assembly science and engineering aims toward predictive materials design of colloidal and soft matter, with current emphasis on shape, packing, and assembly pathways. Dr. Glotzer received her PhD in Physics from Boston University.

Al Hero is the John H. Holland Distinguished University Professor of Electrical Engineering and Computer Science and the R. Jamison and Betty Williams Professor of Engineering at the University of Michigan in Ann Arbor. He is also the Co-Director of the University's Michigan Institute for Data Science. His group focuses on building foundational theory and methodology for data science and engineering. He is developing theory and algorithms for data collection, analysis, and visualization that use statistical machine learning and distributed optimization. These approaches are being applied to network data analysis, personalized health, multi-modality information fusion, data-driven physical simulation, materials science, dynamic social media, and database indexing and retrieval. Some of his many honors include the Society Award and Technical Achievement Award from the IEEE Signal Processing Society. Dr. Hero received his PhD in Electrical Engineering from Princeton University.

David Higdon is a Professor in the Social Decision Analytics Laboratory at the Biocomplexity Institute of Virginia Tech. He previously spent 10 years as a scientist or group leader of the Statistical Sciences Group at Los Alamos National Laboratory. He is an expert in Bayesian statistical modeling of environmental and physical systems, combining physical observations with computer simulation models for prediction and inference. His research interests include space-time modeling; inverse problems in hydrology and imaging; statistical modeling in ecology, environmental science, and biology; multiscale models; parallel processing in posterior exploration; statistical computing; and Monte Carlo and simulation based methods. Dr. Higdon received his PhD in Statistics from the University of Washington.

Jens Hummelshoj is the Materials Design and Discovery Program Manager at the Toyota Research Institute. His research focuses on electronic structure theory, catalysis informatics, energy storage, method scaling relations, analytical applications, and dimensional data warehouses. Dr. Hummelshoj received his PhD in Theoretical and Mathematical Physics from the Technical University of Denmark.

Bryce Meredig is the Cofounder and Chief Science Officer at Citrine Informatics. His research focuses on enabling a future in which every materials scientist and chemist can use large-scale data and machine learning routinely in their day-to-day work. His research and development team focuses on machine learning method development specifically for the physical sciences; machine learning for explicitly integrating experiment, computation, and theory; sequential learning; and data structures and standards. Dr. Meredig received his PhD in Materials Science and Engineering from Northwestern University.

Rebecca Nugent is the Associate Department Head, Director of Undergraduate Studies, and a Professor in the Department of Statistics and Data Science at Carnegie Mellon University. Her work is focuses on high dimensional clustering and classification methodology. Recent projects include developing record linkage algorithms and frameworks that incorporate atypical features such as network structure across records. Another of her primary application interests is differentiating learning strategies and trajectories for online courses. She is very active in statistics and data science education with extensive experience in program and curriculum development. A few of her many honors include the American Statistical Association Waller Education Award for innovation in statistics education, and the William H. and Frances S. Ryan Award for Meritorious Teaching from Carnegie Mellon University. Dr. Nugent received her PhD in Statistics from the University of Washington.

Patrick Riley is a Principal Engineer at Google. He and his team collaborates with external scientists to apply Google's knowledge and experience in machine learning and other computational approaches to important scientific problems. His work includes applications in drug discovery, quantum chemistry, materials science, and nuclear fusion. Dr. Riley received his PhD from Carnegie Mellon University in Artificial Intelligence in Multi-Agent Systems.

231564651354123142654123564132145795614253201230369700101024465789541
203354302326457669546231042537694852701364589754623021346597885400036
326154978012364785946545210034262154879856540012487956413425316224579
864312564372154659784562312465579832659551242754935829600200134270132
975201375219870161758064137698505247325316325978150243597810245687653

Planning Committee Biographies

Michelle Chang is an Associate Professor at the University of California, Berkeley in the Departments of Chemistry and Molecular and Cell Biology. Her research group works at the interface of enzymology and synthetic biology and studies biological fluorine chemistry, formation of mixed-valent nanomaterials by directional-sensing bacteria, and processes involved in developing synthetic biofuel pathways. She has received many awards, including the Dreyfus New Faculty Award, TR35 Award, Beckman Young Investigator Award, NSF CAREER Award, Agilent Early Career Award, NIH New Innovator Award, DARPA Young Faculty Award, Camille Dreyfus Teacher-Scholar Award, 3M Young Faculty Award, Arthur C. Cope Scholar Award, and Pfizer Award in Enzyme Chemistry. Dr. Chang received her PhD in chemistry from the Massachusetts Institute of Technology and did postdoctoral training in the Department of Chemical and Biochemical Engineering at University of California, Berkeley.

Leo Chiang is Associate Technology Director at The Dow Chemical Company and leads the Chemometrics and Big Data Analytics Implementations for Manufacturing. He has developed and implemented several data analytics techniques to solve complex manufacturing problems; his work has resulted in 11 Dow Manufacturing Technology Center Awards. In 2016, he received the Dow R&D Excellence in Science Award in recognition of his scientific achievement in industrial research. Dr. Chiang is active in the American Institute of Chemical Engineers (AIChE) and served as 2014-2016 Computing and Systems Technology (CAST) director, the 2016 CAST 10E programming chair, and the 2017-2018 AIChE spring meeting program chair. He was recognized by the AIChE with the 2016 Herbert Epstein Award for his leadership on Big Data Analytics technical programming and the 2016 Computing Practice Award for his leadership in the development and application of methods in analytics for batch and continuous processes. Dr. Chiang has contributed to 20 externally refereed papers and 12 proceedings papers, given over 70 conference presentations and university lectures, and co-authored two books. Dr. Chiang received a PhD from the University of Illinois at Urbana-Champaign in chemical engineering.

Bruce Garrett is the Director of the Chemical Sciences, Geosciences, and Biosciences Division in the Office of Basic Energy Sciences (BES) of the Department of Energy (DOE). His expertise is in theoretical and computational chemistry with a focus on reaction rate theory and its application to gas and condensed phase chemical reactions, gas-to-particle nucleation kinetics, and mass transfer across interfaces. He leads a team that is responsible for managing a broad portfolio of experimental, theoretical, and computational research to provide fundamental understanding of chemical transformations and energy flow in systems relevant to DOE missions. Before joining BES, he was Chief Scientist for Chemical Sciences at Pacific Northwest National Laboratory (PNNL) where he managed the Physical Sciences Division and served as the point of contact for the PNNL BES Chemical Sciences, Geosciences, and Biosciences programs. His research has produced more than 220 publications, and recognition of his scientific contributions includes being named Fellow of the American Physical Society, the American Association for the Advancement of Science, and the Royal Society of Chemistry. Dr. Garrett received a PhD in chemistry from the University of California, Berkeley, and was a postdoctoral specialist at the University of Minnesota.

Carlos Gonzalez is Chief of the Chemical Sciences Division at the National Institute of Standards and Technology (NIST). He leads a team that provides measurement science, standards, technology, and data that enable scientists to advance chemical sciences in various research areas, including human health assessment, food science, and exposure science. He is a co-author on numerous publications that focus on various analytical and computational approaches and applications. Dr. Gonzalez received his PhD in chemistry from Wayne State University and was a postdoctoral scholar at Carnegie Melon University.

John Gregoire leads the high-throughput experimentation group at Caltech University and is the Thrust Coordinator for Photoelectrocatalysis in the Department of Energy Joint Center for Artificial Photosynthesis and an Associate Director in the National Science Foundation Computational Sustainability Network. He is an expert in the discovery of energy-related materials, particularly solar fuels materials and electrocatalysts for fuel synthesis and conversion, and he is a strong advocate for further adoption of artificial intelligence in materials research. Dr. Gregoire received his PhD at Cornell University and conducted postdoctoral research at Harvard University in the fields of computer science, mathematics, physics, and materials science.

Angela K. Wilson is the Acting Director of the Division of Chemistry at the National Science Foundation. She is also the John A. Hannah Distinguished Professor of Physical, Theoretical, and Computational Chemistry at Michigan State University (MSU). At MSU, her research group focuses on developing and applying computational methods in such research areas as heavy element chemistry, drug design, disease mechanisms, catalysis, and green chemistry. Her group's efforts are primarily focused on the development of ab initio approaches that aim for accurate prediction of thermochemical properties across the periodic table. Prior to MSU, Dr. Wilson was the Regents Professor of Chemistry at the University of North Texas. Dr. Wilson received her PhD in chemical physics from the University of Minnesota.



CHEMICAL SCIENCES ROUNDTABLE