Modeling Science, Technology and Innovation

NSF Conference Report

Scientific Organizers:
Katy Börner and Staša Milojević
Department of Information and Library Science
School of Informatics and Computing
IU Network Science Institute (IUNI)
Indiana University, Bloomington, IN 47405

Scientific Advisors:
James Evans, Associate Professor of Sociology, The University of Chicago
Susan Fitzpatrick, President, James S. McDonnell Foundation
Richard B. Freeman, Herbert Ascherman Chair in Economics, Harvard University
Jerome Glenn, CEO, The Millennium Project
Jason Owen-Smith, Professor of Sociology, University of Michigan
Caroline Wagner, Ambassador Milton A. and Roslyn Z. Wolf Chair in International Affairs & Director, Battelle Center for Science and Technology Policy, Ohio State University
Uri Wilensky, Professor of Learning Sciences, Computer Science, and Complex Systems, Northwestern University

Local Organizer:
Daniel O'Donnell
Cyberinfrastructure for Network Science Center
Indiana University, Bloomington, IN

Venue:
National Academy of Sciences in Washington, D.C.

Dates:
May 17-18, 2016

Acknowledgements

The organizers would like to thank Patsy Powell with the National Academies for providing an excellent space for this conference to take place; Dawn Pass, Sodexo catering for providing every meal on time; Sara Bouchard, CNS Senior Graphic Designer for designing all printed materials and the website; Jill Taylor, CNS Center Assistant for taking care of travel reimbursements; the team from TV Worldwide for recording the conference; Rebecca Reeseman and Layla Hashemi for serving as rapporteurs for the panels; all conference advisors, presenters, and participants for their expert contributions. This conference was supported in part by the National Science Foundation's Science of Science and Innovation Policy program under Grant No. SBE-1546824, NETE Federal IT, Thomson Reuters, Indiana University Network Science Institute, and the Cyberinfrastructure for Network Science Center at Indiana University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Cover Image Credits

Photo of conference participants by Andreas Bueckle, Indiana University

Executive Summary

In a knowledge-based economy, science and technology are omnipresent and their importance is undisputed. Equally evident is the need to allocate resources, both monetary and human, in an effective way to foster innovation (Ahrweiler et al., 2015; Watts & Nigel, 2014). In the preceding decades, science policy has embraced data mining and metrics to gain insights into the structure and evolution of science and to devise metrics and indicators (Hicks et al., 2015), but it has not invested significant efforts into mathematical, statistical, and computational models that can predict future developments in science, technology, and innovation (STI). STI models make it possible to simulate the diffusion of ideas and experts, to estimate the impact of population explosion and aging, to explore alternative funding schemas, or to communicate the probable outcomes of different policy decisions.

Advances in computational power combined with the unprecedented volume and variety of data on science and technology developments (e.g., publications, patents, funding, clinical trials, stock market, social media data) create ideal conditions for the advancement of computational modeling approaches that can be empirically validated and used to simulate and understand the structure and dynamics of STI and to augment human decision making.

An NSF-funded conference "Modelling Science, Technology, and Innovation" was held at the National Academy of Sciences Building in Washington DC in May 2016. More than 100 participants from 67 institutions and seven countries attended the conference; 62 experts from academia, government, and industry presented their work. This report discusses key conference findings regarding challenges, insights, and opportunities associated with the usage of mathematical, statistical, and computational STI models.

Key challenges, insights, and opportunities identified during the conference are discussed in this report and can be summarized as follows:

Key Challenges

Challenges exist in five areas:

Fundamental Research

- Few models are validated using empirical data.
- STI success criteria are not clearly defined, e.g., when is science successful?
- Few agencies support STI modelling efforts.
- Federating and cleaning existing open or proprietary data requires resources.

Applied Research

- Most scholars design models to publish; few build policy-relevant models.
- Most policy makers do not engage in modeling efforts yet their expert input is needed to design models that make a difference.
- Few models are production-strength, i.e., validated, well documented, 24/7 service.
- Few active partnerships among academia, government, and industry exist.

Cyberinfrastructure

- Unconnected silos of data and code repositories exist in different areas of science.
- Scholarly results are published in many different journals and are hard to find. Few special issues, reviews, general textbooks exist.
- ModSTI scholars and practitioners attend many different conferences, might not know about each other.

Education

- There is a need for improved “model literacy” via formal and informal education combined with more effective communication by researchers on the power of models.
- A community of teachers/students that share data, code, results, training materials, etc. is missing.

Outreach

- Engagement with the public is very important. Scientists need to be more actively involved in communicating with the public and engaging with other communities.

Key Insights and Opportunities

Insights and opportunities can be grouped into five areas:

Model Needs and Implementation

There are diverse reasons why models are not used; among them are:

1) Model development process is not open, transparent, and inclusive enough to create and maintain buy-in from the very beginning with all stakeholders.
2) Discordance between what is needed by the decision-makers to make the decisions and what is given by the modelers.
3) Limitations of the input data (garbage in, garbage out).
4) Untenable assumptions.
5) Wrong outcome measures and metrics.
6) Models contain too much math.
7) Models are too complicated (tension between ease of use and the reality that most phenomena and mechanistic processes are complex and nonlinear and involve lags, dynamic feedback, etc.).
8) Reductionist vs. holistic thinking.
9) Problem space definition—either too narrow or too wide.
10) Tension for a single answer when in reality “it depends” almost always holds true. Context is (almost) everything.

Major takeaways

- Successful modeling teams require close collaboration and active partnership between (policy) decision makers and researchers to ensure the usefulness of the models and increase the chances for their adoption.
- Successful models require testing, iterative improvements, and a community of users.
- Models can inform decision making at the federal level but are also valuable at the regional, state, county, municipal, or institution level.

- Models can be used in “evaluation” exercises but are even more valuable when used in support of “situational awareness, proactive steering”.
- Simple models (e.g., simple data entry, simple easy-to-produce graphic output) are easier to implement in practice than complex models, which are very difficult to validate and the usefulness of which is difficult to determine.
- There is a need to identify and call out bad models since their existence and wide dissemination can harm the reputation of modeling efforts in general.

Data Infrastructure

- Data quality, coverage, and richness are improving rapidly and support the design and validation of detailed models.
- Many teams are spending 80% of overall STI modeling effort on data preparation. The setup of data repositories and joint data curation efforts should be explored to increase the amount of time available for model development. Furthermore, easy access to relevant data will support reproducibility.
- As many high-quality datasets are held by industry or government, close academia-industry-government partnerships seem desirable.

Code Repository and Standards

- Well-documented tools are needed that would allow decision makers to run their own models.
- Models need to be reliable and results have to be reproducible.
- STI modeling community should aim to adopt modeling guidelines (e.g., those developed by Interagency Modeling and Analysis Group (IMAG)[1]) and aim to create a shared data and model code infrastructure.

Visualization and Communication of Modeling Results

- It is important to communicate data quality, model complexity, and modeling results clearly to different stakeholders.
- Telling visual stories, augmented with high quality data, is powerful.
- More advanced visualizations of model results can be used to have decision makers “fly the future” before writing a check.

Funding

- Modelling needs increase with reduction of budgets, significant increases in the number of researchers, exponential growth of scientific productivity, larger team sizes, and higher interdisciplinarity.
- While few agencies and organizations have active funding programs on STI models, modeling of STI programs might be supported analogous to STI program evaluation.

[1] https://www.imagwiki.nibib.nih.gov

Introduction

This section motivates and discusses computational models of STI that were the focus of the conference, provides details on conference participant selection and engagement, and explains the structure of the remainder of this conference report.

Modelling Science, Technology, Innovation

Models of science, technology, and innovation (STI) aim to inform (policy) decision making in education, energy, healthcare, security and other sectors. The aim of these models is not to replace but rather to empower experts to make informed decisions when selecting reviewers, picking the best proposals for funding, or when making resource allocation decisions. They are a new kind of "macroscope tool" (de Rosnay, 1979) that can be used to derive key insights from big data in support of evidence-based policy. As Kevin Finneran, National Academies of Sciences, Engineering, and Medicine noted in his presentation: "If retail has figured out how to optimize sales by using models, then there is likely a market in government for practical decisions."

Some models are optimized to make recommendations, e.g., IBM Watson suggesting reviewers for a set of proposals, without much information on the type of match or the matching process. Other models aim to capture the true structure and dynamics of complex STI systems; they simulate the diffusion of ideas and experts, estimate the impact of population explosion and aging, or communicate the probable outcomes of different policy decisions. In sum, they help answer resource allocation or multi-faceted strategic questions. The latter models are often used in a team setting where small multi-disciplinary groups investigate and debate alternative futures together.

Computational models are already well established and widely used in a number of fields such as meteorology to predict weather and storms; epidemiology to predict and prevent pandemics; and climate to predict future scenarios and set carbon prices. Industry also extensively uses computational models to optimize operations, management, production, distribution, and marketing. Early adopters of data-driven decision making (most notably Target, Walmart, and Amazon) now dominate their sectors. Those who were slow to invest and then did so in isolated aspects of the organization (most notably Sears, Kmart, and Barnes & Noble) are heading towards bankruptcy.

Advances in computational power combined with the availability of relevant data (e.g., publications, patents, funding, clinical trials, stock market, and social media) create ideal conditions for the implementation of computational modeling approaches that can be empirically validated and used to simulate and understand future developments within STI and to pick desirable futures. Interactive data visualizations that show probable futures in response to different (policy) decisions or external events can help stakeholders discuss model assumptions, design, and output. Ideally, stakeholders get to "drive the future before it is implemented" (Rouse, 2014, 2015); they can quickly explore different policy options and discard those that lead to unexpected, undesired consequences (Watts & Gilbert, 2014; Ahrweiler et al, 2015). However, designing effective interfaces that let different stakeholders communicate and explore different scenarios is non-trivial.

C.D. (Dan) Mote, Jr., President of the National Academy of Engineering, in his opening remarks rightly pointed out that more data and more advanced models are frequently the wrong answers to the right questions. He argued that it is important to ask and answer: "Are we modeling the system correctly?", and even more importantly, "Are we modeling the right system?" Mote provided an illustrative story of an effort that brought running water to a remote village only to learn later that the women preferred to walk (a mile and a half, to and from the village, to get water, and haul it back, several times a day) as this was their only chance to get away from the men and to discuss candidly what was going on in the village. More data on water, weather, or cost of operation and maintenance would not have led to a better solution to this problem. Bringing the water nearer, but outside of the village, might have been more useful. However, such a solution required understanding of local culture and not just input/output measures. This story highlights the importance of humanities and social sciences expertise for the development and implementation of models that make a positive difference in the world. Without adequately accounting for human behavior and human factors models will be rather limited. Data, computing, and visualization tools—no matter how advanced they might be—cannot compensate for inadequate model assumptions.

Conference Participant Selection and Engagement

Conference organizers Katy Börner and Staša Milojević (School of Informatics and Computing, Indiana University Bloomington) worked closely with the conference Advisory Board members (James Evans, Associate Professor of Sociology, The University of Chicago; Susan Fitzpatrick, President, James S. McDonnell Foundation; Richard B. Freeman, Herbert Ascherman Chair in Economics, Harvard University; Jerome Glenn, CEO, The Millenium Project; Jason Owen-Smith, Professor of Sociology, Barger Leadership Institute Professor of Organizational Studies, University of Michigan; Caroline Wagner, Ambassador Milton A. and Roslyn Z. Wolf Chair in International Affairs Director, Battelle Center for Science and Technology Policy, Ohio State; and Uri Wilensky, Professor of Learning Sciences, Computer Science, and Complex Systems, Northwestern University) to identify 30 potential conference participants. The selection process was guided by the principles of inclusiveness with the goal to identify experts with a wide range of disciplinary backgrounds engaged in current STI modeling efforts and policy makers. Efforts were also made to achieve gender diversity. Due to the large interest in conference, organizers decided to open up the participation and allow up to 100 additional experts to attend and present talks. The experts who were interested in presenting their models submitted abstracts which were reviewed by the conference organizers. The final agenda included: remarks by C.D. (Dan) Mote, Jr., President of the National Academy of Engineering and E. William Coglazier, AAA Center for Science Diplomacy; two keynotes by Alex "Sandy" Pentland, MIT and Dame Wendy Hall, University of Southampton, UK; four panel sessions with 16 presenters (presenting case studies; discussing (a) funding opportunities, (b) data, algorithms, infrastructure, and (c) policy issues); four talks sessions showcasing models of science, innovation, and technology with 14 presenters; and two 2-minute flash talks sessions with 25 presenters (Full Conference Agenda is available in Appendix B).

The conference featured talks, panels, tours, and extensive discussions during the reception and breaks, see Figure 1.

Figure 1: (Top-left) John Walsh and Richard Freeman in the Models of Innovation session. (Top-right) Ben Shneiderman asking questions from the audience. (Bottom-left) Participants exploring complex systems Living Architectures. (Bottom-right) Keynote by Alex (Sandy) Pentland.

Twitter was used as a back channel by more than 100 participants to communicate conference insights. The network of tweets between Sunday, 15 May 2016 at 22:38 UTC to Tuesday, 24 May 2016 at 01:14 UTC was plotted using NodeXL and can be seen in Figure 2. The network comprises 102 Twitter user nodes whose tweets contained "ModSTI", or who were replied to or mentioned in those tweets. The graph contains 195 directed edges of three types: "replies-to" relationships, "mentions" relationship in a tweet, and self-loops. The graph's nodes were grouped using the Clauset-Newman-Moore cluster algorithm; the graph was laid out using the Harel-Koren Fast Multiscale layout algorithm, for details see http://bit.ly/22mO5v4.

As can be seen, there are six major communities (named G1 to G6, with nodes that are color coded) that are weakly connected with each other. The G1 cluster in the lower left—with dark blue labels for Twitter handles—has two very active nodes: ct_bergstrom (Carl Bergstrom, one of conference speakers) and eigenfactor (See http://www.eigenfactor.org). The G2 cluster in the top left—with light blue labels—features harayasushi (one of conference speakers) and garybcross (one of conference attendees) but also nsf, openscience, etc. Many of the active nodes connect to each other—bridging cluster boundaries—to interconnect conference organizers, e.g., katycns (Katy Börner), advisors, speakers, and others. This clustering by scientific discipline, geography, and institution type is typical for MODSTI research and practice as are the synergistic linkages across cluster boundaries.

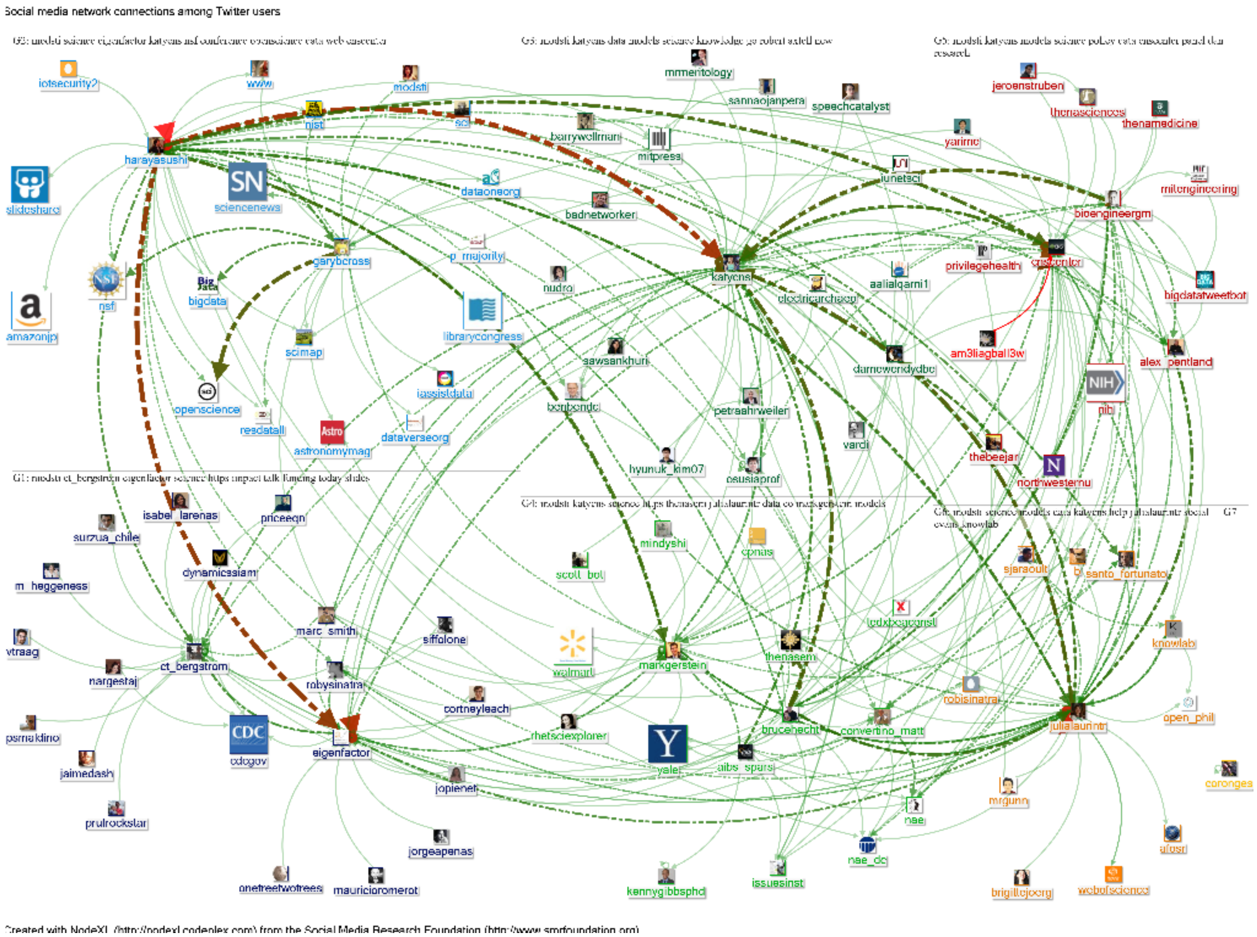

Figure 2: Twitter network of users communicating conference events and using #ModSTI. Top influencers are @katycns, @harayasushi, @ct_bergstrom, @garybcross, @julialaurintr, @cnscenter, @eigenfactor, @markgerstein, see http://bit.ly/22mO5v4 for details.

Report Structure

The remainder of the report is structured as follows. The next section provides an overview of exemplary mathematical, statistical, and computational models that can be used to plan and forecast STI, a listing of major model types is presented as well as a discussion of three exemplary STI models. Next, we present the significant challenges associated with the specification, development, and implementation of models that were identified during the conference. Subsequently, we discuss STI modelling opportunities that were presented at the conference by various speakers and participants. The report concludes with an outlook for computational models of science, technology, and innovation, including the development of a repository of STI models among others. The appendix contains information on the project team—including organizers, scientific advisors, and rapporteurs—as well as conference agenda, abstracts for all the presentations, and biographies of all presenters.

ModSTI Examples

In 2007, *Issues in Science and Technology* published "The Promise of Data-Driven Policymaking" by Daniel Esty & Reece Rushing. In 2016, the same magazine published "Data-Driven Science Policy" (Börner, 2016). The articles point out that in the corporate sector, a wide variety of data-driven approaches are used to boost profits, including systems to improve performance and reliability, evaluate the success of advertising campaigns, and determine optimal pricing. They argue for the need and discuss the premise of data-driven decision making in STI (including policy making)—using large-scale, high-quality datasets and computational means to inform human decision makers.

During the two-day conference, participants presented and discussed a wide range of mathematical, statistical, and computational models that were developed and implemented in a variety of settings (see Figure 3 for an overview of key terms such as data, research, new, impact, models, and systems). While the emphasis was on the power of models to advance future decisions, the presentations and discussions emphasized the usefulness of models to simulate, explain, and communicate the past, present, and future.

Figure 3. Word frequency computed from all the abstracts of all the presentations at the conference.

STI Methods and Tools

The book *Models of Science Dynamics* published in 2012 provides a unique review of major model classes (from population dynamics models to complex network models) accessible to science policy researchers and practitioners. The conference built on the book, but aimed to create a more up-to-date inventory of exemplary mathematical, statistical, and computational models that can be used to forecast STI in different settings. Table 1 lists the methods and tools presented at the conference.

Table 1. List of methods and tools for different settings presented at the conference

METHODS	SETTINGS	TOOLS	PRESENTER

Machine learning	Health (analysis of clinical trials)	IBM Watson (http://www.ibm.com/watson)	Richard Ikeda, NIH
Text mining	Health (analysis of clinical trials)	IBM Watson (http://www.ibm.com/watson)	Richard Ikeda, NIH
Natural language processing	Health (analysis of clinical trials)	IBM Watson (http://www.ibm.com/watson)	Richard Ikeda, NIH
	Cognitive extent of science		Staša Milojević, Indiana University
Systems analysis	Health (global health planning – vaccines)	SMART Vaccines (http://www.nap.edu/smartvaccines)	Guru Madhavan, National Academies of Sciences, Engineering, and Medicine
	Health economics (modeling projections of cases with and without interventions)	CDC FLUAID Special Edition Software to aid State and Local Planners with Estimating the State Level Impact of 2009 N1H1 Influenza (http://www.cdc.gov/h1n1flu/tools)	Martin Meltzer, CDC
Multi-model system	Climate (prediction of climate extremes)		Venkatachalam "Ram" Ramaswamy, NOAA
Regression models	Economics of innovation (role of non R&D innovation)		John P. Walsh, Georgia Tech
	Science to business knowledge bridges		Lynne G. Zucker, UCLA
	Intellectual lineages (reproductive success of scholars)		Daniel McFarland, Stanford University
	Measuring innovation (hedonic pricing and innovation index for products)		Richard B. Freeman, Harvard University
Policy flight simulators	Evidence-based decision making		William Rouse, The Stevens Institute of Technology
Agent-based simulations	Collective allocation of science funding		Katy Börner, Indiana University
	Evaluating Horizon 2020 Policy Interventions	SKIN model (http://cress.soc.surrey.ac.uk/skinwp)	Petra Ahrweiler, EA of Technology and Innovation Assessment GmbH, Germany
	Emergence of research teams		Staša Milojević, Indiana University
	The evolution of economic goods and services		Robert Axtell, George Mason University

Event sequence analytics	Health (Innovation trajectories)	EventFlow (http://hcil.umd.edu/eventflow)	Ben Shneiderman, University of Maryland
Network models	The fitness of ideas in science and inventions		Brian Uzzi, Northwestern University
	Citation dynamics		Santo Fortunato, Aalto University, Finland
	Reasoning pathways in science		James Evans, University of Chicago
	Publication patterns	PubNet (http://pubnet.gersteinlab.org)	Mark Gerstein, Yale University
Empirically guided mathematical models	Individual careers (Q-model)		Robert Sinatra, CNS, Central European University, Hungary
	Attention decay in science		Santo Fortunato, Aalto University, Finland
	Citation inflation		Santo Fortunato, Aalto University, Finland
Game theory	Why scientists chase big problems		Carl Bergstrom, University of Washington

STI Models Presented at the Conference

Many different models were presented at the conference (see slides and video files of presentations linked from http://modsti.cns.iu.edu). The presentations of STI models at the conference were grouped by the types of phenomenon they aim to capture. Below we provide brief descriptions of eleven and longer discussion of three exemplary STI models presented during four 15-minute talks sessions.

Models of Science

Daniel McFarland used dissertation data to model intellectual lineages via "the system of faculty reproduction" showing how the dynamics of faculty reproduction differs over time and across disciplines.

Brian Uzzi modeled knowledge networks to enhance our understanding of scientific breakthroughs. He found that the breakthrough paper rate is narrowing and scientists take longer to make their first discoveries. He also found a link between the age of information and scientific discovery.

Santo Fortunato examined the consequences of "publish or perish" climate in academia, which has led to exponential growth of papers. He showed that such a growth in turn has led to attention decay and inflation of citations and argued that it should be taken into account in models of science dynamics.

Roberta Sinatra presented a model that predicts and quantifies patterns of individual scientific impact. She found that successful scientific careers look different, i.e., they do not follow a singular pattern and are the result of luck combined with quality.

Carl Bergstrom used a framework he calls "new economics of science" to develop a model that examines how the incentives created by contemporary scientific institutions lead scientists to allocate research efforts across problems. He cautioned that by allowing for sharing of partial results, Open Science can slow down the solution of a particular problem by deterring entry of important actors.

James Evans used a hypergraph of millions of scientific articles from MEDLINE to model "how science thinks." He found that the contemporary science has become risk-averse, less efficient at discovery, and more detached from real health problems.

Mark Gerstein discussed the importance of mapping collaboration in large distributed projects in order to facilitate knowledge creation and diffusion.

Models of Technology

Robert Axtell presented a model of the emergence of technological epochs and policies that foment them. He modeled the evolution of economic goods and services as a stochastic process of recombination conducted by purposive agents and presented the resulting technological lineages.

Models of Innovation

Ben Shneiderman showcased the power of event sequence analytics and EventFlow tool using as an example model of innovation trajectories in health domain.

John Walsh, based on his model of the economics of innovation, cautioned that by outsourcing production the U.S. might also be outsourcing innovation given that non-negligible innovation happens outside R&D departments of companies.

Lynne Zucker proposed a novel idea that knowledge does not travel via diffusion, but rather through learning. She emphasized the importance of collaboration between academia and industry that results in increasing publication leading to "disclosed science."

STI Models Detailed

Below we discuss three STI models that were presented at the conference in more detail. Example 1 describes how teams in various fields have evolved over time and what it is they contribute to contemporary science. Example 2 proposes radical changes to the current funding system. Both of these models were empirically validated and a high correlation was found between the simulated datasets and the structures and dynamics found in publication and funding data. Example 3 presents innovative ways to communicate the implications of policy decisions to policy makers before any policy is implemented.

Example 1: The Importance of Small Teams in the Big Science Era

Contemporary science is a collaborative effort within an intricate network of people, institutions, concepts, and technology. Many projects are of such complexity or scope that they require joint efforts of many individuals with diverse expertise, reaching team sizes of few hundreds. Furthermore, studies suggest that large interdisciplinary teams are more likely to produce high impact work.

Only 50 years ago, the situation was very different. Most papers were written by single authors and the largest co-author teams did not exceed ten members. How did this change in the production of knowledge occur? How do science teams form and what processes lead to their expansion? What makes a successful team?

Research team size distribution lies at the heart of our understanding of collaborative practices and research productivity. As Figure 1 shows, knowledge production today is qualitatively different from that of earlier times: “little science” performed by individuals or small groups of researchers is to a large part superseded by “big science” efforts by large teams that span disciplinary, institutional, and national boundaries.

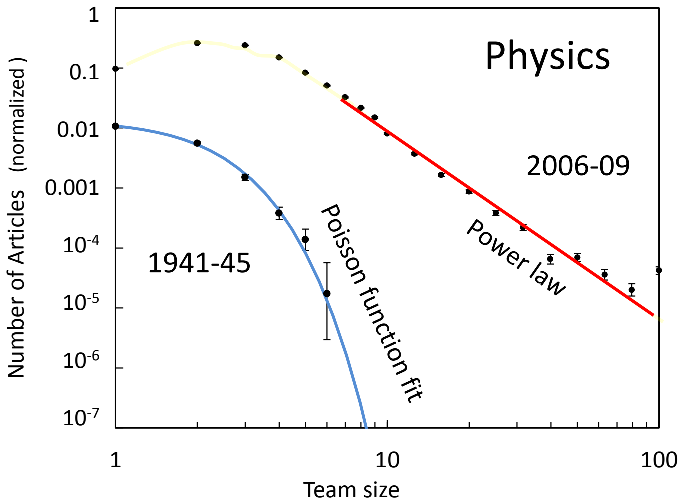

Figure 4. Change in the distribution of research team sizes in physics from a Poisson distribution to one dominated by a fat tail (a power law). In 1941-1945, for each paper with five authors, there were one thousand single-authored papers (blue). In 2006-2009, there are as many papers with five authors as there are single authored papers (red), and very large teams are not uncommon.

Staša Milojević at Indiana University developed a model of how teams emerge and grow, which accurately reproduces the change seen in Figure 4, and predicts how teams will evolve in the future. The model shows that team formation was, and remains, a Poisson process that results in relatively small, core teams (including single investigator teams) that are necessary to carry out certain types of research. The model also simulates the emergence of larger teams over the last 50 years in all fields of science albeit with varying pace and magnitude of change. According to the model, each big team originates from a small team; while some small teams do not change in size, others quickly accumulate additional members proportionally to the past productivity of team members developing into big teams. Surprisingly, the model shows that relatively small teams dominate knowledge production in most fields, so that cumulatively they still contribute more new knowledge than large teams. These findings are of key importance to policy, because they show that increased funding emphasis on large teams may undermine the very process by which large, successful teams emerge.

Example 2: Crowdsourcing Funding Allocation

Johan Bollen and colleagues at Indiana University argue that scholars “invest an extraordinary amount of time, energy and effort into the writing and reviewing of research proposals,” and that funding agencies are consuming resources that could be more productively used to conduct and finance research. In a 2014 paper, they use NSF and Taulbee Survey data to provide a simple calculation of return on investment for scholars in computer science. The calculation quickly reveals that the return on investment is negative: Four professors working four weeks full-time on a proposal submission at labor costs of about $35,000; given a CISE funding rate of 21% about five submission-review cycles might be needed, resulting in a total expected labor cost of $175,000. The average NSF grant is $164,526 per year of which U.S. universities charge about 50% overhead leaving about $109,684 for research. That is, the four professors lose $175,000-$109,684 =$65,316 of paid research time by obtaining a grant and U.S. universities might like to forbid professors to apply for grants—if they can afford to forgo the indirect dollars. Note that this simple calculation does not cover any time spent by scholars to review proposals. In 2015 alone, NSF conducted 231,000 proposal reviews to evaluate 49,600 proposals.

Bollen et al. then go on to propose a FundRank model to (partially) replace the current process of government research funding allocation by expert-based crowdsourcing. In the new system, each eligible scholar (e.g., all eligible to submit NSF and NIH grants today) receives a certain dollar amount each year, let us say $100,000. S/he then needs to give a certain fraction, e.g., 50%, to colleagues that are most deserving by logging into a centralized website and entering names and amounts. That is, scholars collectively assess each other's merit and they "fund-rank" other scholars, with highly ranking scholars receiving the most funding.

Instead of spending weeks writing and reviewing proposals, scholars are now incentivized to spend time communicating the value and impact of their past, current, and planned work so that others can judge their contributions. Using a fully digital system, conflicts of interest can be easily identified and honored; networks of mutual favors can be detected automatically, and results shared publicly.

FundRank was implemented using the recursive PageRank algorithm pioneered by Sergey Brin and Larry Page in 1998. Using PageRank, the "importance" (here reputation, value, impact) of a scholar depends not only on the number of scholars that vote for him/her but also their importance. The more that important scholars link to a person, the more important the person must be. The FundRank model was validated using citation data from 37 million papers over 20 years as a proxy for how each scientist might distribute funds in the proposed system. Simulation results show funding patterns that have a similar distribution compared to NSF and NIH funding for the past decade—at a fraction of the cost required by the current system.

Example 3: Explore the Future Before Writing a Check

Policy decision makers need to understand and trust modelling results or they will not use them in practice. Visualizations of the modelling process and modelling results have proven invaluable for affording a strong intuitive feel for the predictions and insights models provide. William Rouse's team at the Stevens Institute of Technology has been working closely with the National Academies of Engineering and Medicine to implement so called "policy flight simulators" that let decision makers fly the future before they write the check. In his book "Modeling and Visualization of Complex Systems and Enterprises: Explorations of Physical, Human, Economic, and Social Phenomena," Rouse details existing models, associated visualizations, and the utility of policy flight simulators for enabling evidence-based decision making.

Figure 6 shows key decision makers—senior executives—in front of a large tiled wall filled with evidence that shows how different strategies will likely affect merger and acquisition scenarios among New York City's 66 hospital corporations that are driven mostly by the elements of the Affordable Care Act. During the immersive and active session—that cannot be properly reproduced by a still image—decision makers and data experts argue, simulate, and debate competing perspectives and possible compromises. Using the simulator, they were surprised to find that major player's strategies, relative to their primary competitors, very strongly affect the "pecking order" resulting over the coming decade.

For example, one of the leading hospitals that has been in the top one or two for a long time will need to carefully consider competitors' strategies or, much to their surprise, they could fall out of the top five. After the simulators experience, the hospital leadership paid much closer attention to competitor actions.

Figure 6: Healthcare Delivery Ecosystem of New York City

Another policy flight simulator session focused on adoption of automobile powertrain technologies, comparing internal combustion, hybrid, electric, and hydrogen systems. As expected, modelling results show an increased adoption of electric vehicles as both the federal subsidies and federal and state investments in battery charging infrastructure increase. This has long been true for federal subsidies of railroads, aviation, nuclear power, and ethanol. Surprising, however, is the impact of (Corporate Average Fuel Economy) standards. With fuel costs very low, Americans are buying more pickup trucks and large SUVs, which can have profit margins approaching $10,000 per vehicle. In order to secure these profits, while also meeting the CAFE standards, manufacturers have to incentivize sales of economy cars, often losing $2,000 per vehicle. As economy cars become cheaper, they take away sales from hybrid and electric vehicles, undercutting government incentives.

ModSTI Challenges

The conference aimed to identify the diverse challenges associated with the usage of mathematical, statistical, and computational models in STI in decision making. Here, we discuss key challenges that designers and users of computational models face in five aspects of modeling STI: fundamental research, applied research, cyberinfrastructure, education, and outreach. Note that many of these challenges can be phrased as opportunities.

Fundamental Research

Research on STI is carried out by researchers in a wide range of disciplines: economics, social science, information science, science policy, scientometrics/bibliometrics, physics, and science policy among others that develop mathematical, statistical, and computational models of different types (stochastic, agent-based, epidemics, game-theoretic, network. etc.). One of the impeding factors in moving forward fundamental research is freely available access to good quality data that will reduce data curation efforts currently done by each individual team and to allow reproducibility (one of the most-wanted traits of models as identified by conference participants, see discussion of Data and Code in Subsection "Opportunities"). Lack of obvious continuous sources for funding for this type of research was identified as an additional challenge.

In addition, researchers who do STI modeling publish in a wide range of venues, addressing different audiences. As became evident at the conference, current research efforts and results are not universally known to the researchers (let alone policy makers). Such a state slows the scientific progress and can possibly lead to unnecessary "reinvention of the wheel." Conference participants genuinely enjoyed being in a truly intellectually diverse environment, which helped them shed new light to the problems they were grappling with, but also forced them to think and talk about their own research in a new way. There was a sense that similar events in the future would help greatly advance the fundamental research efforts (see listing of upcoming conferences in Section "ModSTI Outlook").

When discussing his views on policy-relevant research, David Goroff emphasized the importance of posing good questions, rather than focusing on outcomes. He also called for moving from descriptive to normative theories. One of the major research challenges is the development of multiscale models—covering the micro (individual) to macro (population) levels—and understanding the appropriateness of particular models for particular scales. STI modelling experts should aim to learn from other sciences that use systems-science approaches (see depiction of multiscale modeling in biomedical research in Figure 7).

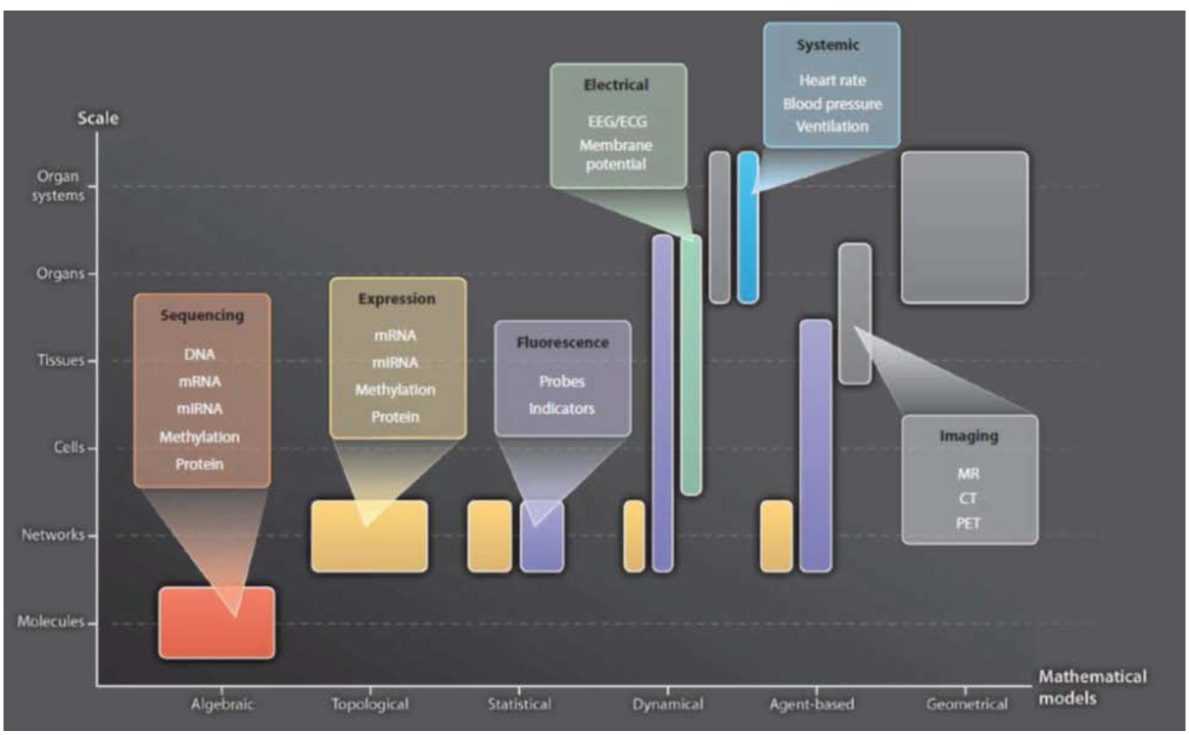

Figure 7. Multiscale modeling in biomedical research. Credit: Raimond L. Winslow, *Computational Medicine: Translating Models to Clinical Care.*

Applied Research

STI models are also developed within different government institutions and agencies and often lack wider exposure. The “Case studies” presentations were especially helpful in providing insights into the possibilities and challenges of carrying out applied research using modeling. Guru Madhavan, for example, emphasized the importance of systems analysis approaches. He reiterated the importance of taking into account cultural factors mentioned in the opening remarks by Dan Mote. He also emphasized the importance of building tools, such as the one developed by his group for the measurement of the importance of vaccines, SMART Vaccines, which are used by decision makers rather than model builders. There was a general agreement that there is often poor communication between model builders and users/stakeholders, at all stages from the initial design (what question is being asked, what assumptions are being made, what measures and metrics to use, etc.) to the interpretation and application of results to the real-world problem. This is further exasperated by a less than open and transparent modeling process that does not create and maintain buy-in from the very beginning of a project.

Cyberinfrastructure

Cyperinfrastructure, e.g., data and model repositories but also computing and visualization infrastructures, are highly beneficial for advancing STI modeling efforts. Many sciences have setup billion-dollar international data infrastructure and distributed computing systems in close collaboration with government and industry partners. Examples are meteorology (e.g., weather forecasts, including hurricane and tornado prediction), epidemiology (e.g., predicting the next pandemic and identifying best intervention strategies), climate research (e.g., to predict future scenarios or to set carbon prices), or financial engineering (e.g., stock trading and pricing predictions). No such infrastructure exists yet for the study and management of STI, yet 80 percent of project effort is commonly spent on data acquisition, cleaning, interlinkage, and preprocessing. Furthermore, modeling STI resources have been spent on individual project level, despite the experiences in natural sciences where building a general infrastructure of commonly used data available to all has led to major advances (e.g., climate studies, astronomy, etc.). Successful STI modeling requires validation, iterative improvement, and a community of users; all of which could be provided via appropriate infrastructure. However, building such an infrastructure will require active partnerships among academia, government, and industry. Sandy Pentland argued for the need to bring algorithms to data that are either too big for efficient sharing or have serious privacy and security issues.

Education

Education and training was discussed in a number of contexts. There were discussions regarding current “data literacy” and often expressed concern that it is rather low. Going forward, it will be important to introduce computational modeling and example models into formal and informal education. Participants also discussed the increased need for the active involvement of stakeholders into partnerships with modelers—to build simple models that can be understood more easily and validated to help stakeholders determine their usefulness. At the same time, many agreed that there is an urgent need for researchers and other model builders and users to enhance their communication and visualization skills.

Outreach

Modeling results need to be communicated to different stakeholders. Conference participants agreed that simple models and tools that are easy to understand and use and visualizations of model results that anyone can understand are key to the adoption and usage of models. Participants also emphasized the importance of storytelling and the art of communicating major results/recommendations in a clear and simple message. Kevin Finneran, National Academies of Sciences, Engineering, and Medicine, especially emphasized the importance of communication with non-scientists and provided excellent examples of how such communication can be achieved.

ModSTI Insights and Opportunities

Key insights gained from the conference presentations and discussions as well as rather timely opportunities for advancing R&D on and the implementation of STI models are presented here.

Several speakers shared their insights on why it is difficult to develop and use models of STI. An example is *REASONS MODELS ARE NOT (YET) USED* by Stephen Marcus, National Institutes of Health:

- Incomplete buy in of stakeholders; poor communication
- Limitations of input data (garbage in, garbage out)
- Untenable assumptions
- Wrong outcome measures and metrics
- Too much time needed
- Changing or new realities
- Discordance between info needed by decision maker and information provided by modeler
- Models contain too much math
- Models are too complicated, even with recent advances in visualization
- Unresolved tension between reductionist and holistic thinking and between simple and complex models
- Boundary space that was either too narrowly or too widely defined; unresolved tension between problem and solution focused thinking
- Desire for a single answer when in reality "it depends" almost always holds true
- Models lack "face validity"; failed efforts to replicate and validate model
- Unresolved differences in expectations about modeling process and what can be realistically achieved
- Model results that are unwelcome
- Summary – poor communication and failure to find enough common ground between modeler and user

and *WHAT ARE THE BARRIERS TO USE OF MODELS FOR DECISION SUPPORT* by Bill Valdez, Consultants International Group, Inc.:

- Absence of Funding for Data Collection/Modeling
- Lack of Career Staff Expertise with Models
- Complexity of Use
- Short Attention Span of Political Leadership
- Institutional Inertia
    - Expert Judgement

    - Budget/Procurement Processes
- Fear of Transparency
- Deep Distrust of Social/Economic Sciences

Martin Melzer, CDC and others pointed out the importance of making 100% sure one can answer the following questions:

- What is the question?
- Who is the audience?
- What is the appropriate perspective (health, society, academia)?
- What is the most appropriate way to communicate results?

before talking about what data should be used or which algorithms should be implemented.

Venkatachalam "Ram" Ramaswamy, National Oceanic and Atmospheric Administration added two more questions for addressing the policy input challenge:

- How do they want to know it?
- How should they respond to it?

Based on a detailed review of all presentations and prior work, we identified five major opportunity areas that look particularly important and promising: modeling needs and implementation, data, code, visualization, and funding. These five are detailed subsequently.

Modeling Needs and Implementation

Modeling research and development strongly depend on a detailed understanding of the problem at hand and the range of actions a decision maker can take. If the wrong problem is modeled (see cautionary tale by C.D. (Dan) Mote, Jr., President of the National Academy of Engineering, in Section "Introduction") or if suggested actions are infeasible (e.g., doubling the U.S. R&D funding budget) then model utility will be low.

Several speakers noted that there is a major difference between statistical significance and business relevance. Nachum Shacham, PayPal pointed out model costs using an example of "false positives" (unidentified malicious users that cost PayPal money) versus "false negatives" (valued customers with blocked accounts that cost PayPal reputation and might lead to bad press). He pointed out that nobody should trust the result of any one model but should take note if five different models predict the same result.

Kaye G. Husbands Fealing, Georgia Tech suggested using a model similar to MakerSpaces for the development of computational models that truly address the STI questions and policy issues. She argued for "speed dating" as a means to connect stakeholders and scientists/analysts, to look at collective problems (not just one-offs), and to develop a feasible and sustainable bridge of communication. Ultimately, experts need to work across disciplinary and institutional (academia, industry, government) boundaries to exploit synergies and to arrive at modeling results that are greater than the sum of their parts. Model developers (e.g., in academia and industry) should aim to "room in" with model users (policy and other decision makers).

Ben Shneiderman, University of Maryland argued for the need to combine basic and applied (contract) work, see *The New ABCs of Research. Achieving Breakthrough Collaborations* (Shneiderman, 2016).

Martin Meltzer, Centers for Disease Control and Prevention pointed out the high value of usable, simple models that answer real questions.

There was a major discussion of black-box models such as IBM Watson technology presented by Richard Ikeda, NIH versus models that help people understand the system as presented by Petra Ahrweiler, EA of Technology and Innovation Assessment GmbH, Germany, William Rouse, Stevens Institute of Technology and others.

Computational models will need to be vetted by experts and earn the trust of the scientific policy making community before many start using them in practice. The key to building trust is transparency and the engagement of all stakeholders in the design and application of STI models.

Bill Valdez, The Consultants International Group pointed out that the primary audience for STI models is the science policy community. This community has a coherent structure with rules and processes. For example, the science policy community can be viewed as a pyramid, with those occupying the top (primarily Hill staffers, OMB, OSTP, and federal government agency leaders) wielding the most power simply because they control budgets and make the day-to-day decisions about the direction of science policy. The next layer is the university community (including Association of American Universities (AAU)), Association of Public and Land-grant Universities (APL), which exerts some influence, albeit indirectly, on the top policy makers. The next layer would be the private sector, including the major foundations (Kaufmann, Gates), associations (American Association for the Advancement of Science (AAAS)), Chambers of Commerce, and corporations (Raytheon, Battelle) that care about science policy. Next layer is state/local governments and below that is the general public. Each of these stakeholder groups have their own interest in STI models and models should be developed with those needs in mind.

Valdez also noted that the different policy offices have different ability to absorb/implement models and there is a considerable resistance to the adoption of new tools and approaches in general. He pointed out that the Federal Government is the largest, most complex organization in the world, yet it is poorly understood and continues to use outdated decision support tools and processes. Models could be extremely useful when making resource allocation decisions, the promotion of agency missions, or crisis management. Systems dynamic modeling is considered the way to go, yet not much has changed over the last decade when these approaches were first suggested, see listing of key reasons in the beginning of this section.

In sum, the primary funder/sponsor of models should be the Federal government. A case must be made to federal policy makers that they need models in support of objective, data-driven decision making. Financial buy-in from the science policy community will likely lead to closer government-researcher collaborations in support of ever more useful and actionable models. By broadening current NSF, NIH and other funding for ModSTI R&D to many if not all of the 17 agencies with S&T programs would result in a substantial, dedicated modeling budget.

Data Infrastructure

A common theme across all the presentations was the importance of high-quality and-high coverage data for high quality modeling results. Currently, many teams are cleaning, interlinking, and processing the very same data, often in slightly different ways—foregoing the ability to replicate results across team sites. There was consensus that while having "big data" on science and technology dynamics is extremely important for answering certain questions, "more data" is not and should not be the answer to modeling questions. Going forward, data sharing and joint data curation efforts should be explored and the setup of data repositories seems desirable. James Onken talked about the ongoing effort within NIH for data integration via linked data. Richard B. Freeman emphasized the importance of using scraped information from websites such as Amazon and cell phone data in addition to traditional survey data and to get better measures of innovation in economic statistics. Also, while a large number of modelers are using unstructured data, Dame Wendy Hall and others have emphasized the importance of the creations of ontologies and structured data.

Given that many high-quality datasets are held by industry (e.g., Web of Science and Scopus publication data, LinkedIn expertise profile data, Twitter or Instagram data) it seems highly desirable to work closely with industry.

Code Repository and Standards

Equally important are efficient means to share STI model code. Some teams are using GitHub.com but STI models will be hard to find among millions of open source projects.

Many conference participants agreed the time is ripe to focus energy and resources on building cyberinfrastructure and research community that support systematic research and (tool) development efforts. Instead of creating a new repository, it seems beneficial to build on and extend (or interlink) existing model repositories. Model repositories are commonly created by academic researchers, government institutions, or publishers.

Academic repositories are typically associated with a tool, e.g.:

- Agent Modeling Platform (AMP) project provides "extensible frameworks and exemplary tools for representing, editing, generating, executing and visualizing agent-based models (ABMs) and any other domain requiring spatial, behavioral and functional features." (http://www.eclipse.org/amp).
- GAMA is a "modeling and simulation development environment for building spatially explicit agent-based simulations." (https://github.com/gama-platform)
- NetLogo is a "multi-agent programmable modeling environment. It is used by tens of thousands of students, teachers and researchers worldwide. It also powers HubNet participatory simulations." (http://ccl.northwestern.edu/netlogo)
- MASON is a "fast discrete-event multi-agent simulation library core in Java, designed to be the foundation for large custom-purpose Java simulations, and also to provide more than enough functionality for many lightweight simulation needs. MASON contains both a model library and an optional suite of visualization tools in 2D and 3D." (http://cs.gmu.edu/~eclab/projects/mason)
- The Repast Suite is a "family of advanced, free, and open source agent-based modeling and simulation platforms that have collectively been under continuous development for over 15 years." (http://repast.sourceforge.net)

They might also be created for specific research projects:

- For example, the SIMIAN project is funded by the Economic and Social Research Council to promote and develop social simulation in the UK. SIMIAN uses the SKIN model (Ahrweiler, Pyka, & Gilbert, 2004). (https://github.com/InnovationNetworks/skin).

Modeling efforts are also supported by scholarly societies:

- OpenABM is a “node in the CoMSES Network, providing a growing collection of tutorials and FAQs on agent-based modeling.” (https://www.openabm.org)

Government institutions aim to support sharing of datasets or tools. NSF’s SciSIP program maintains a listing of “Datasets, Graphics & Tools” pertinent to the Science of Science Policy (SOSP) community at http://www.scienceofsciencepolicy.net/datasets_tools.

Interagency Modeling and Analysis Group (IMAG)[2] and the Multiscale Modeling Consortium aim to grow the field of multiscale modeling in biomedical, biological and behavioral systems, to promote model sharing and the development of reusable multiscale models, and disseminate the models and insights arrived from the models to the larger biomedical, biological, and behavioral research community, among others. The Predictive Model Index lists over 100 reusable, sharable models in support of reproducible science, see presentation by Grace Peng, NIH.

The Centers for Disease Control and Prevention (CDC) made the “H1N1 Flu (Swine Flu): Preparedness Tools for Professionals” software available at http://www.cdc.gov/h1n1flu/tools. The page was developed during the 2009-2010 H1N1 pandemic, it has not been updated, and is being archived for historic and reference purposes only.

Publishers aim to ensure replicability of work by asking authors to submit datasets and models. Examples are The *Journal of Artificial Societies and Social Simulation (JASSS,* http://jasss.soc.surrey.ac.uk/JASSS.html), an interdisciplinary journal for the exploration and understanding of social processes by means of computer simulation; published since 1998, *JASSS* recommends authors to upload model code and associated documentation to the CoMSES Net Computational Model Library. In June 2016, the CoMSES library features 352 agent-based models.

Industry has long embraced big data and advanced data mining, modelling, and visualization algorithms. Computational models are widely used in online recommendation services (e.g., those provided by Amazon or Netflix), by financial and insurance companies (e.g., to detect credit card fraud, estimate fees). Many companies use models internally to support strategic decision making and to guide investment decisions. While code is typically proprietary, close industry-academia-government collaborations are likely beneficial for all parties involved.

Visualization and Communication of Modeling Results

Global operation rooms that provide visualizations of current data and predictions of possible futures are commonplace in meteorology, finance, epidemiology, or defense and might soon be commonplace in support of funding, strategic intelligence, or policy decision making.

William Rouse, Stevens Institute of Technology showcased how operation rooms can be used to support STI decision making, see also Example 3 in Section “ModSTI Examples.” Rouse’s team uses a combination of commercial off-the-shelf tools (e.g., AnyLogic, D-3, Excel, R, Simio, Tableau, and Vensim) rather than writing software from scratch. This practice can enable creating a prototype interactive environment within a week or two, which in turn allows rapid user feedback and easy mid-course corrections.

[2] https://www.imagwiki.nibib.nih.gov

Ben Shneiderman, University of Maryland demonstrated EventFlow, a novel tool for event sequence analytics that includes a timeline display showing all individual records and their point and interval events as well as an aggregated view of all the sequences in the dataset (http://hcil.umd.edu/eventflow) and NodeXL (http://nodexl.codeplex.com; see also Figure 2 in Section "Introduction"). He vividly argued for the need to understand data quality before any type of data analysis is conducted or visualizations are rendered. Blind usage of data is dangerous.

Kevin Finneran, National Academies of Sciences, Engineering, and Medicine argued for the importance of storytelling, i.e., merging data with narrative, when communicating (the value of) research results. He noted that "Stories are the primary way to connect to policy-makers" and "Data can be used to support the stories."

Katy Börner, Indiana University and her team are developing and prototyping "science forecasts," a news show that communicates local and global developments in science, technology, and innovation to a general audience. In Spring 2015, a pilot episode was recorded featuring a moderator that explains trends using an animated map of science (analogous to a weather forecast) and a zoom into a specific research result on 'using Twitter for detecting episodes of depression' presented by Johan Bollen who is interviewed by Fred Cate, both faculty at Indiana University. A still image of the news can be seen in Figure 8.

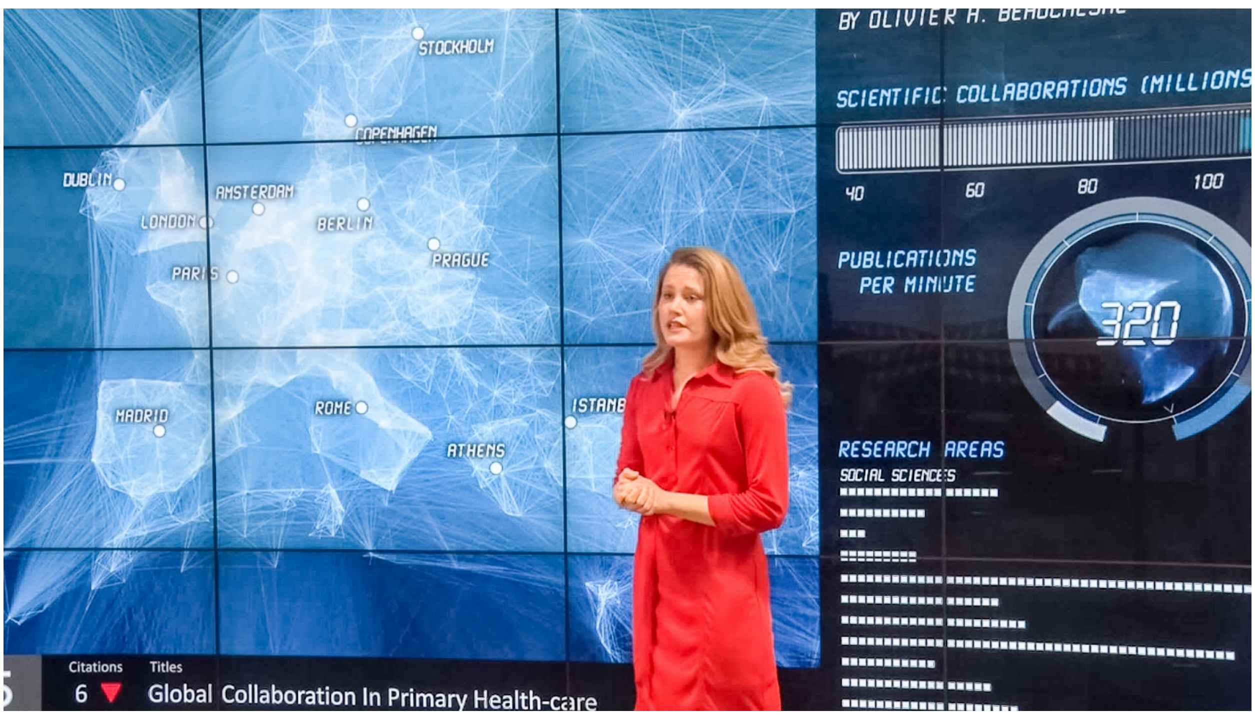

Figure 8: "Science Forecast," recorded at IU, presents interviews and animated maps of scientific activity in a manner similar to weather forecasts. The program demonstrates the power of data and visual analytics to provide up-to-date news on science trends and developments.

Funding

Basic research on STI models is supported by the Science of Science & Innovation Policy (SciSIP) program[3] at NSF and the Scientific Workforce Analysis and Modeling (SWAM) #RFA-GM-14-011 (U01), Modeling Social Behavior #PAR 13-374, Systems Science and Health in the Behavioral and Social Sciences #PAR 15-048 programs at NIH. The Alfred P. Sloan Foundation and the James S. McDonnell Foundation have active funding programs in this area. Representatives from all these agencies were present at the conference.

However, the implementation and validation of computational models is costly. A key insight from the conference is the fact that just like it is common to set aside 10% of the overall budget for program evaluation, it seems appropriate to set a certain percentage of overall costs aside for computational modelling efforts. The concrete percentage amount will depend on an institutions interest (or mandate) to take all knowledge about a system, e.g., the science system, into account and to make decisions that lead to desirable futures.

ModSTI Outlook

One two-day event will not suffice to bridge the gap between academic research ambitions, industry capabilities, and model needs by policy makers. A more continuous, long-term discussion and close collaboration is required to arrive at truly useful models that are widely adopted by science policy makers.

A conference working group listserv was setup to accommodate interests of conference participants to share info on data, models, publications, events and continue the conversations in this research area.

There are a number of upcoming events that will bring together experts with a deep interest in STI indicators, analysis, and modelling. We hope many of the ModSTI Conference participants will be able to attend these events in 2016 or in future years:

2016

- Sept 19-21, OECD Blue Sky Forum on Science and Innovation Indicators Conference, Ghent, Belgium. This conference is organized every 10 years by the Organisation for Economic Co-operation and Development (OECD).
- Sept 19-22, Conference on Complex Systems, Amsterdam.
- Oct 26-27, Multiscale Modeling Consortium Meeting, Washington, D.C.
- Oct 31, Bibliometrics and Research Assessment: A symposium for librarians and information professionals, NIH, Bethesda, Maryland.
- Nov 29-30, International Congress on Agent Computing, Research Hall, George Mason University, Fairfax, Virginia, USA. (Organized by Robert Axtell and Joshua Epstein)

2017

- Feb 16-20, AAAS Annual Meeting, Boston, MA.
- June 19-23, Network Science Conference, Indianapolis, IN.
- Oct 26-30, International Conference on Scientometrics & Informetrics, Wuhan, China.

[3] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501084

Results from the Conference will be presented at the OECD Blue Sky Forum on Science and Innovation Indicators Conference.

There is a forthcoming Special issue of *Scientometrics* entitled *Simulating the Processes of Science, Technology, and Innovation* edited by Katy Börner, Bruce Edmonds, Staša Milojević, and Andrea Scharnhorst that will come out end of 2016.

References

Ahrweiler, P., Pyka, A., & Gilbert, N. (2004). Simulating knowledge dynamics in innovation networks (SKIN). In R. Leombruni & M. Richiardi (Eds.*), Industry and labor dynamics: The agent-based computational economics approach*. Singapore: World Scientific Press.

Ahrweiler, P., Gilbert, N., & Pyka, A., eds. (2015). *Joining Complexity Science and Social Simulation for Innovation Policy*. Cambridge Publishers.

Ahrweiler, P., Schilperoord, M., Pyka, A., & Gilbert, N. (2015). Modelling research policy: Ex-ante evaluation of complex policy instruments. *Journal of Artificial Societies and Social Simulation*, 18: 4, 5. http://jasss.soc.surrey.ac.uk/18/4/5.html

Bollen, J., Crandall, D., Junk, D., Ding, Y., & Börner, K. (2014). From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review. *EMBO Reports,* 15 (1): 1-121.

Börner, K. (2016, Spring). Data-driven science policy. *Issues in Science and Technology*, 32 (3).

Börner, K., Boyack, K.W., Milojević, S., & Morris, S. (2012). An introduction to modeling science: Basic model types, key definitions, and a general framework for comparison of process models. In A. Scharnhorst, K. Börner, & P. van den Besselaar (Eds.), *Models of science dynamics: Encounters between complexity theory and information sciences* (pp. 3-22). Berlin: Springer.

Börner, K., Glänzel, W., Scharnhorst, A., & van den Besselaar, P., eds. (2011). Modeling science: Studying the structure and dynamics of science. *Scientometrics,* 89 (1).

de Rosnay, J. (1979). *The macroscope: A new world scientific system*. New York: Harper & Row. 73.

Esty, D., & Rushing, R. (2007, Summer). The promise of data-driven policymaking. *Issues in Science and Technology,* 23 (4).

Freeman, R.B. (2011). The economics of science and technology policy. In: Fealing KH, Lane JI, Marburger III JH, Shipp SS, editors. *The Science of Science Policy: A Handbook*. Stanford: Stanford University Press. pp. 85-103.

Goffman, W., & Newill, V.A. (1964). Generalization of epidemic theory. An application to the transmission of ideas. *Nature*, 204: 225 – 228.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520: 430-431.

Milojević, S. (2014). Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences (PNAS)*, 111: 3984-3989.

Rouse, W.B. (2014). Human interaction with policy flight simulators. *Journal of Applied Ergonomics*, 45 (1): 72-77.

Rouse, W.B. (2015). *Modeling and visualization of complex systems and enterprises: Explorations of physical, human, economic, and social phenomena*. Hoboken, NJ: John Wiley & Sons.

Scharnhorst, A., Börner, K., & van den Besselaar, P., eds. (2012). *Models of science dynamics: Encounters between complexity theory and information science*. Springer Verlag.

Shneiderman, B. (2016). The new ABCs of research: Achieving breakthrough collaborations. Oxford University Press.

Star, S.L. (1995). Introduction. In: Star S.L., editor. *Ecologies of knowledge: Work and politics in science and technology*. Albany, NY: State University of New York Press. pp. 1-35.

Watts, C. & Gilbert, N. (2014). *Simulating innovation. Computer-based tools for re-thinking innovation*. London: Edward Elgar.

Winslow, R. L. (2012) Computational medicine: Translating models to clinical care. Sci Transl Med, 4 (158): 158rv11.

Wuchty S., Jones B.F., & Uzzi B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316: 1036-1039.

APPENDIX

A. Project Team

Organizers

Dr. Katy Börner is the Victor H. Yngve Distinguished Professor of Information Science and the director of the Cyberinfrastructure for Network Science Center in the School of Informatics and Computing at Indiana University. She serves as a principal investigator for the grant that funded the Modeling Science, Technology, and Innovation Conference.

Dr. Stasa Milojević is an Associate Professor of Information Science in the Department of Informatics at the School of Informatics and Computing at Indiana University. She serves as a co-principal investigator for the grant that funded the Modeling Science, Technology, and Innovation Conference.

Daniel O'Donnell is the project manager for the Cyberinfrastructure for Network Science Center at Indiana University. He served as the organizer of local conference activities for the conference.

Samantha Tirey is a grants specialist at Indiana University and supported the grant submission to NSF. She also participated in the initial logistical and financial planning of the conference.

Please see **Appendix C** for complete biographies.

Scientific Advisors

James Evans, Associate Professor of Sociology, The University of Chicago

Susan Fitzpatrick, President, James S. McDonnell Foundation

Richard Freeman, Herbert Ascherman Chair in Economics, Harvard University

Jerome Glenn, CEO, The Millennium Project

James Owen-Smith, Professor of Sociology, Barger Leadership Institute Professor of Organizational Studies, and Research Professor in the Institute for Social Research (ISR) Survey Research Center (SRC), University of Michigan

Caroline Wagner, Ambassador Milton A. and Roslyn Z. Wolf Chair in International Affairs & Director, Battelle Center for Science and Technology Policy, Ohio State University

Uri Wilensky, Professor of Learning Sciences, Computer Science, and Complex Systems, Northwestern University

Please see Appendix C for complete biographies.

Rapporteurs

Rebecca Reesman works at CNA, Inc, a federally funded research and development center located in Arlington, Virginia, where she analyzes an array of technological issues for the Department of Defense, primarily for the Department of the Navy. Dr. Reesman holds a bachelor of science in physics and statistics from Carnegie Mellon University and a PhD in physics from the Ohio State University. Beginning in the fall of 2016, Dr. Reesman will be an American Institute of Physics Congressional Fellow where she will assist in science and technology policy issues on the Hill.

Layla Hashemi is an Adjunct Professor in the Department of History and Political Science at Montgomery College Rockville and is currently seeking her PhD in Public Policy at George Mason University's School of Politics, Government, and International Affairs (SPGIA). She is a member of Montgomery College's Peace and Justice Community, a part-time faculty representative for the Center of Teaching and Learning (CTL) and currently serves as an Executive Committee member at the Montgomery College chapter of Local SEIU 500. Hashemi has worked at various governmental and non-governmental organizations including Forum 2000 (Prague, Czech Republic based NGO focused on building a stronger global civil society), the NYU School of Law, and the New York State Unified Court System. She earned her M.A. in International Relations and Comparative Politics at New York University with a concentration on Middle Eastern and Islamic Studies.

B. Conference Agenda

The conference featured talks and panel discussions by representatives from government agencies, university administrators, and other science policy makers on current and future STI model needs. The conference agenda with links to slides and recordings of all talks is available online at http://modsti.cns.iu.edu/agenda.

Tuesday May 17th

8:00 – 8:30 *Breakfast*

8:30 – 9:00 Welcome and Opening Remarks by C.D. (Dan) Mote, Jr., President of the National Academy of Engineering

9:00 – 9:30 Setting the Stage by Katy Börner, Indiana University

9:30 – 11:00 **Case Studies**
Government and policy researchers and staff present computational models they have implemented to optimize internal processes and to improve agency decision making.

11:00 – 11:30 *Break*

11:30 – 12:30 Keynote by Alex "Sandy" Pentland, MIT

12:30 – 1:30 *Lunch*

1:30 – 2:00 Two-Minute Flash Talks by Leading Experts

2:00 – 3:00 **Funding Opportunities**
Government and private foundations discuss how they fund people, projects, and infrastructure in support of R&D on validated and trusted STI models and standards.

3:00 – 4:30 **Models of Innovation**
Academic experts from different disciplines discuss diverse approaches to model the birth, diffusion, and adoption of innovations in science and technology. Temporal dynamics, diffusion trajectories, and the impact of interventions are covered.

4:30 – 5:00 *Break and Group Photo*

5:00 – 6:30 **Models of STI**
Researchers present models that help us understand the inner working of STI and/or that aim to address the needs of science policy makers.

6:30 Reception with Remarks by E. William Colglazier, AAAS Center for Science Diplomacy
Tour of Sentient Chamber Living Architecture Installation

8:00 PM *Adjourn*

Wednesday May 18th

7:30 – 8:00 *Breakfast*

8:00 – 9:30 **Data, Algorithms, and Infrastructure**
High quality predictions require access to high quality and high coverage data. Just like local data is of little value for global weather predictions; data for just one institution or country is of limited value when aiming to make STI predictions.

9:30 – 10:00 Break

10:00 – 11:30 **Models of Science**
Exactly how can models of science inform decision making in academia, government, and industry? Leading experts present their models and discuss limitations to predictability.

11:30 – 12:30 **Policy Issues**
Hear first-hand from policy makers the types of issues they are dealing with and what kinds of models and model results would help them to be informed decision-makers.

12:30 – 1:30 *Lunch and Tour of Sentient Chamber Living Architecture Installation*

1:30 – 2:00 Two-Minute Flash Talks by Leading Experts

2:00 – 3:00 Keynote by Dame Wendy Hall, University of Southampton, UK

3:00 – 3:30 *Break*

3:30 – 5:00 **Models of Science & Innovation**
Learn how models of science and innovation can improve decision making and how computer simulations can help understand the impact of (policy) decisions on future developments.

5:00 – 5:30 Closing Remarks and Next Steps by Staša Milojević, Indiana University

5:30 *Adjourn*

C. Conference Abstracts

Organized in the sequence of speaker presentations, see Appendix B. Conference Agenda.

Richard Ikeda

National Institutes of Health

"NIH Experiments with IBM Watson"

NIH uses text mining technologies (https://report.nih.gov/rcdc/process.aspx) to produce its Categorical Spending reports (https://report.nih.gov/categorical_spending.aspx).

Consequently, NIH is keenly interested in advances in text mining technologies that would improve capabilities and utility.

At its core, Watson is a robust implementation of a next generation technology for deriving facts and meanings from publications, data sets or other information sources. Watson "reads" documents and parses the text into keywords and sentence fragments in order to identify relationships among the different bits of information.

NIH has pursued a Watson "proof-of-concept" in collaboration with IBM and the latest experiments with Watson will be discussed.

Guru Madhavan

National Academies of Sciences, Engineering, and Medicine

"Systems Analysis for Global Health Planning"

Planning tools based on narrow efficiency metrics (e.g., cost-effectiveness) miss out on a number of important factors that often underpin final policy decisions. A comprehensive systems analysis approach is needed to improve our programmatic effectiveness in global health, especially with preparedness, response, and resilience as challenged by recent disease outbreaks. I will discuss a platform concept for strategic policy planning—expanding upon Strategic Multi-Attribute Ranking Tool for Vaccines (or SMART Vaccines), a decision support software based on multi-criteria systems analysis developed by the National Academies of Sciences, Engineering, and Medicine, now being enhanced into a web-based application for broad use by the U.S. Department of Health and Human Services. I will explore the broader potential of such a systems platform to consider and formally include many other factors affecting short and long-term planning and response (especially when disasters affect vulnerable communities), and demonstrate how a collaborative, transparent policy decision support system across stakeholders could be developed and deployed for public benefit.

Martin Meltzer

Centers for Disease Control and Prevention

"Using graphs and maps to aid public health decision making during an emergency response"

Since 2009, the CDC has participated in emergency responses ranging from small domestic outbreaks involving a few hundred people, to international public health emergencies affecting millions of individuals from around the world. Each response required unique skills from the CDC'S Modeling Task Force to address public health officials' questions and to help them make informed decisions. This presentation will discuss the roll of the Modeling Task Forces in the CDC's Incident Management structure and some of the models used to assist officials in making decisions about the potential size of the public health crisis, how effective interventions could be, and what resources are required.

Matteo Convertino

University of Minnesota

"Enhanced Adaptive Management for Population Health: Integrating Ecosystem Dynamics and Stakeholder Mental Models"

Ecosystem health issues abound worldwide with environmental implications and impact for animal and human populations. The complexity of addressing problems systemically in the policy arena on one side, and the lack of use of computational technologies for quantitative public policy on the other side has determined a worsening of ecosystem health.

Maria Larenas

National Institutes of Health

"Impact of NIH Funded Postdoctoral Training on Future Career Outcomes"

Training future scientists is critical for all federal scientific agencies and for the future of scientific research in the United States. The National Institutes of Health (NIH), the federal scientific agency tasked with enhancing health and reducing disability and illness through research advancements, dedicates significant funding to training the next generation of biomedical researchers. NIH invests in postdoctoral research training to increase independent research and professional experiences and to promote future research career success for emerging biomedical scientists. Given exponential increases in biomedical postdoctoral appointments since 1995, and widely-held expectations that these opportunities enhance future research career success, this paper presents new evidence about the impact of postdoctoral fellowship participation on research career outcomes. We estimate the causal effects of F32-NIH postdoctoral fellowship programs on the probability to continue doing research in areas of NIH interest using a fuzzy regression discontinuity design. We examine a dichotomous RPG award outcome and find that receiving an individual postdoctoral fellowship award at the margin increases the probability of receiving a RPG grant in about 6 percentage points after 4 years or more of their last F32 application.

Jeroen Struben

McGill University

"CSR-Mainstreamed Innovation: A Model of Market Transformation for Scaled Solutions to Socio-Economic Inequity"

Corporate social responsibility (CSR) has emerged over the years as a mitigation strategy to unanticipated negative externalities of industrial technologies and markets. However, the scope, scale and impact of what has been possible through CSR in addressing these major societal challenges is clearly insufficient, as are the efforts deployed by governments and actors from the not-for-profit (NFP) sector. We develop and argue a computational model of convergent innovation (CI), a cross-sectoral approach to mainstream the societal issues targeted by CSR into core for-profit (FP) activities, placing them upfront as a driver of commercially successful technological innovation, business strategy, and market transformation, while having FP actors join governments and NFP actors to enact behavioral change and ecosystem transformation at scale. CI also entails social and institutional innovation to enable such a shift in the drivers of supply and demand at market level and in broader society. Using socio-economic inequity in access to healthy food in industrial Western society as a context, this paper lays the foundations for the dynamic modeling of equitable nutrition market transformation in the agri-food sector. We first deconstruct the existing ecosystem to specify the major inertial forces constraining change. We then use a stylized behavioral dynamic model to simulate interventions for single-actor and collective actions by FP and NFP actors and governments and examine economic change and inequity reduction outcomes over time. Results show that the economic viability of lasting social change requires cross-sectoral convergence between CSR-mainstreaming business strategy and market transformation and actions by NFP actors and government.

Julie Mason

National Institutes of Health, National Cancer Institute, Center for Cancer Training

"Labor and Skills Gap Analysis of the Biomedical Research Workforce"

The United States has experienced an unsustainable expansion of the biomedical research workforce over the past three decades. This has led to myriad of consequences, including imbalance in the number of researchers and available tenure track faculty positions; extended postdoctoral training periods; rising age of investigators at first NIH R01 grant; and exodus of talented individuals seeking careers beyond traditional academe. Without accurate data on the biomedical research labor market, challenges will remain in addressing these issues and in advising trainees of viable career options and necessary skills to be productive in their careers. Herein, we analyzed workforce trends, integrating both traditional labor market information and real-time job data. We generated a profile of the current biomedical research workforce, performed labor gap analyses of occupations in the workforce at regional and national levels, and assessed skill transferability between core workforce occupations and complementary occupations. We conclude that although supply into the workforce and the number of job postings for occupations within that workforce have grown over the past decade, supply continues to outstrip demand. Moreover, we identify technical and foundational skill gaps in the workforce through analysis of real-time job postings. Addressing these skill gaps could potentially equip trainees for multiple career pathways beyond academic research and lead to a more sustainable workforce.

Timothy Slaper

Indiana Business Research Center

"Driving Regional Performance: Theory and Measurement in Innovation Research"

What drives regional innovation and economic performance? This paper reviews the major topical areas in the study of innovation and applies those theories and findings to developing measures for innovation and economic performance across regions. Attempts to create indices for innovation have focused almost exclusively on countries or states, thus precluding the ability to analyze the drivers of regional development. In addition, these indices are often under-appreciated by academics and researchers because many of the measures used to construct the indices lack theoretical and empirical support.

Our research corrects two deficiencies. One, our geographic unit of analysis for the innovation measures is the county. Using U.S. county definitions enables regional analysis across state boundaries for a single year as well as consistent metropolitan boundary definitions over time. Two, we operationalize key innovation concepts such as knowledge spillovers, venture capital, foreign direct investment, business formation, human capital and technology diffusion, in addition to the traditional drivers and measures of economic performance. The authors also propose metrics for measuring the performance of industry clusters on a county and regional basis as well as classifying patent technologies into a dozen major patent categories. This paper is, in effect, the motivation and theoretical foundation for the Innovation Index 2.0 — currently in beta. (http://www.statsamerica.org/innovationindex/Default.aspx)

Kenneth Gibbs

NIH/NIGMS

"Biomedical Workforce Diversity: Missing Linkage Between Underrepresented Minority Talent Pool and Basic Science Departments in Medical Schools"

African-American/Black, Hispanic/Latino, and American Indian/Alaska Native scientists are poorly represented on the faculty of basic science departments in MD-granting medical schools, despite decades of efforts to increase faculty diversity. The authors utilized the National Science Foundation Survey of Earned Doctorates and the Association of American Medical Colleges (AAMC) Faculty Roster to describe changes in the participation of scientists from underrepresented minority (URM) and well-represented (WR) backgrounds in the populations of (i) biomedical Ph.D. graduates, and (ii) full-time assistant professors in basic science departments at MD-granting medical schools between 1980-2013. These data were used to impute faculty-hiring trends, and to calibrate a system dynamics model of the progression from Ph.D. graduates to assistant professorships in basic science departments. The size of the potential candidate pool was significantly associated with the number of assistant professors hired each year for scientists from WR ($r2=0.48$, $p<0.0001$) but not URM backgrounds ($r2=0.12$, $p>0.07$). Between 2005-2013, data indicated that there were 5,824 biomedical Ph.Ds. awarded to URMs, and a 7% growth in the total population of assistant professors, but no growth in the population of URM assistant professors. The system dynamics model explained significant variance in faculty hiring ($r2=0.79$). The model predicted that given current trends in transition from Ph.D. to assistant professors, URM faculty representation would remain below 8% as late as 2080—even in the context of exponential growth in the population of URM Ph.D. graduates. Faculty diversity cannot be achieved by relying primarily on increasing the pool of URM Ph.D. graduates. Efforts to increase faculty diversity must make meaningful linkages between URM Ph.D. graduates and faculty hiring.

Xiaoran Yan

Indiana University

"Graph transformations of scholarly networks"

We introduce an umbrella framework for defining and characterizing an ensemble of dynamic processes on graphs. It leads to intuitive linear transformations of graphs which can represent the flow of different dynamic processes, including consensus, random walks, as well as information diffusion over networks.

We show some empirical examples of how such transformations can be applied in scholarly networks where additional data is available, such as geographic and disciplinary maps. The goal is to produce multi-layered scholarly networks that better capture underlying scientific activities.

Bruce Hecht

Analog Devices

"Framework for Scalable Sensors for the Internet of Things and People"

The growth in scalable sensors is driving opportunities for things and people to work together in expanded methods. Platforms that combine data sets with machine learning algorithms, curated by people and used by networks of people, are transforming significant aspects of the world in which we live. An example is the combination of online maps, traffic data, handheld devices, and the GPS network that is generating navigation tools as well as leading to new transportation services for drivers, passengers, cargo, and autonomous vehicles. Looking forward, by integrating networks at exponential scale from sensors through signal processing, edge computation, and cloud computing, the power of linking across these domains is enabling transformative change to agriculture, manufacturing, healthcare, and education. To approach this problem at multiple scales of operation, techniques of systems engineering are emerging that incorporate hardware, software, as well as the socio-technical system of designers, operators, regulators, and users. These techniques build on those developed for Safety Design and Analysis developed by Nancy Leveson, of MIT Engineering Systems Division, and application to complex systems engineering curated by Anna McGowan, NASA senior engineer for complex systems design. Highlights of platform requirements and design will be presented together with two illustrative examples from smart agriculture and for the Factory of the Future.

Masaru Yarime

University of Tokyo and University College London

"Modeling Innovation Systems to Address Grand/Societal Challenges: A Case of Smart Cities"

Science, technology, and innovation are critical components of our efforts to tackle societal challenges we face today. Smart cities would be considered to be a key field in which a variety of science and technological knowledge need to be integrated effectively to address the combined target of energy security and environmental protection. A smart city would involve an advanced technological system for efficient electricity supply and applications, incorporating all the behavior of the actors involved, including generators, distributors, technology developers, and consumers, through an intelligent information network. As a smart city integrates a diverse mixture of hardware as well as software involving a large amount of various kinds of data in a complex way, different approaches would be possible for creating and implementing innovation on smart cities in practice, depending on the economic, social, and environmental factors, such as energy efficiency, operating cost, environmental impact, resilience to external shocks and disturbances, and accessibility and inclusiveness to end users. This study examines the innovation systems of smart cities in Japan, Europe, and the United States. Approximately 200 projects on smart cities are analyzed with regard to the knowledge domain, actors involved, and institutional environment. Information was collected through various sources, such as project reports, academic articles, corporate reports, trade journals, and web sites. Interviews were conducted with relevant stakeholders, including academia, firms, industry association, and government organizations. Network analysis is conducted to identify key stakeholders involved in innovation and to analyze the relationships between them. The innovation systems of smart cities are modeled based on the primary functions involved in the dynamic processes. These include the creation of future visions based on science, setting of concrete and practical goals and targets, joint scenario making with stakeholders, securing active participation and serious engagement of stakeholders, collection and analysis of data on societal needs and demands, development of new technologies and systems through social experimentation at universities as living laboratories, assessment of impacts with transparency, objectivity, neutrality, legitimation of innovation in society, provision of effective feedback to decision makers, incorporation into institutional design, and contribution to agenda setting at regional, national, and global levels. The model developed is applied to explain the differences observed in the direction and process of innovation on smart cities between Japan, Europe, and the United States. In Japan a strong focus is placed on sophistication of application technologies for extensive use of home appliances and electric vehicles. In Europe an emphasis is placed on establishing a basic infrastructure in which information about the behavior of all the stakeholders is collected and distributed among the stakeholders appropriately so that the various objectives of the electricity grid are achieved in a more equitable way. In the United States a strong interest can be observed in creating and maintaining security through improvement in resilience against physical as well as virtual threats. These asymmetries in conceptualizing and implementing smart cities reflect the differences in how knowledge development, stakeholder networks, and institutional environment interact in dynamic and systemic manners.

Philip Beesley

School of Architecture, University of Waterloo

"Living Architecture Systems"

Philip Beesley's Living Architecture System Group at the University of Waterloo is exploring new kinds of building systems that raise fundamental questions about how architecture might behave in the future.

Might future buildings begin to know and care about us? Might they start, in very primitive ways, to become alive? This experimental new work draws together multiple disciplines that include next-generation lightweight structures, interactive robotics, and synthetic biology in pursuit of a kind of architecture that comes close to being alive. Visualizing this responsive architecture presents formidable challenges, and it also offers striking opportunities for thinking and working with complex systems.

Recent projects are composed of towering transparent acrylic arches and flexible silicon, creating quilt-like patterns and composite structures. Custom glasswork vessels housing synthetic biology and translucent filtering elements expand the skeletons to form hovering surfaces that interplay with shadow and light. Distributed sensors and mechanisms are controlled by arrays of microprocessors that give these environments the power to sense and perceive, reacting to the presence of visitors with mechanic curiosity and by delicate waves of light, motion and choruses of murmuring sounds. The work is being created by a group of architects, engineers, scientists, and artists from Canada, the U.S., and Europe within the Living Architecture Systems Group. Their design methods are being used to train new generations of architects and engineers, providing them with skills to work with complex and interconnected sustainable environments. For more information: www.lasg.ca

Alina Lungeanu

Pennsylvania State University

"Team Assembly in New Emerging Fields: A Computational Modelling Approach"

In this paper we use a multi-level, multi-theory framework to study the influence of compositional, relation, and ecosystem mechanisms on the assembly of scientific teams in interdisciplinary fields. Specifically, we test the effects of these mechanisms on the assembly of interdisciplinary scientific teams using a novel hybrid agent-based and systems dynamics computational model fitted using data collected from 533 teams and 1,696 researchers working in the scientific field of Oncofertility from its inception in 1996 until 2010. We found that when a new field emerges, team assembly is influenced by the reputation and seniority of the researchers, prior collaborators, prior collaborators' collaborators, and the prior popularity of an individual as a collaborator by all others. We also found that individuals are more likely to assemble into an Oncofertility team when there is a modicum of overlap across its global ecosystem of teams. The ecosystem is defined as the collection of teams that share members with other teams that share members with the Oncofertility team. The impact of the assembly mechanisms vary over the 15-year period with clear different trends prior and after 2007. It is noteworthy that the changes that appear around 2006-2007 coincide with the National Institutes of Health (NIH) funding the creation of the Oncofertility Consortium. This illustrates the impact of external events such as funding on individual motivations to form teams. This research indicates that the NIH's funding initiative to create a national Interdisciplinary Research Center (IRC) on Oncofertility had the intended consequence of facilitating an assembly of teams among those universities funded by the IRC, but had the unintended consequence of chilling research collaborations within the larger community of oncofertility not funded by the IRC.

Ben Shneiderman

University of Maryland

"The New ABCs of Research: Achieving Breakthrough Collaborations"

View more information about *The New ABCs of Research: Achieving Breakthrough Collaborations* on Amazon**.**

Kalev Leetaru

GDELT Project

"Quantifying, Visualizing, and Forecasting Global Human Society Through "Big Data": What it Looks Like to Compute on the Entire Planet"

What happens when massive computing power brings together an ever-growing cross-section of the world's information in realtime, from news media to social media, books to academic literature, the world's libraries to the web itself, machine translates all of that material as it arrives, and applies a vast array of algorithms to identify the events and emotions, actors and narratives and their myriad connections that define the planet to create a living silicon replica of global society? The GDELT Project (http://gdeltproject.org/), supported by Google Ideas, is the largest open data initiative in the world focusing on cataloging and modeling global human society, offering a first glimpse at what this emerging "big data" understanding of society looks like. Operating the largest open deployments of streaming machine translation, sentiment analysis, global geocoding, and event identification, coupled with perhaps the world's largest program to catalog local media, the GDELT Project monitors worldwide news media, emphasizing small local outlets, live machine translating all coverage it monitors in 65 languages, flagging mentions of people and organizations, cataloging relevant imagery, video, and social posts, converting textual mentions of location to mappable geographic coordinates, identifying millions of themes and thousands of emotions, extracting over 300 categories of physical events, using deep learning to quantify visual narratives alongside the textual world, and making all of this available in a free open data firehose of human society. This is coupled with a massive socio-cultural contextualization dataset codified from more than 21 billion words of academic literature spanning most unclassified US Government publications, the open web, and more than 2,200 journals representing the majority of humanities and social sciences research on Africa and the Middle East over the last half century. Used by governments, NGOs, scholars, journalists, and ordinary citizens across the world to identify breaking situations, map evolving conflicts, model the undercurrents of unrest, explore the flow of ideas and narratives across borders, and even forecast future unrest, the GDELT Project constructs a realtime global catalog of behavior and beliefs across every country, connecting the world's information into a single massive ever-evolving real-time network capturing what is happening around the world, what its context is and who is involved, and how the world is feeling about it, every single day. Here is what it looks like to conduct data analytics at a truly planetary scale and the incredible new insights we gain about the daily heartbeat of our global world.

Venkatachalam "Ram" Ramaswamy

National Oceanic and Atmospheric Administration

"Prediction of Climate Extremes for Decision-making"

The mantle of understanding and predicting the state of weather and climate is a principal mandate of the National Oceanographic and Atmospheric Administration. NOAA's mission objectives are Science, Service, and Stewardship, with the responsibility of providing credible, trustworthy forecasts of the state of the weather and climate system for the nation on timescales ranging from daily to seasonal to decadal to centennial. NOAA carries its mission through observations, and through scientific understanding and prediction using mathematical formulation of the processes and interactions occurring in the Earth's systems (atmosphere, oceans, land, and ice) that is solved on high-performance computers.

Before credible forecasts can be produced, the mathematical models have to engage in rigorous science that evaluates the theoretical and observational knowledge about the components of the system, the uncertainties in the science, and the range of solutions possible given the natural variations and the forced changes on the system. Ensemble solutions for the modeling are performed in which a wide variety of possible initial states of the system have to be accounted for in order to have realistic predictions with probabilistic outlooks. Even more challenging is the prediction of extremes (e.g., heat waves, excess or deficit rainfall, hurricanes) occurring in the days, seasons, and years ahead and that cause destruction of life and property. This challenge has to be addressed with increased rigor since numerous societal sectors need authoritative information to address population and economic risks.

The task of advancing the science of predictions to produce high-quality, user-friendly data requires sustained expertise at an outstanding level. This, in turn, necessitates maintaining the highest standards through steady recruitment and retention of a creative, skilled workforce, and through advances in observational and computational infrastructure. Actionable information for practical decision-making also requires advances in technical skills to manipulate the increasingly vast amounts of climate data, and the concomitant need to improve visual algorithms for easy discernment of the significance of the science-based information.

The outcomes of weather and climate prediction science feed into national policy decision inputs, and also serve the nation in bilateral and multilateral exchanges. Thus, the NOAA activities include dissemination to national and international bodies such as the National Climate Assessment, Intergovernmental Panel on Climate Change, Federal, state, local and tribal agencies, and combination sectoral bodies such as the Western Governors Association. The paramount need is actionable information which facilitates risk assessment by the various sectors, and integrates with other factors besides climate that must be factored in sectoral decisions. The technology to propagate increasingly useful climate information, with quantified uncertainties, relies on the feedback from the stakeholders e.g., estimates of the gains from the information received, with accompanying calibration in the management of expectations.

Alex "Sandy" Pentland

MIT

“Social Physics: Data and Innovation”

We are building what are probably the world’s largest, most complete data resources for understanding human behavior and social evolution. As part of the UN Sustainable Development Goals program, we have worked with Orange, AFD (France), PARIS21, UN, etc., to develop and deploy the OPAL (Open Algorithms) system that combines live data from telcos, banks, health, etc., to enable every nation to have a constant, up-to-date picture of aggregate human behavior. As part of the Kavli HUMAN project we have begun to measure everything from microbiome to social distances to medical records for 10,000 New Yorkers over the next 30 years, providing the first statistically balanced, contextually rich picture of human health and development. Similar efforts are underway in financial and legal spheres. These efforts are already revealing new computational, actionable insights into innovation and the business of science, and suggest a new path for government to ensure progress of prosperity and human understanding.

Stephen Marcus

National Institutes of Health

“Behavioral Science Modeling at NIGMS”

Modeling and simulation are powerful tools for studying health and the role of research and funding programs in achieving health objectives. Almost all research involves some kind of conceptual model but these are often implicit, based on unstated assumptions, and have unknown relationships to data as well as unverified consequences. Much of the value of mathematical, statistical, and computational modeling comes from its forcing us to make our assumptions explicit. Moreover, the role and purpose of modeling are not limited to prediction or forecasting. Models also help us “explain observations, understand system dynamics, illuminate uncertainties, offer options for interventions, set boundaries of parameters and outcomes, discipline our thinking, and identify new questions.” The social science perspective is critical to understanding health, at both the phenomenological and mechanistic level. Not least because of the powerful direct effects of behavior on health but also because of the complex feedback between emotional, cognitive, psychological, and social phenomena and dynamics and biological phenomena and processes. Exploring interactions and emergent behaviors should include study of the context in which the behaviors and the system exist. This systems-level, multi-scale exploration requires a transdisciplinary, team science approach. To this end, the Social and Behavioral Modeling Research Program at NIGMS supports research in basic behavioral and social modeling that applies computational and systems approaches to the study of bio-related behavioral and social phenomena at multiple scales, ranging from genes to populations. The program encourages applications that bring together transdisciplinary teams of scientists with expertise in the behavioral and social sciences and in the computational and quantitative sciences. As part of its continuing effort to employ data-driven, scientifically rigorous tools to plan and evaluate programs to develop and maintain a strong and diverse scientific workforce, NIGMS is also pursuing a systems-based approach to studying the scientific workforce. The Scientific Workforce Analysis and Modeling (SWAM) program is a concerted effort focused on understanding the underlying dynamics that produce successful scientists, examining strategies for increasing the diversity of the scientific workforce, identifying questions in need of research, and guiding the collection and analysis of data. In addition, other modeling initiatives across the NIH will be discussed, including extramural funding programs and in-house activities.

Riq Parra

Air Force Office of Scientific Research

"Opportunities for Partnership with the Air Force Office of Scientific Research (AFOSR)"

As an integral component of the Air Force Research Laboratory (AFRL), the Air Force Office of Scientific Research (AFOSR) has the responsibility to discover, shape and champion basic research that profoundly impacts the future Air Force. The talk will provide an overview of the agency and its research focus areas.

Daniel Goroff

Alfred P. Sloan Foundation

"The Productivity of the Scientific Enterprise: A Research Agenda"

What is the rate of return on investments in basic research? How do the incentives, institutions, and inner workings of the research system affect those returns? Why is the market for scientists and engineers similar or different from other labor markets? Ideally, how would a social planner organize scientific research, rewards, and risk-taking? The Alfred P. Sloan Foundation actively supports rigorous research projects on questions like these about the economics of science.

Brian Pate

United States Department of Defense

"Awareness and Forecasting for Countermeasures to Weapons of Mass Destruction"

Safeguarding the U.S. and its allies from weapons of mass destruction (WMD) requires a coordinated approach to global situational awareness and assessment of technologies and threats, as well as cultivation of improvements in the ability to forecast emerging science and technology trends and their potential impact within this arena. In order for the Defense Threat Reduction Agency (DTRA) to meet these current and future challenges, a portion of current activities focus on improving the anticipation of technology-driven emerging and disruptive WMD threats as well as potential countermeasure approaches. I will share selected highlights of these activities within the overall context of DTRA's mission, together with opportunities for new collaborations to address these challenges.

John Walsh

Georgia Tech

"Inventing while you work: Knowledge, non-R&D learning and innovation"

Intuition, judgment, creativity are basically expressions of capabilities for recognition and response based upon experience and knowledge (p. 128–129) (Simon, 1997). Workers gain experience and knowledge in the course of their normal jobs. Therefore, innovative ideas can be generated from knowledge built from learning opportunities across the firm (not just the R&D lab). Employees working for different functions (R&D and outside of R&D) in an organization have different work practices and build their learning through different processes. Moreover, the relative effectiveness of learning by different work practices for innovation is contingent on the nature of knowledge, characterized by generality (i.e., high mobility/transferability) and visibility (i.e., tighter links between actions and outcomes). Using multiple datasets combining public and private data and focusing on births of innovations, this study shows how the nature of knowledge affects differences in the innovation productivity of R&D and non-R&D work. The paper concludes with a discussion of the implications of these insights for innovation management and policy.

Lynne Zucker

University of California, Los Angeles

"Science-Business Knowledge Bridges"

Many paths to sharing or selling access to scientific knowledge have been documented, but which ones really matter? Though many different bridges exist, we focus those central to the "corporate-academic" model. This model emphasizes attracting the best and brightest scientists, providing them with a commensurate increase in autonomy including initiation of bench-level collaborations with top university scientists in which valuable tacit knowledge is transferred in both directions. More generally, this basic model of knowledge flow holds whenever two organizations, or person, hold different protected knowledge content (tacit if natural excludable; codified if protected by law, patent or trade secret) which if put together yield significantly higher returns to both organizations/persons compared to alternative use (Zucker and Darby 2014).

We propose and test two strong knowledge bridges as indicators of adoption of the corporate-academic model, whether or not the firm has ever: (a) co-authored an article with a university scientist and (b) applied for (and eventually granted) a patent with non-patent references, where these references are used importantly to cite scientific articles and other scientific materials. Both were robustly positive and statistically significant across four measures of U.S. high-tech firm success (publishing, patenting, obtaining venture capital, and going public) for six broad S&T areas (bio/chem/med, information technology, nanotechnology, semiconductors, other science, and other engineering). Star scientists' publications, as or with firm employees, SBIR grants received, and citation-weighted patents and articles all played comparatively supporting roles in the empirical estimates. We concluded that the most successful high-tech firms have adopted a strategy of operating near the edge of the scientific envelope where high levels of tacit knowledge provide substantial natural excludability reducing or preventing entry of imitators, both by doing basic science themselves with university scientists and by implicitly including this knowledge in patents through non-patent references.

William B. Rouse

Stevens Institute of Technology

"Human-Centered Innovation"

Human-centered design is a process of considering and balancing the concerns, values, and perceptions of all the stakeholders in a design (Rouse, 1991, 2007). This presentation will elaborate on the concepts and principles of human-centered design in the context of innovation. Invention is the creation of a new process or device; innovation is the introduction of change via something new (Rouse, 1992). Motivations for change go far beyond the existence of an invention; consequently, the vast majority of inventions never contribute to innovations (Rouse, 1993, 2001, 2007). Motivations include aspirations for substantial improvements of the validity, acceptability, and viability of new products, services and, in general, solutions. When key stakeholders are not open to purely technical arguments about validity, acceptability, and viability, we have found that interactive visualizations can play an invaluable role (Rouse, 1998, 2007, 2015, 2016). This has led to the general notion of policy flight simulators (Rouse, 2014, 2015, 2016). Numerous examples will be presented to illustrate how interactive visualizations embedded in policy flight simulators have enabled innovations with diverse stakeholder groups.

Daniel McFarland

Stanford University

"Intellectual Movement and Migration"

I will present preliminary results of collaborative research that employs large academic corpora to develop models of intellectual movement and scholarly migration within and between fields. In so doing, we identify multiple resource flows and a larger social structure of science where some fields assume positions of greater influence. We demonstrate that the transfer of knowledge and personnel is episodic and uneven; that interdisciplinarity entails forms of colonization and pollination; and that science policies of funding and institutional support seem to nudge flows along identified paths.

Petra Ahrweiler

EA of Technology and Innovation Assessment GmbH, Germany

"Agent-based Simulation for Science, Technology and Innovation Policy"

Policymaking implies planning, and planning requires prediction – at least some knowledge about the future. This contribution starts from the challenges of complexity, uncertainty, and agency, which refute the prediction of social systems, especially where new knowledge is involved as a radical game-changer. It is important to be aware of the fundamental critiques, approaches, and fields such as Technology Assessment, the Forrester World Models, Economic Growth Theory, or the Linear Model of Innovation received in the past decades. However, agent-based modeling and simulation now provide new options to address the challenges of planning and prediction in social systems: this paper will discuss these options for STI policy with a particular emphasis on the contribution of the social sciences, both in offering theoretical grounding and in providing empirical data. Fields such as Science and Technology Studies, Innovation Economics, Sociology of Knowledge/Science/Technology, etc., inform agent-based models in a way that realistic representations of science, technology and innovation policy worlds can be brought to the computer. These computational STI worlds allow scenario analysis, experimentation, policy modeling and testing prior to any policy implementations in the real world. This contribution will illustrate this for the area of STI policy using examples from the SKIN model.

Ben Shneiderman

University of Maryland

"Innovation Trajectories: A basis for understanding and acting"

The growing availability of information on funding from SBIR and STTR projects, combined with patent and sales data, enable analysis of innovation trajectories that result in success or failure. Using event analytics tools like EventFlow allows us to understand common patterns for success and outliers to guide planning and funding efforts.

James Onken

National Institutes of Health

"Improving the Research Portfolio Data Infrastructure at the National Institutes of Health"

Demonstrating the impact of research investments made years–and sometimes decades–earlier and using that information to predict future trends in science has never been easy. Now, the increased availability of relevant databases, new database technologies, and informatics capabilities create the potential to more readily establish linkages between federal investments in science and long-term outcomes. This talk describes an effort the NIH is making to create a data infrastructure that will facilitate the analysis of NIH research investments and the development of predictive models.

Ian Hutchins

National Institutes of Health

"Utilizing citation networks to explore and measure scientific influence"

The 2013 San Francisco Declaration on Research Assessment decried the widespread and invalid use of Journal Impact Factors for comparing the scientific output of scientists or institutions. The NIH Office of Portfolio Analysis has developed an improved method to quantify the influence of a research article by making novel use of its co-citation network to field-normalize the number of citations it has received. Article citation rates are divided by an expected citation rate that is derived from the performance of articles in the same field and benchmarked to a peer comparison group. The resulting Relative Citation Ratio (RCR) is article-level and field independent, identifies influential papers independently of their publication venue, and thus provides an alternative to the invalid practice of using Journal Impact Factors. We demonstrate that the values generated by this method strongly correlate with the opinions of subject matter experts in biomedical research, and suggest that the same approach should be generally applicable to articles published in all areas of science. A beta version of iCite, our web tool for calculating the RCRs of articles listed in PubMed, is available at https://icite.od.nih.gov.

Richard Freeman

Harvard University

"The Missing Link in How Science and Engineering Affect the Economy"

Studies of the impact of R&D and the work of scientists and engineers on the economy typically relate some measure of R&D (usually a stock created from flows) or of patents to levels or growth of productivity, sales, or profits. But neither R&D nor a patent produces a final product for sale. They produce knowledge that might contain an idea for a new product or process. They are inputs or intermediate outputs that enter a production or profits equation – valuable indicators of innovative activity – but not measures of actual innovation as defined by Schumpeter or the Oslo manual, which require their implementation or commercialization in the market. Measures of actual innovations are the missing link in our understanding of how science and engineering shape the economy. I examine three different ways to gain insight into actual innovations: through questions about the introduction of new products and processes in the NSF's Business Research and Development and Innovation Survey (BRDIS); through web-scraping the attributes, prices, and quantities sold of goods and services on websites; through an Innovation Hunter crowd-sourcing activity in which volunteers search announcements and reports of new products and processes along the lines of the 1982 Small Business Administration study of innovations that provided the data for Audretsch and Feldman's (1996) analysis of the geography of innovation and production.[1] Measures of actual innovations in the form of new products and processes can be produced regularly on a world-wide basis. They have the potential for increasing our understanding of company reports of innovation on BRDIS, the EU Community Innovation Survey, and comparable survey data for China, and of transforming discussions of innovation that rely on aggregate indicators, on the one hand, or on business school case studies of new goods and services introduced in markets. As Lord Kelvin famously said, "When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind." There is a lot new to be discovered in modeling science, technology, and innovation that requires new micro-data. I will give explicit examples of the modes of measuring innovation and the models and research and policy questions they can illuminate.

[1] Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review,*86 (3): 630-40.

Nachum Shacham

PayPal

"Data, platforms and predictive models to enable data-driven actions"

Organizations strive to make data-driven decisions and tailor their actions to the individual needs of millions of customers, suppliers, and partners worldwide. Big data, storage and computation infrastructure, and predictive model algorithms enable this process. Making them work in harmony with the planned actions successful, takes engineering, data science, and business skills. The ingredients are in place: abundance of data, cost-effective infrastructure, and a variety of algorithms are available and improving at a rapid pace. Computers have always been recording every event, activity and signal they measure, making data "the exhaust pipe of computing". These data can now be made available to analysis by a new generation of technologies for collecting, storing, and processing. Streaming technologies shorten the time from data creation to storage. Massively parallel processing systems, like Hadoop and Enterprise Data Warehouse, store petabytes of data; and in-memory processing engines like Spark, support interactive massive computations needed for predictive model training.

Predictive modeling algorithms are considered the key to extracting measurable value from big data. Predicting future activity of each customer, based on a wide variety of features, is a promising field that is increasingly utilized by business, government, and education. Known cases include credit scoring and customer churn that predict a customer's likelihood of paying a loan or terminating a service, respectively. Other models predict the customer's future events like growth, success rate, return value, and engagement. The predictions are often done by supervised learning algorithms that are trained on data comprising multiple aspects of the objects tagged with known results of the metrics of interest. A typical model training fits a function of unknown structure to the available data and score each new case based on the learned model. The scores affect actions like admit/reject, fee level, and incentives to offer.

Though technologies for data movement, storage, and parallel computation for training large scale models are readily available, making the process successful often takes R&D, e.g., in the design of new models. Big data is often observational, distributed across multiple sources, sparse, redundant, partially unreliable or skewed. In short, it is messy, which represents a mismatch between Big data and what is expected by the predictive model as input, thereby requiring data exploration, data munging, and feature engineering to transform the data to the right format.

Data and algorithm must be carefully selected to support the score-based actions. Algorithm selection is done under several tradeoffs like score interpretability vs. accuracy, e.g., single decisions tree vs. ensemble. Other tradeoffs include precision vs. recall to match the costs of different errors and excluding data features to eliminate bias in the actions based on the scores. Technologies supporting large scale predictive models will be reviewed and case studies highlighting tradeoffs and approaches for constructing end-to-end models will be described.

Grace Peng

National Institutes of Health - NIBIB

"Challenges with Model and Data Sharing in Biomedical, Biological and Behavioral Systems"

Over the last decade the number and types of computational models being developed for biomedical research has experienced a healthy increase. The biomedical community is beginning to recognize not only the usefulness of models, but the essential role models play to integrate disparate fields of knowledge, identify gaps and present testable hypothesis to drive experiments. Multiscale modeling, in particular, is at the forefront of making a significant impact in biomedical discoveries, applied science, and medicine.

Over the last decade the U.S., Europe, and Japan have promoted several government funding initiatives for modeling the physiome. In the U.S. a confluence of events resulted in the 2003 formation of the Interagency Modeling and Analysis Group (IMAG) and subsequent release of the first interagency solicitation for multiscale modeling of biomedical, biological, and behavioral systems. That solicitation funded 24 projects, creating the Multiscale Modeling Consortium (MSM) in 2006. The IMAG MSM Consortium, now in its 10th year, has over 100 multiscale modeling related projects. During this time many other multiscale modeling initiatives have emerged from the 9 government agencies of IMAG, with over 80 program directors managing programs for modeling and analysis and biomedical, biological and behavioral systems.

One of the main activities of IMAG is to coordinate the MSM Consortium. The MSM Consortium is run by the investigators in the field. Its mission is to grow the field of multiscale modeling in biomedical, biological and behavioral systems, by 1) promoting multidisciplinary scientific collaboration among multiscale modelers; 2) encouraging future generations of multiscale modelers; 3) developing accurate methods and algorithms to cross the interface between multiple spatiotemporal scales; 4) promoting model sharing and the development of reusable multiscale models; and 5) disseminating the models and insights arrived from the models to the larger biomedical, biological, and behavioral research community.

The MSM Consortium is actively addressing many pressing issues facing the multiscale modeling community. Of particular focus are the challenges of model sharing and model translation. Some pertinent questions: How we improve the accessibility of models by the worldwide community? How can we reproduce published simulations? How can we facilitate model reuse? How do we build credible models? How do we integrate models into clinical practice? The presentation will describe some of the MSM activities around these questions, and the latest IMAG funding initiative, Predictive Multiscale Models for Biomedical, Biological, Behavioral, Environmental and Clinical Research, will also be presented.

Brian Uzzi

Northwestern University

"The Age of Information and the Fitness of Ideas in Science and Inventions"

Though science's knowledge base is expanding rapidly, the breakthrough paper rate is narrowing and scientists take longer to make their first discoveries. Breakthroughs are related to how information is recombined, yet it remains unclear how scientists and inventors forage the knowledge base in search of tomorrow's highest impact ideas. Studying 28 million scientific papers and 5 million U.S. patents, we uncover 2 major findings. First, we identify "Darwin's Conjecture," which reveals how conventional and novel ideas are balanced within breakthrough papers. Second, we find an "information hotspot." The hotspot is that cluster of papers of a certain age distribution in the knowledge base that best predict tomorrow's hits. Together, works that combine knowledge according to Darwin's Conjecture or forage in the hotspot double their odds of being in the top 5% or better of citations. These patterns result in over 250 scientific and technology fields, are increasingly dominant, and outperform other predictors of impact, suggesting a universal link between the age of information and scientific discovery.

Santo Fortunato

Aalto University, Finland

"Attention shifts in science"

Millions of publications are now produced each year by scientists around the world, providing historical traces of current research trajectories and quantifiable links to the past. Because an increasing obsolescence rate of scientific knowledge may lead to a systemic 'reinventing the wheel' syndrome, it is important to quantify and understand citation trends in light of the exponential growth of scientific publications and references, which is analogous to monetary inflation in real economies. We focus on how attention in science varies in time. We consider both the decay of attention towards single publications and the concentration of citations across time. We find that nowadays papers are forgotten more quickly than in the past. However, when time is counted in terms of the number of published papers, the rate of decay of citations is fairly independent of the period considered. We also find a narrowing range of attention – both classic and recent literature are being cited increasingly less – pointing to complex social processes that underly shifts in scientific myopia and hyperopia. To better understand how these patterns fit together, we developed a network-based model of the scientific enterprise, featuring growth, the redirection of scientific attention via publications' reference lists, and the crowding out of old literature by the new, and validate the model against several empirical benchmarks. In particular, we show that shifts in the reference age distribution follow directly from sudden perturbations to the growth rate of scientific output – i.e., the new layer of rapid online publications.

Roberta Sinatra

CNS, Central European University (HU) and CCNR, Northeastern University (USA)

"Quantifying and predicting patterns of individual scientific impact"

Despite the frequent use of numerous quantitative indicators to gauge the professional impact of a scientist, little is known about how scientific impact emerges and evolves in time. In this talk we quantify the changes in impact and productivity throughout a career in science and show that impact, as measured by influential publications, is distributed randomly within a scientist's sequence of publications. This random impact rule allows one to formulate a stochastic model that uncouples the effects of productivity, individual ability and luck, unveiling the existence of universal patterns governing the emergence of scientific success. The model assigns a unique individual parameter Q to each scientist, which is stable during a career and accurately predicts the evolution of a scientist's impact, from the h-index to cumulative citations. Finally, we show that the Q-parameter is more predictive of independent recognitions, like prizes, than cumulative citations, h-index or productivity.

Staša Milojević

Indiana University

"The role of small teams in big science era"

Efforts from large collaborative teams, with anywhere between 20 and 1000 members, have led to major scientific and technological breakthroughs, and studies have shown that their work has higher citation impact than the works of small teams and individual researchers. Is small-team mode of knowledge production a thing of the past? Is society wasting resources by continuing to fund individual researchers who primarily work alone or with only a few collaborators? The key to answering these questions lies in understanding the principles that govern team formation and their evolution. I have modeled the sizes of research teams in several fields over the last 60 years, since the emergence of first large teams. According to the model, which is able to reproduce the empirical data remarkably well, each large team originates from a small team. While many small teams stay small, some quickly accumulate additional members proportionally to the past productivity of team members, developing into larger teams, and allowing them to grow even faster. However, even today, the small teams remain the necessary seeds for the formation of larger teams. Furthermore, I show that the topics covered by large teams represent only a subset of themes that small teams work (60% in physics). Thus the small teams appear to be critical in maintaining the intellectual diversity and expanding the frontiers of science, and may serve as the incubators for the topics that big teams work on.

Bill Valdez

Consultants International Group, Inc.

"Overcoming Resistance to Modeling and Simulation as Decision Support Tools"

The 2008 Science of Science Policy (SoSP) Roadmap identified developing decision support tools for science policy makers as one of its highest priorities. Modeling and simulation are among the most advanced decision support tools being developed by the SoSP community, but the acceptance of those tools by senior policy makers has not been as widespread as was originally envisioned by the Roadmap. This presentation will focus on the reasons why that has occurred and how to gain more acceptance for these high-end tools.

Kevin Finneran

National Academies of Sciences, Engineering, and Medicine

"Merging data with narrative to communicate the value and values of research"

The National Academies have an enduring interest in making the public and policymakers aware of the value of research to society and have explored many ways of delivering this message effectively. The institution organized two Sackler colloquia on the Science of Science Communications. An important theme that permeated both was the effectiveness of using engaging narrative in explaining how science works and why it's valuable. Incorporating the human dimension of research builds public trust and credibility by making the public aware of the values that underlie science. Magazine articles such as those by Atul Gawande in the New Yorker and popular books such as The Immortal Life of Henrietta Lacks by Rebecca Skloot have demonstrated that large numbers of non-scientists will read about science if it is communicated in a well-wrought story.

On the analytic front, in recent years, science policy scholars have capitalized on the availability of new data mining and analysis tools to develop a more detailed picture of how science is done, how its results are disseminated, and how we understand the larger social benefits that emerge. The DBASSE report "Capturing Change in Science, Technology, and Innovation: Improving Indicators to Inform Policy" addressed this topic. This work is of great interest to professional analysts, but it is not readily accessible to the general public or policymakers.

The Academies' Committee on Science, Engineering, and Public Policy (COSEPUP) has been concerned that public understanding of the value of fundamental research is particularly tenuous, with possibly serious implications for long-term federal support for this research. The committee has struggled with the question of how to tell the story of how scientific research contributes to important social goals such as a more productive economy, better public health, a more sustainable environment, and enhanced national security.

COSEPUP is planning a SciSIP-funded workshop in September that will assemble experts in the narrative and data-driven approaches with the research experts who serve on COSEPUP to discuss how these various perspectives can be merged to define a template for a type of communication that encompasses the appeal of narrative, the rigor of new analytic data, and the understanding of how science works in practice.

Kaye G. Husbands Fealing

Georgia Tech

"Matchmaking: SciSIP PIs – Meet – Policymakers"

Abstract requested.

Yasushi Hara

National Graduate Institute for Policy Studies

"Scenario-based Economic Model Approach to evaluate the impact of the Internet of Things: For the Creation of Policy Options of Science, Technology and Innovation Policy"

The 21st century marks the prosperity of Internet of Things (IoT) with the stream of technology change that facilitates the production and capital stock of R&D activity that will drastically reshape the social economy structure over the next decades of years. The study aims to develop a recursive dynamic model of science technology, and innovation policy for analyzing social economic impact on multi-sectors including their production, information allocation, and R&D units in order to make the assessments on the science for policies and its social economic factors. By reviewing the economic impact through examining several alternative policy options on the manufacturing sectors with IoT implementation for its information allocation and processioning to accelerate its productivity, the most pressing issues on the overall economic structure could be tackled while the consequences being foreseen. The data used in the model were sourced from input-output table with expansion of the tangible and non-tangible capital investment by considering long/short run block, labor market modeling, value-added and wage determinant, government balance sheet, foreign and the final demand block.

The study interpreted interconnection of exogenous technology scenarios of the policy options to capture the long term structural adjustment and transitions driven by innovation policies in comparison of the baseline of business as usual to derive the impact in the general inter-dependency of economy constituted the multi-sectoral general equilibrium economic model. The model is expected to shed light on the implication of total factor productivity for its process change on the demand side while the productivity improvement in information provision service sector that enlarges the platform-typed business among industries. Such business platform is indispensable for utilizing the cross-sectoral information technology whereas these fundamental factors based on information and system technology of robotic artificial intelligence will construct a new relationship of human kind and machine. The study demonstrated policy options by introducing different level of the processing efficiency index in the activity divisions of marketing, planning, R&D, procurement, operation and sales, conservative, the deviations of economic variables in production process were examined.
The simulation results showed the change on employment and production division along with the IoT advancement of its short/long-run effect. For manufacturing sector, the efficiency improvement would increase the production, public, private R&D investment and consolidate the knowledge stock for the expansion of knowledge infrastructure that rose TFP through the public and private investment, respectively. Moreover, the information management could benefit from outsourcing and externalization, while the cross-sectional platform of information management may thus be established for economic development. Finally, the human resource and higher level professional education revealed an increase especially in information related, service and R&D division in private sector. The dilemma serves as the most challenging direction for the industrial evolution would increase the gap of knowledge that causes technological unemployment remains a concern, the analytical framework in the study is expected to provide evidence-based approach to confront the critical issue by unfolding the process specifically and respond to a trans-formative change agenda for innovation to achieve sustainable development.

Zhao Qu

German Centre for Higher Education and Science Research, Berlin

"What is happening in the new energy vehicles field: a bibliometric review"

New energy vehicle (NEV) is gaining momentum in energy research, and statistical data shows us the importance of this topic for practitioners as well. This paper provides a retrospect to the researches recently completed in NEV's field, and emphatically compares the progress between China and the USA.

The vast NEV literatures published over the past fifteen years are analyzed through bibliometric methods in order to explore the scope of the field, detect current research priorities, and recognize the most prominent papers and authors. This study identifies thematically related clusters of research and shows how the NEV field has evolved through interconnected, yet distinct, subfields. In addition to looking inside the research pattern and evolution path, it explains the differences across countries from the point of national innovation system (NIS) and strategic planning. The paper concludes with a short outlook on the potential frontiers and drawbacks in the on-going NEV's academic research.

Marc Charette

National Heart, Lung, and Blood Institute

"Shifting Demographics among Research Project Grant Awardees at the National Heart, Lung, and Blood Institute (NHLBI)"

The present study has its origins in a series of communications from NHLBI-funded mid-career investigators regarding perceived difficulties in the renewal of their grants. This led us to ask: "Are mid-career investigators experiencing disproportionate difficulties in the advancement of their professional careers?" We find that there has been a significant and evolving shift in the demographics of research project grant (RPG) awardees at NHLBI. Since 1998, the proportion of established investigators (ages 56-70+) receiving research project grant (RPG) awards has been rising in a gradual and linear fashion. At the same time the proportion of early-stage RPG awardees (ages 24-40) fell precipitously until 2006, and then stabilized. During the same period, the proportion of mid-career RPG awardees (ages 41-55), which had been relatively stable through 2006, then began to fall significantly. In examining potential causes of these demographic shifts we have identified certain structural properties within the RPG award process that appear to promote an increasingly older RPG awardee population and continual decreases in the proportion of mid-career RPG investigators. A collateral result of these demographic shifts, when combined with level or declining funding, is a progressive reduction in the number of independent laboratories funded by NHLBI that selectively impacts mid-career investigators.

Kevin Boyack

SciTech Strategies, Inc.

"Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?"

In 1965, Derek de Solla Price foresaw the day when a citation-based taxonomy of science and technology would be delineated and correspondingly used for science policy. A taxonomy needs to be comprehensive and accurate if it is to be useful for policy making, especially now that policy makers are utilizing citation-based indicators to evaluate people, institutions and laboratories. Determining the accuracy of a taxonomy, however, remains a challenge. Previous work on the accuracy of partition solutions is sparse, and the results of those studies, while useful, have not been definitive. In this study we compare the accuracies of topic-level taxonomies based on the clustering of documents using direct citation, bibliographic coupling, and co-citation. Using a set of new gold standards—articles with at least

100 references—we find that direct citation is better at concentrating references than either bibliographic coupling or co-citation. Using the assumption that higher concentrations of references denote more accurate clusters, direct citation thus provides a more accurate representation of the taxonomy of scientific and technical knowledge than either bibliographic coupling or co-citation. We also find that discipline-level taxonomies based on journal schema are highly inaccurate compared to topic-level taxonomies, and recommend against their use.

Arash Baghaei Lakeh

Virginia Tech

"Reproduction of Papers: A Dynamic Cross-Country Analysis of Scientific Production on HIV/AIDS Research"

Our study is a contribution to the science of science production. We employ a dynamic modeling tool (system dynamics) to investigate different growth modes of scientific output across countries. We use academic papers as a proxy for scientific output. As an example case, we have chosen publications in the field of HIV/AIDS. We have gathered data for more than 1,200,000 papers (published between 1984 and 2012) in this field which includes more than 250,000 authorships. A multi-layer system dynamics model is developed to investigate the underlying mechanisms of scientific progress in the case of HIV/AIDS research. In this research, we have introduced a conceptual model for production of academic papers which we called reproduction of academic papers. We have defined this concept with a mathematical model applicable to a generation of academic papers in the scientific community. We have connected this concept with science workforce growth, and international scientific collaborations to explain the patterns of publications in different regions of the world on the topic of HIV/AIDS. The results from our analysis generates insights on different topics of research policy including research priority, distribution of resources, alignment of research with regional challenges, etc.

Giorgio Triulzi

MIT Institute for Data, Systems, and Society

"Predicting Inventors' Movements Across Technology Domains"

Technology influences virtually every aspect of science, economy, and human life, but too little is understood about how new technologies arise or how to rigorously predict them. Technology is invented by inventors, and understanding what leads inventors to invent in a particular technology domain may allow for better prediction of future inventions and technology changes. Analyses of patents are providing increasing understanding of what properties predict successful invention, but there is less consideration of individual inventors' rationales and behaviors, and how the structure of the technology space may condition them to succeed in such inventions. When inventors seek to create new technology they must choose where in the technology space to allocate their efforts and time: Should they work in domains that many others are working on, or in unpopular ones? Should they focus on domains closely related to what they already know, or explore those distant ones from their prior experiences and knowledge? We examine data from 4 million patents and 2.8 million inventors. We created a network map of 629 technology domains and how strongly they relate to each other. We find that inventors are far more likely to successfully enter technology domains that are related to their own individual inventive experience but are also widely popular. Fifty percent of inventors' movements are to the 1% most related and popular domains. We used over a billion data points to build a model to predict individual inventors' future movements in the technology space, which is predictive even when trained only with data from 20 years prior. Moreover, we built a second set of models that predicts inventors' performance in the newly entered domains. We found that inventors who move into related domains patent more than expected when compared to other inventors that start patenting in those domains in the same year, but inventors who successfully move into less related domains receive more citations per patent. Our data-driven predictive models may allow for better forecasting of future inventions and technological changes. Such models are also a tool that inventors can personally use, which uses the inventor's own history to provide a ranked list of domains to consider. These tools may also be most useful for organizations and governments seeking to quantify the expertise of their workforce and what new domains that expertise could be used in.

Victor Yakovenko

University of Maryland

"Economic inequality from statistical physics point of view"

Similarly to the probability distribution of energy in physics, the probability distribution of money among the agents in a closed economic system is also expected to follow the exponential Boltzmann-Gibbs law, as a consequence of entropy maximization. Analysis of empirical data shows that income distributions in the USA, European Union, and other countries exhibit a well-defined two-class structure. The majority of the population (about 97%) belongs to the lower class characterized by the exponential ("thermal") distribution. The upper class (about 3% of the population) is characterized by the Pareto power-law ("superthermal") distribution, and its share of the total income expands and contracts dramatically during booms and busts in financial markets. Globally, data analysis of energy consumption per capita around the world shows decreasing inequality in the last 30 years and a convergence toward the exponential probability distribution, in agreement with the maximal entropy principle. Similar results are found for the global probability distribution of CO_2 emissions per capita. All papers are available at http://physics.umd.edu/~yakovenk/econophysics/. For recent coverage in Science magazine, see http://www.sciencemag.org/content/344/6186/828.

Stephen Kobourov

University of Arizona

"Maps of Computer Science"

Relational data sets are often visualized as networks: objects become the network nodes and relations become the network links. Network visualization algorithms aim to present such data in an effective and aesthetically appealing way. We describe map representations, which provide a way to visualize relational data with the help of conceptual maps as a data representation metaphor. While networks often require considerable effort to comprehend, a map representation is more intuitive, as most people are familiar with maps and ways to interact with them via zooming and panning. We consider map representations of research papers in computer science. Words and phrases from paper titles are the cities in the map, and countries are created based on word and phrase similarity. With the help of heatmaps, we can visualize the profile of a particular conference or journal over a base map of all computer science. Similarly, we can create heatmap profiles for individual researchers or research groups. See http://mocs.cs.arizona.edu for more details.

Scott Dempwolf

University of Maryland, College Park

"Network and temporal models of innovation ecosystems"

Abstract requested.

Mark Orr

Biocomplexity Institute at Virginia Tech

"Information Diffusion on Physical Symbol Systems"

I will discuss a novel program of research on information diffusion that uses human social systems as its basis. Humans, as individual systems, exhibit several properties–motivation, memory, the making of endless predictions, systematic biases, and multiple modes of learning; heterogeneity (individual differences) is the norm that challenges the notion of information diffusion. Bounded by these properties, this fellowship will explore deep questions such as: What is the nature of human information diffusion? And, what are the implications for information diffusion, writ large, and ultimately for complexity itself?

The core orientating framework I propose stems from Herb Simon's conceptualization of artifacts. An artifact is defined, in part, by a system that has a clear boundary or interface to its environment (which may be other artifacts). Physical symbol systems (including computers, minds and brains) are artifacts par excellence. In the human case, the physical symbol system adapted through evolution to solve problems posed by complex environments (both physical and biological in nature). This idea, then, really explores the implications of an abstraction: Diffusion of information in humans is diffusion among a set of artifacts called physical symbol systems.

The secondary, but still central purpose of my program of research, is to use diffusion among physical symbol systems as a window into complexity itself, from first principles so to speak, addressing deep theoretical questions related to nesting/hierarchy, near-decomposability, topological methods, and potentially, the generation of physical symbol systems. These goals may have implications beyond information diffusion into topics such as distributed computing, medicine, and education. Core questions are of this nature: How and under what conditions are physical symbol systems generated? Is a graph-theoretic approach useful for understanding diffusion in physical symbol system, and if so, under what conditions? What is autonomy in distributed information systems?

Julia Laurin

Thomson Reuters

"How can recommendations aid scholarly decision-making?"

This talk will explore how Thomson Reuters applies an augmented intelligence approach to aid researchers with persistent information tasks such as identifying relevant journals for manuscript submission, identifying potential researchers for peer review, and identifying the most salient literature in new research domains. After reviewing solutions currently in production in EndNote and ScholarOne, we will share our latest R&D project in collaboration with the University of Washington. We couple the hierarchical structure of the citation network – which reflects the natural hierarchical structure of scientific domains, fields, subfields, and so forth – with importance scoring based upon a network centrality measure. In this way, we use hierarchical clustering to determine relevance and then recommend papers based upon their importance within these clusters. Thus, we are able to generate a spectrum (or scale) of recommendations for any given topic, paper, or set of key words. We can find papers that are very closely related but perhaps not yet very influential (Expert Recommendations). Alternatively, we can find papers that may be more distantly related but represent foundational contributions to the broader area of research (Classic Recommendations) for researchers new to a field.

Andreas Bueckle (with Katy Börner)

CNS, ILS, Indiana University

"Science Forecasts, S1:E1"

This effort out of Indiana University aims to develop and broadcast "Science Forecasts," a news show that communicates local and global developments in science, technology, and innovation to a general audience. In Spring 2015, a pilot episode was recorded featuring a moderator that explains trends using an animated map of science (analogous to a weather forecast) and a zoom into a specific research result on 'using Twitter for detecting episodes of depression', presented by Johan Bollen, who is interviewed by Fred Cate, both faculty at Indiana University.

Dame Wendy Hall

University of Southampton, UK

"Observatories, data analytics and storm chasing for Web Science research and innovation"

Since the emergence of the World Wide Web in the early 1990s, the Internet has evolved into a critical global infrastructure with vast emergent properties that are transforming society. Web Science is the study of the Web as a socio-technical system. Studying this new eco-system from an interdisciplinary perspective becomes ever more important as the Internet becomes increasingly significant in all our lives and our futures. We are now rapidly moving into a world of data on and about the Web, which gives rise to even more opportunities and challenges. In this talk, we will explore the role of Web

Science in helping us understand the origins of the Web, appreciate its current state and anticipate possible futures in order to address the critical questions that will determine how the Web evolves as a social-technical network. We will discuss the role of observatories, data analytics and computational models in the development of new methodologies for longitudinal research in Web Science, and consider whether it might be possible to forecast future developments by collective action on the sharing and analysis of data associated with major events such as natural disasters, pandemics or even the US election!

Carl Bergstrom

University of Washington

"Why Scientists Chase Big Problems: Individual Strategy and Social Optimality"

Scientists pursue personal recognition as well as collective knowledge. When scientists decide whether or not to work on a hot new topic, they weigh the potential benefits of a big discovery against the costs of setting aside other projects. These self-interested choices can potentially spread researchers across problems in an efficient manner, but efficiency is not guaranteed. We use simple economic models to understand such decisions and their collective consequences. Academic science differs from industrial R&D in that academics often share partial solutions to gain reputation. This convention of "Open Science" is thought to accelerate collective discovery — but extending our model to allow the option of partial publication, we find that it need not do so. The ability to share partial results influences which scientists work on a problem; consequently, Open Science can slow down the solution of a particular problem if it deters entry by important actors.

Robert Axtell

George Mason University/NICO at Northwestern University

"Emergent Technological Epochs and The Policies that Foment Them"

We model the evolution of economic goods and services as a stochastic process of recombination conducted by purposive agents. There is an initial 'seed set' of goods, each having an intrinsic economic fitness, which is assessed subjectively by individual agents. Each agent adopts some subset of these goods based on its subjective perceptions. The innovation process proceeds by individual agents attempting to invent new goods by combining current goods that it uses. Such attempts at innovation are mostly unsuccessful, insofar as they lead to goods having low economic fitness. However, some inventions do have fitness exceeding other goods in the economy and can therefore be adopted by one or more agents. The adoption process proceeds with each agent considering the new good, developing an idiosyncratic assessment of its value, and accepting it only if this value exceeds the least valuable item it currently holds, in which case it sheds the least valuable item from its holdings. If no agent adopts the new product, then it is considered unsuccessful and has no further possibility of being adopted. If all agents drop a particular good then it no longer exists in the economy and cannot be brought back into existence, unless it is reinvented, which is possible. This simple model has a variety of robust properties. First, agent welfare is monotonically increasing in the model, since agents only adopt subjectively superior products over time. Second, the population of goods is transient, with the initial seed set becoming extinct eventually, and all goods having finite lifetimes. We compare the distribution of invention lifetimes in the model with stylized facts. Third, the number of new inventions adopted by one or more agents is very volatile, and displays clustered volatility. Fourth, the total number of goods in the economy over time is very irregular, displaying periods of relative stasis in which few inventions are successful, punctuated by periods of rapid technological progress in which there is dramatic change in the goods being used. These episodes of technological change are often instances of 'creative destruction' insofar as one or a few successful inventions can lead to the extinction of some larger number of other, older goods. Finally, each realization of the model generates a graph of technological antecedents that has somewhat peculiar properties, which we characterize, and from which the lineages of current technologies can be readily determined. The relation of this model to other models of evolution is described. While this model is rather abstract, it is, in many ways, much more realistic than conventional economic models. Its manifold policy conclusions will be discussed.

James Evans

University of Chicago

"How Science Thinks (and How to Think Better)"

I explore how modeling the scientific process can create opportunities for improving it. I begin by demonstrating how the complex network of modern biomedical science provides a substrate on which a scientist–-and indeed science as a whole-–thinks, and its consequences for ongoing scientific discovery and human health. Using millions of scientific articles from MEDLINE, I show how science moves conservatively from problems posed and questions answered in one year to those examined in the next. Along the way, I show how contemporary science "changes its mind"; how it has become more risk-averse and less efficient at discovery; and how the atmosphere of its own internal puzzles have largely decoupled it from health needs. We use this as an opportunity to demonstrate how much more efficient strategies can be found for mature fields, which involve greater individual risk-taking than the structure of modern scientific careers supports, and propose institutional alternatives that maximize a range of valuable objectives, from scientific discovery to robust understanding to technological advance.

Mark Gerstein

Yale University

"Analyzing the Structure of Genomic Science"

The emergence of collective creative enterprise, such as large scientific consortia, is a unique feature in modern scientific research, especially in genomic areas. Recent examples include the ENCyclopedia of DNA Elements (ENCODE) consortium, annotating the human genome and the 1000 Genomes consortium, generating a catalog of uniformly called variants for the biomedical community. To ensure that the scientific community can benefit from these efforts, it is important to understand the connections between consortium members and researchers outside of the consortium. To address the issue, we analyzed the temporal co-authorship network structures of ENCODE and modENCODE consortia. Our analysis revealed their publication patterns showing that the consortium members work closely as a community whereas non-members collaborate in the scale of a few laboratories. We also identified a few brokers playing an important role to facilitate collaborations with outside researchers, which suggests that large scientific consortia should set up formal outreach groups to communicate with outside researchers.

Daifeng Wang, Koon-Kiu Yan, Joel Rozowsky, Eric Pan, Mark Gerstein, Temporal dynamics of collaborative networks driven by large scientific consortia, Trends in Genetics, 2016 http://dx.doi.org/10.1016/j.tig.2016.02.006.

D. Biographies of Presenters

The conference brought together leading experts from economics, social science, scientometrics and bibliometrics, information science, physics, science policy, and other scholarly disciplines that develop or use mathematical, statistical, and computational models of different types to inform STI (policy) making. Biographies of presenters are given here on alphabetical order.

Petra Ahrweiler

Petra Ahrweiler is the Director of the European Academy of Technology and Innovation Assessment, a joint research centre of the Federal German state of Rhineland-Palatinate and the German Aerospace Center. Ahrweiler also holds a professorship for Technology and Innovation Assessment at Johannes Gutenberg University Mainz in Germany. Her main research interests are innovation networks in knowledge-intensive sectors such as ICT and biotech, issues of science in society, responsible research and innovation, and policy modelling for complex social systems using methods such as social network analysis and agent-based simulation. She has long experience as principal investigator and co-ordinator of international projects on innovation networks, for example the EU-projects on “Simulating Self-Organizing Innovation Networks (SEIN)”, “Network Models, Governance, and R&D Collaboration Networks” (NEMO) or “Governance of responsible Research and Innovation” (GREAT). Ahrweiler holds various research awards and is member of a number of advisory boards in both governmental and academic organisations.

Robert Axtell

Robert Axtell holds a public policy degree and works at the intersection of computer science, economics, game theory, artificial intelligence, finance, and politics, focusing on modeling human social phenomena in software. In his work, he tries to use all available data in models which may run at full (1-to-1) scale with the phenomena under study. His MIT Press book, “Growing Artificial Societies: Social Science from the Bottom Up,” is a citation classic. His forthcoming book, “Dynamics of Firms from the Bottom Up: Data, Theories and Models” by The MIT Press is due out at the end of 2017; it uses data on all six million American firms that have employees to synthesize a model of the entire U.S. private sector and draws many important policy conclusions. Axtell has also built models of science and innovation processes.

Arash Baghaei Lakeh

Arash Baghaei Lakeh is a PhD candidate in the Grado Department of Industrial & Systems Engineering at Virginia Tech. He is interested in tackling complex interdisciplinary problems which involve social, economic, technical, cultural, and environmental aspects. He has used dynamic modelling tools as well as experimental design in the past. He has worked on a variety of topics including science policy, complex systems, health care, and safety.

Philip Beesley

Philip Beesley is a practicing visual artist, architect, and Professor in Architecture at the University of Waterloo and Professor of Digital Design and Architecture & Urbanism at the European Graduate School. Beesley's work is widely cited in contemporary art and architecture, focused in the rapidly expanding technology and culture of responsive and interactive systems. Beesley was educated in visual art at Queen's University, in technology at Humber College, and in architecture at the University of Toronto. He serves as the Director for the Living Architecture Systems Group, and as Director for Riverside Architectural Press. His Toronto-based practice, Philip Beesley Architect Inc., combines the disciplines of professional architecture, science, engineering, and visual art. The studio's methods incorporate industrial design, digital prototyping, instrument making, and mechatronics engineering. His work was selected to represent Canada at the 2010 Venice Biennale for Architecture, and has received distinctions including the Prix de Rome, VIDA 11.0, FEIDAD, Azure AZ, and Architizer A+.

Carl Bergstrom

Carl T. Bergstrom is a Professor in the Department of Biology at the University of Washington. Bergstrom's research uses mathematical, computational, and statistical models to understand how information flows through biological and social systems. His recent projects include contributions to the game theory of communication and deception, use of information theory to the study of evolution by natural selection, game-theoretic models and empirical work on the sociology of science, and development of mathematical techniques for mapping and comprehending large network datasets. In the applied domain, Bergstrom's work illustrates the value of evolutionary biology for solving practical problems in medicine and beyond. These problems include dealing with drug resistance, handling the economic externalities associated with anthropogenic evolution, and controlling novel emerging pathogens such as the SARS virus, Ebola virus, and H5N1 avian influenza virus. He is the coauthor of the college textbook "Evolution," published by W. W. Norton and Co., and teaches undergraduate courses on evolutionary biology, evolutionary game theory, and the importance of evolutionary biology to the fields of medicine and public health. Bergstrom received his Ph.D. in theoretical population genetics from Stanford University in 1998; after a two-year postdoctoral fellowship at Emory University, where he studied the ecology and evolution of infectious diseases, he joined the faculty at the University of Washington in 2001.

Katy Börner

Katy Börner is the Victor H. Yngve Distinguished Professor of Information Science in the Department of Information and Library Science, School of Informatics and Computing, Adjunct Professor at the Department of Statistics in the College of Arts and Sciences, Core Faculty of Cognitive Science, Research Affiliate of the Center for Complex Networks and Systems Research and Biocomplexity Institute, Member of the Advanced Visualization Laboratory, Leader of the Information Visualization Lab, and Founding Director of the Cyberinfrastructure for Network Science Center at Indiana University in Bloomington, IN. She is a Visiting Professor at the Royal Netherlands Academy of Arts and Sciences (KNAW) in The Netherlands and a Visiting Professor (Gastprofessur) and Mercator Fellow in User-Centered Social Media, at the University of Duisburg-Essen in Germany. She is a curator of the international *Places & Spaces: Mapping Science* exhibit. She became an American Association for the

Advancement of Science (AAAS) Fellow in 2012 and an Royal Society for the Encouragement of Arts, Manufactures and Commerce (RSA) Fellow in 2016.

Kevin Boyack

Kevin W. Boyack is President of SciTech Strategies, Inc., and has been with the company since summer of 2007. Prior to this, he spent 17 years at Sandia National Laboratories where he worked in various areas including combustion (experimental and modeling), transport processes, socio-economic war gaming, and science mapping. Since joining SciTech, his work has centered on developing more accurate global maps of science. He has published over 40 articles dealing with various aspects of science mapping and related metrics. Current interests include detailed mapping of the structure and dynamics of science, technology, and altruism; merging of multiple data types and sources; and development of advanced metrics.

Andreas Bueckle

Andreas Bueckle is a Ph.D. student in Information Science at Indiana University as well as a professional videographer and photographer. His academic interests revolve around visual learning systems for programming. In that context, he currently does research on how video games function as feedback systems for game programmers. To that end, he co-teaches a novel class on video game programming at Indiana University. As a professional videographer and photographer, he has been hired by various clients for the past six years. He has worked on video and photo projects on three continents, with a focus on documentary as well as nature, especially social issues and nature photography. Check out work samples on andreas-bueckle.com.

Marc Charette

Marc Charette is a Program Director in the Cardiovascular Sciences Division at the National Heart, Lung, and Blood Institute (NHLBI). Charette's responsibilities include management of a portfolio of extramural grants and coordination of the Vascular Interventions/Innovations and Therapeutic Advances (VITA) Program. VITA is an NHLBI translational initiative that seeks to stimulate early-stage biomedical product development. Charette is particularly interested in the processes that promote the translation of basic research discoveries into promising product candidates and in portfolio analysis and the development of policies to promote high impact science. He has a Ph.D. in genetics from the University of Chicago and did postdoctoral research at the Harvard Medical School. Prior to NHLBI, Charette was an executive in the biotech industry developing novel drug candidates for various diseases and disorders.

William Colglazier

Dr. E. William Colglazier is Editor-in-Chief of Science & Diplomacy and Senior Scholar in the Center for Science Diplomacy at the American Association for Advancement of Science. He served as the fourth Science and Technology Adviser to the U.S. Secretary of State from 2011 to 2014. From 1994 to 2011, he was Executive Officer of the U.S. National Academy of Sciences (NAS) and National Research Council

(NRC). In 2015 he received the Burton Forum Award from the American Physical Society for "outstanding contributions to the public understanding or resolution of issues involving the interface of physics and society" and the Order of the Rising Sun, Gold Rays with Neck Ribbon, from the Japanese government. In 2016, he became co-chair of the 10-member committee appointed by the UN Secretary General to support the Technology Facilitation Mechanism to promote the role of science, technology, and innovation for achieving the Sustainable Development Goals.

Matteo Convertino

Matteo Convertino is the principal investigator of the HumNat Lab. He is involved in the promotion of complexity science and engineering design of natural and human systems for population health. In a broader perspective, this effort is committed to the diagnosis, etiognosis, and prognosis of diseases via smart and multiscale global system science and art-in-science. Convertino's deep interest is in the identification of the fundamental factor interactions (”processes”) leading to observed patterns by integrating system biology/ecology (with particular focus on the environmental dynamics, e.g. ecohydrological dynamics, in systemic macro-epidemiology) and in the translation of that knowledge to applications for stakeholders via decision science and engineering methods. The quest for universalities, system states, and state transitions via tipping points is a key in the research of Convertino. Theories and models that have been developed are: Optimal Transmission Networks, Morphological Effective Systemic Epigraph, Information-theory based Global Sensitivity and Uncertainty Analysis, MaxEnt Model in geomorphology and epidemiology, Portfolio Decision Models for Enhanced Adaptive Management, Reverse Engineering Traceback Model, and Game-based Mental Modeling. The conversion of these models to software (STEM and DECERNS) is ongoing as well as ”science as art” initiatives. Fun fact: How did Convertino get interested into this? His hometown: Venice! He started to be interested in the design of bridges, the dynamics of water ecosystems, and later on in people dynamics and how that can be analyzed by combining methods used to design bridges and water ecosystems. In one word: connectomics, i.e., how everything is connected to everything.

Scott Dempwolf

C. Scott Dempwolf is an Assistant Professor in the Urban Studies and Planning Program at the University of Maryland, College Park, and Director of the UMD—Morgan State Center for Economic Development. Dempwolf is also affiliated with the National Center for Smart Growth. His research focuses on understanding innovation ecosystems including the networks of people and organizations that comprise them and the activities that they engage in. Scott uses network and temporal analysis tools to visualize and analyze these innovation ecosystems and their patterns of innovation activities. In addition to aiding in the development of new innovation metrics, this research has applications in research portfolio management, technology transfer, economic development, public policy and corporate strategy. In 2015, Dempwolf founded Tertius Analytics, LLC, a consulting firm with a longer-term goal of commercializing applications using his unique analytic technologies. Before pursuing his Ph.D. in Urban and Regional Planning at the University of Maryland, Dempwolf practiced community and economic development for over 20 years at the neighborhood, city, county, and regional levels. He joined the UMD faculty in 2012. Dempwolf holds a Masters in Community and Regional Planning at Temple University; and a Bachelor's from the Massachusetts Institute of Technology.

James Evans

James Evans is Associate Professor of Sociology at the University of Chicago, member of the Committee on the Conceptual and Historical Studies of Science, Senior Fellow at the Computation Institute, Director of Knowledge Lab (knowledgelab.org) and Director of the Computational Social Science program (macss.uchicago.edu). His work explores the sources, structure, dynamics, and consequences of modern knowledge. Evans is particularly interested in the relation of markets to science and knowledge more broadly, and how evolutionary and generative models can inform our understanding of collective representations, experiences and certainty. He has studied how industry collaboration shapes the ethos, secrecy and organization of academic science; the web of individuals and institutions that produce innovations; and markets for ideas and their creators. Evans has also examined the impact of the Internet on knowledge in society. His work uses natural language processing, the analysis of social and semantic networks, statistical modeling, and field-based observation and interviews. Evans' research is funded by the National Science Foundation, the National Institutes of Health, the Mellon and Templeton Foundations and has been published in *Science, PNAS, American Journal of Sociology, American Sociological Review, Social Studies of Science, Administrative Science Quarterly* and other journals. His work has been featured in *Nature, The Economist, Atlantic Monthly, Wired, NPR, BBC, El Pais, CNN* and many other outlets.

Kaye G. Fealing

Kaye Husbands Fealing is an economist who comes to the Ivan Allen College of Liberal Arts from the Hubert H. Humphrey School of Public Affairs at the University of Minnesota. During the course of her career, she has built a distinguished record of achievements in scholarship and education, as well as in national and international leadership and service. Her areas of expertise include international trade policy; science, technology, and innovation policy in specific contexts; knowledge generation and the development of networks. Husbands Fealing developed models to measure science innovation and to measure the impacts of market forces and policy on the access of women and minorities to employment and careers in science, technology, engineering, and mathematics (STEM) areas. She has held named professorships at two institutions and served as president of the National Economic Association. She developed the National Science Foundation's (NSF) Science of Science and Innovation Policy program and co-chaired the Science of Science Policy Interagency Task Group. At NSF, she also served as an economics program director. She was a visiting scholar at Massachusetts Institute of Technology's Center for Technology Policy and Industrial Development, where she conducted research on NAFTA's impact on the Mexican and Canadian automotive industries and research on strategic alliances between aircraft contractors and their subcontractors.

Kevin Finneran

Kevin Finneran is director of the Committee on Science, Engineering, and Public Policy (COSEPUP) at the National Academy of Sciences (NAS) in Washington, DC, and editor-in-chief of *Issues in Science and Technology*, a quarterly policy magazine published jointly by NAS, Arizona State University, and the

University of Texas at Dallas. Previously, Finneran was Washington editor of *High Technology* magazine, a correspondent for the *London Financial Times* energy newsletters, and a consultant on science and technology policy. His clients included the National Science Foundation, the Office of Technology Assessment, the U.S. Agency for International Development, and the Environmental Protection Agency. He is a fellow of the American Association for the Advancement of Science and the author of "The Federal Role in Research and Development" published by the National Academy Press in 1985 and a contributing author to "Future R&D Environments: A Report to the National Institute of Standards and Technology," National Academy Press, 2002.

Susan Fitzpatrick

Susan M. Fitzpatrick is President of the James S. McDonnell Foundation, St. Louis, Missouri. The McDonnell Foundation is one of a limited number of international grant-makers supporting university-based research in biological, behavioral, and complex systems sciences through foundation-initiated programs. As President, Fitzpatrick serves as JSMF's Chief Executive Officer. Fitzpatrick received her Ph.D. in Biochemistry and Neurology from Cornell University Medical College (1984) and pursued post-doctoral training with in vivo NMR spectroscopic studies of brain metabolism/function in the Department of Molecular Biochemistry and Biophysics at Yale University. Fitzpatrick joined the James S. McDonnell Foundation in 1993 as the Foundation's first Program Officer. She was promoted to Program Director in 1997 and to Vice President in 2000. Fitzpatrick is an adjunct associate professor of Neurobiology and Anatomy and Occupational Therapy at Washington University School of Medicine (St. Louis) and teaches neuroscience in both lectures and seminars. Fitzpatrick lectures and writes on issues concerning applications of neuroscience to clinical problems, the translation of cognitive science to educational settings, the role of private philanthropy in the support of scientific research, and on issues related to the public dissemination of and understanding of science. Fitzpatrick serves on the board of the Ontario Brain Institute, is a member of the American Occupational Therapy Foundation Science Council, and is a member of International Advisory Council of the Rotman Institute for Philosophy. Fitzpatrick is a past member of the board of the American Association for the Advancement of Science, the American Occupational Therapy Foundation, and is a Past-President and former Chair of the Board of the Association for Women in Science.

Santo Fortunato

Santo Fortunato is Professor of Complex Systems at the Department of Computer Science of Aalto University, Finland. Previously he was director of the Sociophysics Laboratory at the Institute for Scientific Interchange in Turin, Italy. Fortunato got his Ph.D. in Theoretical Particle Physics at the University of Bielefeld in Germany. He then moved to the field of complex systems, via a postdoctoral appointment at the School of Informatics and Computing at Indiana University. His current focus areas are network science, especially community detection in graphs, computational social science and science of science. His research has been published in leading journals, including *Nature, PNAS, Physical Review Letters, Reviews of Modern Physics, Physics Reports* and has collected about 15,000 citations (Google Scholar). His review article "Community detection in graphs" (Physics Reports 486, 75-174, 2010) is the most cited paper on networks of the last years. He received the Young Scientist Award for

Socio- and Econophysics 2011, a prize given by the German Physical Society, for his outstanding contributions to the physics of social systems.

Richard Freeman

Richard B. Freeman is Ascherman Professor of Economics at Harvard University. He directs the Science and Engineering Workforce Project at the National Bureau of Economic Research, is Faculty Co-Director of the Labor and Worklife Program at the Harvard Law School, and Co-Director of the Harvard Center for Green Buildings and Cities. His research interests include the job market for scientists and engineers; the transformation of scientific ideas into innovations; Chinese labor markets; income distribution and equity in the marketplace; forms of labor market representation, and shared capitalism.

Mark Gerstein

Mark Gerstein, Ph.D. is the Albert Williams Professor of Biomedical Informatics. His lab (http://gersteinlab.org) was one of the first to perform integrated data mining on functional genomics data and to do genome-wide surveys. His tools for analyzing motions and packing are widely used. Most recently, he has designed and developed a wide array of databases and computational tools to mine genome data in humans, as well as in many other organisms. He has worked extensively in the 1000 genomes project in the SV and FIG groups. He also worked in the ENCODE pilot project and currently works extensively in the ENCODE and modENCODE production projects. He is also a co-PI in DOE KBase and the leader of the Data Analysis Center for the NIH exRNA consortium. In these roles, Gerstein has designed and developed a wide array of databases and computational tools to mine genomic data in humans as well as in many other organisms.

Kenneth Gibbs

Kenneth (Kenny) Gibbs, Jr. is a Program Analyst in the Office of Program Planning, Analysis and Evaluation (OPAE) at the National Institute of General Medical Sciences (NIGMS). Prior to joining NIGMS, Dr. Gibbs was a Cancer Prevention Fellow at the National Cancer Institute, and an AAAS Science & Technology Policy Fellow at the National Science Foundation in the Directorate for Education and Human Resources (EHR). Dr. Gibbs completed his Ph.D. in the Immunology program at Stanford University, and received his B.S. in biochemistry & molecular biology from the University of Maryland, Baltimore County.

Jerome Glenn

Jerome C. Glenn co-founded and directs The Millennium Project, a leading global participatory think tank, which produces the State of the Future reports for the past 20 years, Futures Research Methodology 3.0, and the Global Futures Intelligence System. He invented the “Futures Wheel”, a futures assessment technique and concepts such as conscious-technology, transInstitutions, tele-nations, management by understanding, feminine brain drain, just-in-time knowledge, nodes as a management concept for interconnecting global and local views and actions, and definitions of environmental security, collective Intelligence, and scenarios. He wrote about information warfare in the late 1980s in his book “Future Mind.” He sent his first email in 1973 and was hired by the Quakers’ action arm to help organize the environmental movement New England 1971. In the mid-1980s, he was instrumental in getting x.25 packet switching in 29 developing countries which was key to their later getting low cost access to the Internet. More recently, he led the design and implementation of collective intelligence systems for the Global Climate Change Situation Room in South Korea, the Prime Minister’s Office of Kuwait, and now the Global Futures Intelligence System and ECISIS for Egypt. Other current work includes: Future Work/Technology 2050; the EC’s 2050 scenarios on innovation, research, and higher education; and the public’s roles in preventing individuals from deploying future weapons of mass destruction. He was instrumental in naming the first Space Shuttle the Enterprise and banning the first space weapon (FOBS) in SALT II. He has published over 150 future-oriented articles, spoken to over 300 organizations, written several books, including “Future Mind,” “Linking the Future,” and co-author of “Space Trek,” and more information on his research is available at www.millennium-project.org.

Daniel Goroff

Daniel Goroff is Vice President and Program Director at the Alfred P. Sloan Foundation, a private philanthropy that supports breakthroughs in science, technology, and economics. He is Professor Emeritus of Mathematics and Economics at Harvey Mudd College, where he served as Vice President for Academic Affairs and Dean of the Faculty. Before that, he was a faculty member at Harvard University for over twenty years. Daniel Goroff has twice worked for the President’s Science Advisor in the White House Office of Science and Technology Policy, most recently as Assistant Director for Social, Behavioral, and Economic Sciences.

Dame Wendy Hall

Wendy Hall, DBE, FRS, FREng is Professor of Computer Science at the University of Southampton, UK, and was Dean of the Faculty of Physical Science and Engineering from 2010 to 2014. She was Head of the School of Electronics and Computer Science (ECS) from 2002 to 2007. One of the first computer scientists to undertake serious research in multimedia and hypermedia, she has been at its forefront ever since. The influence of her work has been significant in many areas including digital libraries, the development of the Semantic Web, and the emerging research discipline of Web Science. She is currently the Executive Director of the Web Science Institute at Southampton and Managing Director of the Web Science Trust. She was President of the ACM from 2008-2010, a member of the UK Prime Minister’s Council for Science and Technology from 2004-2010 and a founding member of the Scientific Council of the European Research Council. She is currently a member of the Global Commission on Internet Governance and the World Economic Forum’s Global Council on AI and Robotics. She holds many fellowships including Fellow of the Royal Society, Fellow of the Royal Academy of Engineering, and Fellow of the ACM. She is currently the Kluge Center Chair of Technology and Society 2016 at the Library of Congress.

Yasushi Hara

Yasushi Hara holds an Associate's Degree from the Toyota National College of Technology (2004), a B.A. in Graduate School for Economics (2006), and a M.A. in Faculty of Economics, Hitotsubashi University (2009). Hara was aDC1 Fellow of the Japanese Society for the Promotion of Science from 2009-2012 and a Research Associate at the Institute of Innovation Research, Hitotsubashi University from 2012-2015. Hara worked as a specialist at the Science for RE-Designing Science, Technology and Innovation Policy Center (SciREX Center) at the National Graduate Institute for Policy Studies from 2015 and worked at the IT startup Clara Online, Inc. from 2002 to 2009. Hara's research focuses on the social construction of technology and the role of scientists, international comparative analysis for national innovation systems, and evaluating the economic impact of ICT systems.

Bruce Hecht

Bruce Hecht received M.A.Sc. and B.A.Sc. degrees in Electrical Engineering from the University of Waterloo, Ontario, Canada. Originally from Montreal, Quebec, he joined Analog Devices in 1994, where he is currently with the Worldwide Quality Systems Engineering Group in Wilmington, MA, USA. His interests are in the design of all kinds of electronic systems for medical, automotive, industrial, consumer, and communications systems. View his website at http://www.analog.com.

Ian Hutchins

Ian Hutchins is a Data Scientist in the Office of Portfolio Analysis within the Division of Program Coordination, Planning, and Strategic Initiatives, Office of the Director, National Institutes of Health. He leads development teams to make scientific portfolio analysis tools; conducts trans-NIH portfolio analysis using bibliometrics, statistical programming, and text mining; and teaches scientific portfolio analysis courses for agency staff. Prior to this, he investigated the self-assembly of neural circuits in models of neurological disorders and of healthy brain development. He holds a Ph.D. in Neuroscience and a B.S. in Genetics from the University of Wisconsin-Madison.

Richard Ikeda

Richard Ikeda is the director of the Office of Research Information Systems (ORIS) in the NIH OD Office of Extramural Research. In this position, he oversees the operations of the NIH electronic Research Administration (including eRA's IMPAC II and Commons systems), which support the mission-critical function of grants administration for NIH and other federal granting components. He also manages the ORIS Office of Data Quality, which is responsible for the integrity of the data stored in IMPAC II, and the Research, Condition, and Disease Categorization Program, which is responsible for categorizing NIH research activities. Prior to joining NIH in 1999 as an NIGMS program director, with a portfolio of research in the fields of enzymology and wound healing, Rick served on the faculty of the Department of Chemistry and Biochemistry at the Georgia Institute of Technology (Georgia Tech). He has a Ph.D. in Chemistry from the California Institute of Technology (Caltech).

Stephen Kobourov

Stephen Kobourov is a Professor of Computer Science at the University of Arizona. He completed B.S. degrees in Mathematics and Computer Science at Dartmouth College in 1995, and a Ph.D. in Computer Science at Johns Hopkins University in 2000. He has worked as a Research Scientist at AT&T Research Labs, a Humboldt Fellow at the University of Tubingen in Germany, and a Distinguished Fulbright Chair at Charles University in Prague.

Maria Larenas

Maria I. Larenas has extensive expertise in the analysis and management of complex data sets as well as a solid understanding of the theoretical issues in areas of labor economics and economics of education. She had worked as a research economist in different organizations such as the research group of the Nobel Laureate James Heckman, RCF Economic and Financial Consulting in Chicago, SENDA (Chilean government agency of prevention and rehabilitation of alcohol and drugs abuse), and International institutions (IADB and World Bank). Larenas holds a B.A. and M.A. in Economics from her native country (Chile) and a M.A. in Public Policy from the University of Maryland. She is interested in the analysis and policy implications of school-to work transitions and their impacts on long-term outcomes.

Julia Laurin

Julia Laurin is Vice President, Head of Product Management at Thomson Reuters Intellectual Property & Science where she oversees product strategy for Government & Academia solutions including Web of Science, InCites, Journal Citation Reports, and Converis. As Head of Product Management, Laurin works on crafting software, content and services solutions that support the scholarly eco-system. Laurin has a strong professional interest in the potential of multi-institution collaborations and industry initiatives such as VIVO and ORCID to accelerate innovation, as well as R&D efforts to support visibility and greater understanding into the structure and dynamics of science. Before joining Intellectual Property & Science, Julia worked in product development for the Legal division of Thomson Reuters, where she focused on expert witness products and services. Laurin holds a J.D. from UC Berkeley and B.A. in International Relations from Wellesley College. She lives in Pittsburgh, Pennsylvania, with her husband and three daughters.

Kalev Leetaru

Kalev H. Leetaru is one of Foreign Policy Magazine's Top 100 Global Thinkers of 2013. He is a Senior Fellow at the George Washington University Center for Cyber & Homeland Security and a member of its Counterterrorism and Intelligence Task Force. In 2015-2016, he worked as a Google Developer Expert for the Google Cloud Platform. From 2013-2014, he was the Yahoo! Fellow in Residence of International Values, Communications Technology & the Global Internet at Georgetown University's Edmund A. Walsh School of Foreign Service, where he was also an Adjunct Assistant Professor. His work has been profiled in *Nature, The New York Times, The Economist, BBC, Discovery Channel* and the presses of more than 100 nations, while he has been an invited speaker throughout the globe, from the United Nations to the Library of Congress, Harvard to Stanford, Sydney to Singapore. In 2011, *The Economist* selected his Culturomics 2.0 study as one of five science discoveries deemed the most significant developments of 2011. Leetaru's work focuses on how innovative applications of the world's largest datasets, computing platforms, algorithms and mind-sets can reimagine the way we understand and interact with our global world. More on his latest projects can be found at http://www.kalevleetaru.com or http://blog.gdeltproject.org.

Alina Lungeanu

Alina Lungeanu is a Research Scientist at the Population Research Institute, Pennsylvania State University. Her research focuses on understanding the social dynamics of collaboration and its impact on performance in various contexts including science, health communities, and business. To address these issues, Lungeanu uses computational social science approaches including advanced social network analytic techniques, agent based modeling, machine learning, and text analytics.

Guru Madhavan

Guru Madhavan, Ph.D., is a program director at the National Academies of Sciences, Engineering, and Medicine where he has led the R&D of SMART Vaccines—a prioritization software tool to help reduce barriers for vaccine innovation. He serves as a technical adviser to the U.S. Department of Health and Human services in the development of a fully web-based SMART Vaccines 2.0. Madhavan received his M.S. and Ph.D. in biomedical engineering, and an M.B.A. from the State University of New York. His professional experience includes working in the medical device industry as a research scientist developing cardiac surgical catheters for ablation therapy. He is a vice-president of IEEE-USA of IEEE, and has received the Innovator Award and the Cecil Medal from the presidents of the National Academies. He has been named as a distinguished young scientist by the World Economic Forum. Madhavan has co-edited six books, and is author of “Applied Minds: How Engineers Think” published by W.W. Norton.

Stephen Marcus

Stephen Marcus is a program director in the Division of Biomedical Technology, Bioinformatics, and Computational Biology. Marcus is developing and leading a new extramural research program focused on social and behavioral modeling, but also looking more broadly at systems science approaches to health. He also leads predoctoral T32 and K99/R00 Pathway to Independence Awards in biostatistics and bioinformatics and computational biology and the Mentored Quantitative Research Career Development Awards (K25) program. Before coming to NIGMS, he served as a senior epidemiologist in the Tobacco Control Research Branch, Behavioral Research Program, Division of Cancer Control and Population Sciences for the National Cancer Institute. He earned a B.S. in public health engineering and health sciences from Northwestern University, an M.P.H. in epidemiology and medical care administration from the University of Michigan, Ann Arbor, and a Ph.D. in epidemiology from the University of North Carolina at Chapel Hill.

Julie Mason

Julie Bronder Mason is the Associate Director of the Center for Cancer Training (CCT) at the National Cancer Institute (NCI), where she leads training program development and evaluation, workforce analysis, and strategic planning. With over 12 years of experience at the NCI, Mason recently led a labor and skills gap analysis of the U.S. biomedical research workforce. Before joining CCT, she was a health science analyst in NCI's Office of Science Planning and Assessment, where she managed program evaluations and performance reporting. Mason earned a Ph.D. in Pharmacology and Toxicology from the Medical College of Virginia, and performed postdoctoral research in the NCI Laboratory of Molecular Pharmacology. Several notable awards Mason received include the NIH Plain Language Award, Award of Merit for service on the NIH American Recovery and Reinvestment Act committee, and the American Evaluation Association's Best Paper Award for her career development program outcome evaluation.

Daniel McFarland

Daniel McFarland studies the social and organizational dynamics of educational systems like schools, classrooms, universities, and intellectual disciplines. In particular, he performed a series of studies on classroom organization and interaction; on the formation of adolescent relationships, social structures, and identities; on interdisciplinary collaboration and intellectual innovation; and on relational sociology. McFarland's broad research interests have drawn him into a variety of interdisciplinary collaborations with linguists and computer scientists which in turn has led to studies of big data and methodological advances in social networks and language modeling.

Martin Meltzer

Martin I. Meltzer is the Lead of the Health Economics and Modeling Unit (HEMU) and a Distinguished Consultant in the Division of Preparedness and Emerging Infections at the Centers for Disease Control and Prevention (CDC) in Atlanta, GA. He received his undergraduate at the University of Zimbabwe and his graduate degrees from Cornell university. He led the modeling teams supporting CDC's response to the 2009 H1N1 influenza pandemic, including producing monthly estimates of cases, hospitalizations and deaths, as well as estimating impact of the vaccination program and use of influenza anti-viral drugs. Among others, he led the modeling activities that estimated the residual risk associated with the 2012 contaminated steroid injectable products that caused fungal meningitis among patients and CDC's 2014 West Africa Ebola Response Modeling Unit. Examples of his research include estimating the impact of influenza pandemics, the modeling of potential responses to smallpox as a bioterrorist weapon, and assessing the economics of controlling diseases such as rabies, dengue, hepatitis A, meningitis, Lyme, and malaria. He is an associate editor for *Emerging Infectious Diseases*. At CDC, he supervises a number of post-doctoral health economists.

Staša Milojević

Staša Milojević is an Associate Professor of Informatics, a core faculty of Cognitive Science, and a fellow of the Rob Kling Center for Social Informatics at Indiana University, Bloomington. She is co-founder and past chair of ASIST SIG METRICS and serves on the editorial boards of *Scientometrics* and *BioScience*. Her research endeavors to elucidate the dynamics of science as a social and an intellectual (cognitive) endeavor by approaching science as a heterogeneous system comprised of people, ideas, documents, instruments, institutions, and policies situated in an historical context. Specifically, she strives to produce comprehensive, yet straightforward models based on theoretical and empirical findings from a wide range of fields (science, technology, and society (STS), science of science, information science, network science, economics, sociology, philosophy, history, etc.) in order to uncover and explain the fundamental principles that govern contemporary science. She received a Ph.D. in Information Studies at University of California, Los Angeles.

Dan Mote

C. D. Mote, Jr. is President of the National Academy of Engineering and Regents Professor, on leave, from the University of Maryland, College Park.

Dr. Mote is a native Californian who earned his BS, MS, and PhD degrees at the University of California, Berkeley in mechanical engineering between 1959 and 1963. After a postdoctoral year in England and three years as an assistant professor at the Carnegie Institute of Technology in Pittsburgh, he returned to Berkeley to join the faculty in mechanical engineering for the next 31 years. He and his students investigated the dynamics, stability, and control of high-speed rotating and translating continua (e.g., disks, webs, tapes, and cables) as well as biomechanical problems emanating from snow skiing. He coined the area called "dynamics of axially moving materials" encompassing these systems. Fifty-eight PhD students earned their degrees under his mentorship.

James Onken

James Onken is Senior Advisor to the NIH Deputy Director for Extramural Research and Director of the Office of Data Analysis Tools and Systems within the NIH Office of Extramural Research (OER). He has been conducting portfolio analyses and program evaluations at the NIH for over 27 years, holding positions at the National Institute of Mental Health and National Institute of General Medical Sciences before moving to OER. He previously held positions at AT&T Bell Laboratories, Decisions and Designs, Inc., and the U.S. Government Accountability Office. He holds M.S. and Ph.D. degrees in psychology from Northwestern University, and an MPH with a concentration in biostatistics from George Washington University.

Mark Orr

Mark Orr was originally trained as a cognitive psychologist at the University of Illinois at Chicago. He received augmentation to this training with postdoctoral fellowships in computational modeling (Carnegie Mellon), neuroscience (Albert Einstein College of Medicine), and epidemiology/complex systems (Columbia University). Over the past decade, he has become heavily involved in understanding dynamic processes and drivers of risky behavior and decision making, primarily in a public health context, at the scale of the individual and populations. Currently, he is PI on an NSF-funded project that explores a highly novel approach for understanding the diffusion of attitudes on social networks in the context of health behavior (see http://www.nsf.gov/awardsearch/showAward?AWD_ID=1520359). He is currently expanding these ideas into other contexts and for other applications (e.g., DoD, DoE, DHS).

Riq Parra

Riq Parra, Ph.D. joined the Air Force Office of Scientific Research (AFOSR) in 2011 where he currently serves as a program officer with responsibilities over the development and technical oversight of the Ultrashort Pulse Laser-Matter Interactions research program. In this role, he is responsible for articulating the Air Force objectives in fundamental research and finding, reviewing, and funding excellent cutting-edge research at universities to meet those objectives. Before joining AFOSR, Parra spent over seven years at Booz Allen Hamilton serving as a strategic consultant to government science and technology clients, including the Defense Advanced Research Projects Agency (DARPA). In this role, Parra helped formulate and provide technical oversight over twenty basic and applied research programs in a wide range of technical areas including ultrafast lasers, high-energy lasers, atomic physics and nonlinear optics.

Brian Pate

Brian Pate is a Science & Technology Manager at the Joint Science & Technology Office for Chemical & Biological Defense, Defense Threat Reduction Agency. He manages a $13M portfolio focused on (1) interactions relevant to human effects of advanced and emerging weapons threats and (2) identifying and controlling new phenomena to enable breakthrough countermeasures to chemical and biological weapons. He is an adjunct faculty member at the University of Maryland and has recently served as an instructor in biochemistry and in weapons chemistry at Northern Virginia Community College and the U.S. Naval Academy, respectively. Pate earned a Ph.D. in Chemistry from Indiana University, followed by postdoctoral training at MIT in Materials Science & Engineering. Prior to his current role, Pate was employed as a Visiting Scientist at the Air Force Research Laboratory, as a Senior Chemistry Specialist at Dow Chemical, and as a Lecturer in Physics, Chemistry, and Materials Science at Central Michigan University. Pate recently served as Deputy Director for Technology Watch and Horizon Scanning at the Office of Technical Intelligence, Assistant Secretary of Defense for Research and Engineering, and he maintains an active interest in using data-driven analytics to improve S&T understanding, investments, and outcomes.

Grace Peng

Grace C.Y. Peng received the B.S. degree in electrical engineering from the University of Illinois at Urbana, the M.S. and Ph.D. degrees in biomedical engineering from Northwestern University. She performed postdoctoral and faculty research in the department of Neurology at Johns Hopkins University. In 2000, she became the Clare Boothe Luce professor of biomedical engineering at the

Catholic University of America. Since 2002, Peng has been a Program Director in the National Institute of Biomedical Imaging and Bioengineering (NIBIB), at the National Institutes of Health. Her program areas at the NIBIB include mathematical modeling, simulation and analysis methods, and next generation engineering systems for rehabilitation, neuroengineering, and surgical systems. In 2003, she brought together the Neuroprosthesis Group (NPG) of program officers across multiple institutes of the NIH. Also in 2003, Peng lead the creation of the Interagency Modeling and Analysis Group (IMAG), which now consists of program officers from ten federal agencies of the U.S. government and Canada (www.imagwiki.org). IMAG has continuously supported funding specifically for multiscale modeling (of biological systems) since 2004. IMAG facilitates the activities of the Multiscale Modeling (MSM) Consortium of investigators created in 2006. Peng is interested in promoting the development of intelligent tools and reusable models and their application in engineering systems and for solving multiscale physiological problems.

Alex "Sandy" Pentland

Alex “Sandy” Pentland directs the MIT Connection Science and Human Dynamics labs and previously helped create and direct the MIT Media Lab and the Media Lab Asia in India. He is one of the most-cited scientists in the world, and Forbes recently declared him one of the “7 most powerful data scientists in the world” along with Google founders and the Chief Technical Officer of the United States. He has received numerous awards and prizes such as the McKinsey Award from Harvard Business Review, the 40th Anniversary of the Internet from DARPA, and the Brandeis Award for work in privacy. He is a founding member of advisory boards for Google, AT&T, Nissan, and the UN Secretary General, a serial entrepreneur who has co-founded more than a dozen companies including social enterprises such as the Data Transparency Lab, the Harvard-ODI-MIT DataPop Alliance and the Institute for Data Driven Design. He is a member of the U.S. National Academy of Engineering and leader within the World Economic Forum. Over the years, Pentland has advised more than 60 Ph.D. students. Almost half are now tenured faculty at leading institutions, with another one-quarter leading industry research groups and a final quarter founders of their own companies. Together, Pentland and his students have pioneered computational social science, organizational engineering, wearable computing (Google Glass), image understanding, and modern biometrics. His most recent books are “Social Physics” published by Penguin Press and “Honest Signals” published by MIT Press. Interesting experiences include dining with British Royalty and the President of India, staging fashion shows in Paris, Tokyo, and New York, and developing a method for counting beavers from space.

Zhao Qu

Zhao Qu is a Chinese Ph.D. candidate at the Humboldt University of Berlin. Her advisor is Dr. Stefan Hornbostel who is a Professor at the Department of Social Sciences at Humboldt University and Director of the Department “Research System and Science Dynamics” of the German Centre for Higher Education Research and Science Studies (DZHW). Qu’s research focuses on bibliometics and patentometrics. She holds and Engineering Master’s degree and has expertise in scientific management and innovation studies.

Venkatachalam "Ram" Ramaswamy

Venkatachalam ("Ram") Ramaswamy is Director of NOAA's Geophysical Fluid Dynamics Laboratory (GFDL) since 2008. Ram received his undergraduate degree in Physics from Delhi University (India), and Ph.D. in Atmospheric Sciences from the State University of New York at Albany. He was a Fellow in the Advanced Study Program at the National Center for Atmospheric Research. He joined GFDL in 1985, and was a Senior Scientist before becoming Director. His principal interests are numerical modeling of the global climate system and advancing the understanding of the past, present, and future states of climate including weather extremes. He directs one of the world's leading climate modeling centers, with the mission to develop mathematical models for predicting climate. Ram is a Fellow of the American Meteorological Society and American Geophysical Union, a recipient of the Presidential Rank award, and has served on the Intergovernmental Panel on Climate Change and World Climate Research Program.

William B. Rouse

William B. Rouse is the Alexander Crombie Humphreys Chair within the School of Systems & Enterprises at Stevens Institute of Technology and Director of the Center for Complex Systems and Enterprises. His research focuses on understanding and managing complex public-private systems such as healthcare delivery, urban systems, and national security, with emphasis on mathematical and computational modeling of these systems for the purpose of policy design and analysis. Rouse has written hundreds of articles and book chapters, and has authored or edited many books, including most recently "Universities as Complex Enterprises" (Wiley, 2016), "Modeling and Visualization of Complex Systems and Enterprises" (Wiley, 2015), and "Understanding and Managing the Complexity of Healthcare" (MIT Press, 2014).

Nachum Shacham

Nachum Shacham is Director of Data Science at PayPal where he is constructing models and leading a team of data scientists in identifying actionable patterns in large transactional, behavioral, and system performance datasets. Before, he was with eBay, analyzing performance of large data platforms. Prior, he was with SRI, leading research in internet technologies, generation of wireless internet and real-time voice and video communications over mobile networks. As co-founder and CTO of Metreo, he developed models for B2B pricing and subsequently created revenue models for online display and search advertising. Nachum holds B.S. and a M.S. in Electrical Engineering from the Technion and a Ph.D. in EECS from UC Berkeley. Shacham is a Fellow of the IEEE.

Ben Shneiderman

Ben Shneiderman (http://www.cs.umd.edu/~ben) is a Distinguished University Professor in the Department of Computer Science, Founding Director (1983-2000) of the Human-Computer Interaction Laboratory (http://www.cs.umd.edu/hcil), and a Member of the UM Institute for Advanced Computer

Studies (UMIACS) at the University of Maryland. He is a Fellow of the AAAS, ACM, IEEE, and NAI, and a Member of the National Academy of Engineering, in recognition of his pioneering contributions to human-computer interaction and information visualization. Shneiderman is the co-author with Catherine Plaisant of “Designing the User Interface: Strategies for Effective Human-Computer Interaction” (6th ed., 2016), see http://www.awl.com/DTUI. His latest book is “The New ABCs of Research: Achieving Breakthrough Collaborations” (Oxford, February 2016).

Roberta Sinatra

Roberta Sinatra is an Assistant Professor at the Center for Network Science and at the Math Department, Central European University (Hungary), and a visiting Faculty at the Network Science Institute, Northeastern University (USA). She is a theoretical physicist by training, working at the forefront of network and data science, developing novel theoretical methods and analyzing empirical data sets on social phenomena and human behavior. Currently, she spends particular attention on the analysis and the modeling of information and dynamics that lead to the collective phenomenon of success. Sinatra completed her studies in Physics at the University of Catania, Italy and spent time as a visiting research student in Universities and Research centers in Zaragoza (Spain), London (UK), and Vienna (Austria). In 2012, she joined the BarabasiLab in Boston, first as Postdoctoral fellow then starting 2014 as Research Assistant Professor. She has won several awards and grants, in particular a 3-years fellowship by the James S. McDonnell Foundation and a grant from the AirForce for the study of scientific success.

Timothy Slaper

Timothy Slaper {slay-per} is director of economic analysis at the Indiana Business Research Center (IBRC) in Indiana University's Kelley School of Business. His research team is engaged in industry and workforce analysis, economic impact studies, regional economic analyses, demographic estimates and projections, trade and foreign investment analysis, measuring innovation and educational performance and the drivers of economic growth. The team is putting the final touches on “The Innovation Index 2.0” which is a county-based data set and web tool for economic development practitioners as well as policy makers and researchers to assess a region’s innovative capacity and economic performance. All work that Slaper oversees puts analytical tools and practical research into the hands of economic development practitioners to help them address the challenges of economic development in today’s rapidly changing world. Before the IBRC, Slaper served as Senior Economist on the Joint Economic Committee of Congress. He cut his teeth as an economist at the U.S. Bureau of Economic Analysis. Slaper earned his doctorate in economics at The American University in Washington D.C. and is an alumnus of Miami University.

Jeroen Struben

Jeroen Struben is Assistant Professor in the Strategy & Organization Area at the Desautels Faculty of Management Faculty and Fellow of the Marcel Desautels Institute for Integrated Management, McGill University. He is a social and systems scientist with research focused on the dynamics of market formation and transformation towards more sustainable pathways. Empirically, Struben studies energy, alternative fuel vehicle, and nutrition markets. He is particularly interested in the question: How do alternative products, ideas, and practices successfully penetrate in the marketplace or society at large, rather than falter? To examine this, his research focuses on how social processes and evolution of the built environment jointly condition the formation of self-sustaining markets. His research combines empirical, analytical, and systems science-based analysis, producing insights related to coordination, collective action and commitment across organizations, industries and governments. Struben received his Ph.D. at MIT's Sloan School of Management.

Giorgio Triulzi

Giorgio Triulzi is a postdoctoral fellow at the Institute for Data, System, and Society at MIT and an affiliated researcher at UNU-MERIT. His research interests focus on understanding drivers and direction of technological change and their wider effects on the economy. He applies theories and methods from complex system analysis, networks, evolutionary economics and strategy. Triulzi has a Ph.D. in Economics and Policy Studies of Technical Change from UNU-MERIT and Maastricht University.

Brian Uzzi

Brian Uzzi studies the links between social networks, complex systems, and human achievement in the areas of business, science, and the arts. He is the Richard L. Thomas Distinguished Professor of Leadership at the Kellogg School of Management, Northwestern University. He is also co-director of the Northwestern University Institute on Complex Systems (NICO) and is a professor of sociology and professor of management science at the McCormick School of Engineering. Brian has received over 10 scholarly research prizes and 13 teaching awards. He has been on the faculty of Harvard University, INSEAD, University of Chicago, and UC Berkeley where he was the Warren E. and Carol Spieker Professor of Leadership. Media reports featuring his work have appeared in the *WSJ, Economist, Newsweek, NYT, Fortune, Wired*, on Television, and in the *New Yorker Magazine*.

Bill Valdez

Bill Valdez retired from Federal service as a career Senior Executive in 2014 and is now a Senior Vice President at an international consulting firm that specializes in energy, environment and science/technology policy. In addition, he is an Adjunct Faculty at American University's School of Public Affairs and is co-editing "The Handbook of Federal Government Leadership and Administration" to be published in Spring 2016. Mr. Valdez's career with the Department of Energy spanned over 20 years. He has extensive knowledge in the areas of R&D portfolio analysis and evaluation, science and engineering workforce development, national lab policy, science and technology policy, and corporate and strategic planning. Valdez was co-chair of the NSTC Science of Science Policy (SoSP) Interagency Working Group from 2005-2014; was awarded the Presidential Rank Award in 2007, and was elected as a Fellow of the American Association for the Advancement of Science in 2006.

Caroline Wagner

Caroline S. Wagner holds the Ambassador Milton A. and Roslyn Z. Wolf Chair in International Affairs, where she also serves as the Director of the Battelle Center for Science & Technology Policy. As a faculty member at the John Glenn School of Public Affairs, she teaches public policy and leadership. Her area of scholarship is in the field of science and technology and its association to policy, society, and innovation. She has been an advisor to the European Commission, the World Bank, the United States National Science Foundation, the Organization for Economic Cooperation and Development and several governments. As part of the United Nations Millennium Development Project, Wagner served on the Task Force on Science, Technology, and Innovation and co-authored the final report. Wagner earned her Ph.D. from the University of Amsterdam in Science and Technology Dynamics; her M.A. in Science, Technology and Public Policy from George Washington University, and holds a B.A. from Trinity College. She is a Fellow of the American Association for the Advancement of Science (AAAS).

John Walsh

John P. Walsh is Professor of Public Policy at Georgia Institute of Technology. He teaches and performs research on science, technology, and innovation, using a sociological perspective that focuses on organizations and work to explain how research organizations respond to changes in their policy environment. Recent work includes studies of organization and creativity in scientific teams; the effects of knowledge environments on non-R&D innovation; academic entrepreneurship in the U.S. and Japan; and country and industry differences in the role of patents in firm strategy. He is an Editor at *Research Policy*. His work has been published in *Science, American Sociological Review, Research Policy, Social Studies of Science*, and *Management Science*. His work has been funded by the National Science Foundation, the Japanese Society for the Promotion of Science, the Ewing Marion Kauffman Foundation, the Matsushita Foundation and the Japan Foundation. He has consulted for the National Academy of Sciences, the OECD, the European Commission, and the American Association for the Advancement of Science.

Victor Yakovenko

Victor Yakovenko is a Professor at the Department of Physics, University of Maryland, College Park. He is a Russian-born American physicist noted for his work in promoting the subject of econophysics in America. He holds a M.S. in Physics and Engineering from the Moscow Physical-Technical Institute and a Ph.D. in Theoretical Physics from the Landau Institute for Theoretical Physics in Moscow.

Xiaoran Yan

Xiaoran Yan is a Research Scientist at the Indiana University Network Science Institute (IUNI). His research concerns mathematical theories and models of networks, with a focus on community structures and dynamical processes on networks. He worked as a Postdoctoral Research Associate at Information Sciences Institute of University of Southern California. Before that, he was a graduate fellow at Santa Fe Institute. He holds a Ph.D. in Computer Science from University of New Mexico.

Masaru Yarime

Masaru Yarime is Project Associate Professor of Science, Technology, and Innovation Governance (STIG) at the Graduate School of Public Policy of the University of Tokyo, Japan. He also has an appointment as Honorary Reader in the Department of Science, Technology, Engineering and Public Policy (STEaPP) of University College London, United Kingdom. He has been awarded an Abe Fellowship by the United States Social Science Research Council. His research interests focus on public policy, corporate strategy, and institutional design for promoting science, technology, and innovation for tackling societal or grand challenges, including energy, environment, and sustainability. He received B.Eng. and M.S. in Chemical Engineering from the University of Tokyo and the California Institute of Technology, respectively, and a Ph.D. in Economics and Policy Studies of Technological Change from Maastricht University in the Netherlands. Previously, he worked as Senior Research Fellow at the National Institute of Science and Technology Policy.

Lynne Zucker

Lynne Zucker is a Professor in the Departments of Sociology and Policy Studies at UCLA. Her training is in organizational sociology, institutional theory, economic sociology, and social psychology. Zucker's research focuses on the processes and impact of knowledge transmission from basic science to commercial use, especially the impact on economic performance of firms, creation of new organizational populations (some of which become new industries), and on productivity growth. With Michael Darby, she shares an interest in identifying the major mechanisms of knowledge transfer and the institutional infrastructure that cause metamorphic industry change and rapid economic growth. Within the context of basic scientific breakthroughs that are commercially applicable, they are exploring other measures of success such as IPO returns and examining the impact of other means of knowledge transfer such as joint ventures. Zucker and Darby are studying many of the same processes in nanoscience, a newly emerging basic research area with significant commercial potential. To identify institutional infrastructure effects, they are completing a comparative study of biotech in Japan and the U.S. and embarking on a set of major international analyses of the transmission of scientific breakthroughs to commercial use in nanotechnology.