

# ASE

Academy of Science and Engineering

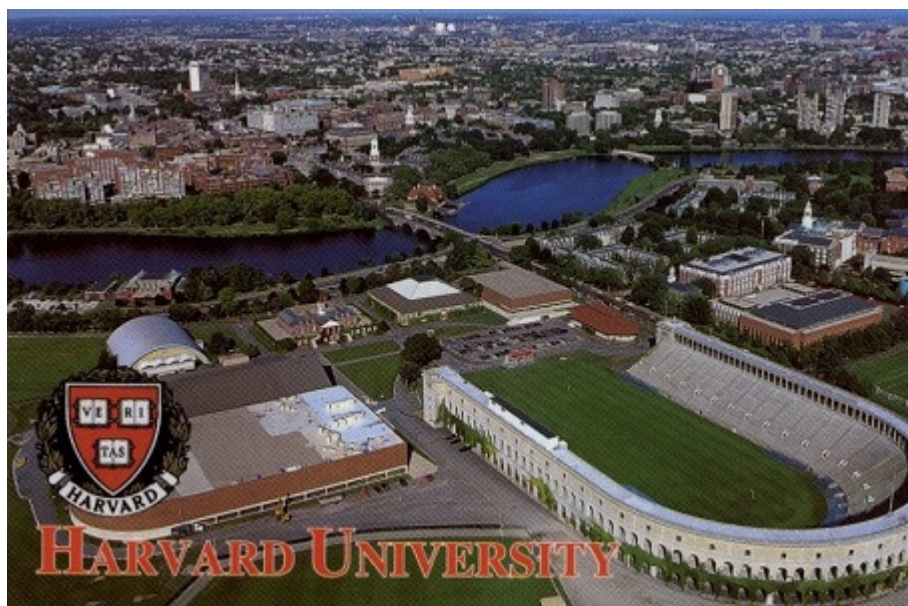
# 2014

**BigData 2014: The Fourth ASE International Conference on Big Data**

**SocialInformatics 2014: The Third ASE International Conference on Social Informatics**

**PASSAT 2014: The Sixth ASE International Conference on Privacy, Security, Risk and Trust**

**BioMedCom 2014: The Third ASE International Conference on Biomedical Computing**



**December 14-16, 2014  
Harvard University,  
1737 Cambridge Street,  
Cambridge, MA 02138, USA**

*Gateway of knowledge,  
imagination, exploration,  
discoveries through open  
access for the next  
generation Digital World*

[www.asesite.org](http://www.asesite.org)

## Table of Contents

<b>Welcome Message from the Conference Chairs.....</b>	<b>Page 2</b>
<b>Conference Committee .....</b>	<b>Page 3</b>
<b>Conference Program/Sessions/Papers .....</b>	<b>Page 5</b>
<b>Speaker Biographical Sketches and Abstracts.....</b>	<b>Page 9</b>
<b>Poster Session .....</b>	<b>Page 26</b>

## **Welcome Message from the Conference Chairs**

On behalf of the Academy of Science and Engineering (ASE), welcome to the Fourth ASE International Conference on Big Data, the Sixth ASE International Conference on Privacy, Security, Risk and Trust, the Third ASE International Conference on Social Informatics, and the Third ASE International Conference on Biomedical Computing. The conference committee has assembled a very relevant program that will provide perspective on the current state of the art and recent advances to address big data, privacy, security, risk, trust, social informatics, and biomedical computing. These ASE conferences are leading international forum for big data, privacy, security, risk, trust, social informatics, and biomedical computing researchers, practitioners, developers, and users to explore cutting-edge ideas and results, and to exchange techniques, tools, and experiences. The topics in this conference continue to address a wide range of important issues. With this year's conference, we continue to receive robust participation from researchers working in areas of big data, privacy, security, risk, trust, social informatics, and biomedical computing.

Various workshops organized in this conference will connect scientist and practitioners focused on improving the predictiveness of organizations or on leading organizations through major transformations. These workshops will provide a medium to begin a collaboration to apply big data, privacy, security, risk, trust, social informatics, and biomedical computing solutions. It will provide a key forum for researchers and industry practitioners to exchange information regarding advancements in the state of art and practice of big data, privacy, security, risk, trust, social informatics, and biomedical computing.

We think that the topics and speakers cover a highly relevant range of material and that this conference is a unique forum that will bring together government, industry and academia to share ideas on these critical topics.

We wish to acknowledge the substantial efforts of the ASE conference and workshop chairs, committee members, speakers, authors, presenters, sponsors, exhibitors, volunteers and, in particular, the Institute for Quantitative Social Science facilities at Harvard University in making this event a reality.

We look forward to the conferences.

Sincerely,

### **Conference Executive Chairs:**

Gil Alterovitz, Harvard/MIT, USA

Ning An, Hefei University of Technology, China

Simon Reay Atkinson, University of Sydney, Complex Civil Systems Group, RAN, Australia

Tom Carsey, University of North Carolina-Chapel Hill, USA

LiWu Chang, US Patent Office, USA

Merce Crosas, Harvard University, USA

Holly Jimison, Northeastern University, USA

Javed Mostafa, University of North Carolina at Chapel Hill, USA

Henning Muller, HES-SO Valais, USA

Michael Ochs, The College of New Jersey, USA

Pavlos Protopapas, Harvard University, USA

Arcot Rajasekar, University of North Carolina-Chapel Hill, USA  
Christoph Riedl, Northeastern University, USA  
Andrey Rzhetsky, University of Chicago, USA  
Paul Thompson, Dartmouth College, USA  
Gregory Tolhurst, University of Sydney, Faculty of Law, Australia  
Jianyong Wang, Tsinghua University, China  
Justin Zhan, North Carolina A&T State University, USA

### **Program Committee**

Andrew A. Adams, Centre for Business Information Ethics, Meiji University, Japan  
Harith Alani, The Open University, UK  
Jay Aslam, Northeastern University, USA  
Anne Bader, The International Cyber Security Dialogue, USA  
Siddhartha Bhattacharyya, University of Illinois, Chicago, USA  
Nicholas Caldwell, University College Suffolk and University of Cambridge, UK  
Agnes Hui Chan, Northeastern University, USA  
Steve Checkoway, Johns Hopkins University Information Security Institute (JHUI SI), USA  
Bruno Crispo, University of Trento, Italy  
Anna Esposito, International Institute for Advanced Scientific Studies, Italy  
Lars Fischer, University of Siegen, Germany  
Leopoldina Fortunati, University of Udine, Italy  
Enrique Frias-Martinez, Telefonica Research, Madrid  
Alain Forget, Carnegie Mellon University, USA  
Simson Garfinkel, Naval Postgraduate School in Monterey, California, USA  
Georg Groh, TU München, Germany  
Greta Greer, American Cancer Society, USA  
Michael Henshaw, Loughborough University, UK  
Patricia Hirl, Syracuse University, USA  
Cindy Hood, Illinois Institute of Technology, USA  
Henric Johnson, Blekinge Institute of Technology, Sweden  
Andrea Kavanaugh, Virginia Tech, USA  
David Kaeli, Northeastern University, USA  
Helmut Krcmar, Technische Universität München, Germany  
Luis Lamb, Institute of Informatics, UFRGS  
Mark Levene, University of London, UK  
Dongman Lee, KAIST, South Korea  
Yuhong Liu, Penn State University, USA  
Aleecia M. McDonald, Stanford University, USA  
Carlo Francesco Morabito, University of Reggio Calabria, Italy  
Peter Mutschke, Leibniz Institute for the Social Sciences, Germany  
Anneli Avatare Nöu, Swedish Institute of Computer Science, Sweden  
Misha Pavel, Northeastern University, USA  
Scott Piao, Lancaster University, UK  
Goran Putnik, University of Minho, Portugal  
Claus-Peter Ruckemann, Leibniz University Hannover and North-German Supercomputing Alliance, Germany  
Viktor Prasanna, University of Southern California, USA

Hamid Rabiee, Sharif University of Technology, USA  
Vaskar Raychoudhury, Indian Institute of Technology, India  
Michael Schermann, Technische Universität München, Germany  
Dale Shoupe, USAF Air Warfare College, USA  
Eunice Santos, University of Texas, USA  
Pravin Shankar, Facebook, USA  
Yan Sun, University of Rhode Island, USA  
Jerzy Surma, University of Massachusetts, USA  
Dan Thomsen, Smart Information Flow Technologies, USA  
Kami Vaniea, Michigan State University, USA  
Gualtiero Volpe, University of Genova, Italy  
Natalie van der Wal, VU University Amsterdam, The Netherland  
May Wang, Georgia Institute of Technology, USA  
Rui Wang, Microsoft Research, USA  
Giorgos Zervas, Boston University, USA

## Conference Program

**December 14, 2014 (Sunday)**

<b>8:00</b>	<b>Registration: The Institute for Quantitative Social Science, Harvard University</b> <b>Address: 1737 Cambridge Street, Cambridge, MA, 02138, USA</b>
	<b>Morning Session</b> <b>Chair: Dr. Justin Zhan/ Dr. Pravin Chopade</b> <b>(Room: Tsai &amp; Belfer)</b>
<b>9:00-9:10</b>	<b>Open Remarks</b>
<b>9:10-9:50</b>	Speech Title: Open Language Grid - Towards a Global Language Service Infrastructure Dr. Toru Ishida Department of Social Informatics Kyoto University
<b>9:50-10:10</b>	Spam Campaign Cluster Detection Using Redirected URLs and Randomized Sub-Domains Abu Awal Md Shoeb, Dibya Mukhopadhyay, Shahid Al Noor, Alan Sprague, Gary Warner
<b>10:10-10:30</b>	A Smart Assignment Technique with Consideration of Multicriteria Reciprocal Judgements Zeynab Bahrami Bidoni, Roy George, Ahmad Makui
<b>10:30-10:50</b>	Management of Optimized Knowledge Diffusion and Creation in Organizational Structured Social Networks Tianshu Chu and Jie Wang
<b>10:50-11:10</b>	User-controlled Privacy for Personal Mobile Data Sharon Paradesi and Lalana Kagal
<b>11:10-11:30</b>	Koonkie: An Automated Software Tool for Processing Environmental Sequence Information using Hadoop Dongjae Kim, Kishori Konwar, Niels Hanson and Steven Hallam
<b>11:30-11:50</b>	Privacy Informer: An Automatic Privacy Description Generator for Mobile Apps Daniela Miao and Lalana Kagal
<b>11:50-12:10</b>	LogitTrust: A Logit Regression-based Trust Model for Mobile Ad Hoc Networks Yating Wang, Yen-Cheng Lu, Ing-Ray Chen, Jin-Hee Cho, Ananthram Swami and Chang-Tien Lu
<b>12:10-13:00</b>	<b>Lunch Speech Title: The Hidden Trade in Personal Medical Data</b> <b>Speaker: Dr. Adam Tanner</b>
	<b>Afternoon Session</b> <b>Chair: Dr. Simon Reay Atkinson/Dr. Yuhong Liu</b>
<b>13:00-13:20</b>	Visualization Techniques for Large-Scale Big Data Networks: Smart Power Grid Survivability in a Complex Operating Environment Pravin Chopade, Kenneth Flurchick, Justin Zhan and Marwan Bikdash
<b>13:20-13:40</b>	Synthetic Generators for Cloning Social Network Data Awrad Mohammed Ali, Hamidreza Alvari, Alireza Hajibagheri, Kiran Lakkaraju and Gita Sukthankar
<b>13:40-14:00</b>	Tweets You Like: Personalized Tweets Recommendation based on Dynamic Users Interests Shaymaa Khater, Hicham Elmongui and Denis Gracanin
<b>14:00-14:20</b>	Sentiment Analysis of Colloquial Arabic Tweets Nagwa El-Makky, Khaled Nagi, Alaa El-Ebshihiy, Esraa Apady, Omneya Hafez, Samar Mostafa, and Shimaa Ibrahim
<b>14:20-14:40</b>	Cultural and Geolocation Aspects of Communication in Twitter Elena Daehnhardt, Yanguo Jing and Nick Taylor
<b>14:40-15:00</b>	Protecting Private Information: Current Attitudes Concerning Privacy Policies Therese Williams, Nitin Agarwal and Rolf Wigand
<b>15:00-15:30</b>	<b>Coffee Break</b>
<b>15:30-15:50</b>	Nemesis: Automated Architecture for Threat Modeling and Risk Assessment for Cloud Computing Patrick Kamongi, Mahadevan Gomathisankaran and Krishna Kavi
<b>15:50-16:10</b>	Risk Characteristics, Mental Models, and Perception of Security Risks Vaibhav Garg and Jean Camp
<b>16:10-16:30</b>	Towards Trusting User-Generated Content in Web Applications Yannis Georgalis and Yuzuru Tanaka
<b>16:30-16:50</b>	Radium: Race-free On-demand Integrity Measurement Architecture

	Srujan Kotikela, Tawfig Shah, Mahadevan Gomathisankaranand Gelareh Taban
<b>16:50-17:10</b>	Using machine learning techniques for traffic classification and preliminary surveying of an attacker's prole Peter Fruhwirt, Sebastian Schrittwieser and Edgar Weippl
<b>17:10-17:30</b>	Msi Windows Installer Security Christian Kadluba, Martin Mulazzani, Lorenz Zechner, Sebastian Neuner and Edgar Weippl
<b>17:30-17:50</b>	Network Performance Rank: An Approach for Comparison of Complex Networks Zeyanab Bahrami Bidoni and Roy George
<b>17:50-18:10</b>	Intelligent Data Management Through The Data Burst (Best Practices). Hussain Alajmi
<b>18:10-18:30</b>	A New Instrument for Moderator Evaluation: Updating Fiedler's Contingency Model for Online Communities Alicia Takaoka
	<b>December 15, 2014</b>
<b>8:00-17:00</b>	<b>Registration &amp; Breakfast</b>
	<b>Morning Session</b>
	<b>Chair: Dr. Mercè Crosas</b>
<b>8:30-9:15</b>	Speech Title: Privacy Tools for Sharing Research Data Speaker: Dr. Salil P Vadhan
<b>9:15-10:00</b>	Speech Title: Big Data Issues with Health Monitoring and Streaming Sensor Data from the Home Speaker: Dr. Misha Pavel
<b>10:00-10:30</b>	Speech Title: How SQL Can Take Us From Hadoop to Energy Independence Ravi Chiruvolu, CEO at Slingshot Power Michael Hoskins, CTO at Actian
<b>10:30-11:00</b>	Speech Title: Big Data versus Big Insight - A Case Study from Call Center Optimization Dr. Bernhard Suhm Director Professional Services, AVOKE Call Center Analytics Raytheon BBN Technologies
<b>11:00-11:30</b>	Speech Title: SciDB: For scientists who want to do science, not computer science Marilyn Matz (CEO and co-founder of Paradigm4) Alex Poliakov (Solutions Architect for Paradigm 4)
<b>11:30-12:30</b>	Panel Title: Big Data, Privacy and Information Technology: Harnessing the Power of the Multi Stakeholder Networks Anne C. Bader, May Wang, Latanya Sweeney, Simon Reay Atkinson, Holly B Jimison
<b>12:30-13:30</b>	<b>Lunch/Best Paper Award/Next Year Conference Announcement</b>
	<b>Afternoon Session</b>
	<b>Chair: Dr. Holly Jimison/Dr. May Wang</b>
<b>13:30-13:50</b>	Constructing a Post Influence-Predicting Model in Social Networking Service Based on NGD and fuzzy C-Means Algorithm Hsiao-Wei Hu and Chia-Wei Huang
<b>13:50-14:10</b>	Incremental and Parallel Association Mining from Evolving Spatial Databases: A Less Iterative Approach on MapReduce Jin Soung Yoo, Douglas Boulware and David Kimmey
<b>14:10-14:30</b>	A Socio-Technical Study on User Centered Trust Notions and Their Correlation to Stake in Practical Information Technology Scenarios Jiun Yi Yap and Allan Tomlinson
<b>14:30-14:50</b>	Policy-based Pre-Processing in Hadoop Yi Cheng and Christian Schaefer
<b>14:50-15:10</b>	Utilizing Dynamic Precedence Criteria To Ensure the Production of Critical Results from Big Data Streams Karen Works and Elke Rundensteiner
<b>15:10-15:30</b>	Dynamic Connected Components in Large Fully Distributed Graphs Ellison Anne Williams and Brant Chee
<b>15:30-15:40</b>	<b>Coffee Break</b>
<b>15:40-16:00</b>	Equity-preserving Management of Privacy Conflicts in Social Network Systems Regina Marin, Guillaume Piolle and Christophe Bidan



16:00-16:20	EStarMom: Extendable Simulator for Trust and Reputation Management in Online Marketplaces Thao Nguyen and Brian D'Auriol
16:20-16:40	"Age shall not wither them": But it Will Change Their Priorities About Protecting Their Information Privacy Anthony Morton
16:40-17:00	A novel Classification-based Hybrid IDS Oscar Rodas, Jose Alvarez, Gerardo Morales, Stephane Maag
17:00-17:20	Towards Leveraging Late-Launch to Create Trustworthy Thin-Terminal Clients Evan Frenn and Craig Shue
17:20-17:40	Adapting Jurisprudential Networks to Stateless Jurisdictions Simon Reay Atkinson and Gregory Tolhurst
17:40-18:00	What do they know about me? Contents and Concerns of Online Behavioral Profiles Ashwini Rao, Florian Schaub and Norman Sadeh
18:00-18:20	Identifiability of Vehicle Tollgate Records: The Milan Tollgate Dataset Nick Manfredi, Darakhshan Mir, Claire Schlenker
18:20-18:40	Seeing Stars from Reviews by a Semantic-based Approach with MapReduce Implementation Pengfei Liu, Xiaojun Qian and Helen Meng
	<b>December 16, 2014</b>
8:00	<b>Registration &amp; Breakfast</b>
	<b>Morning Session</b>
	<b>Chair: Dr. Christoph Riedl</b>
8:30-9:00	Speech Title: Soon Everyone Will Be a "Data Scientist" or a "Data Explorer." What Data Systems Will They Be Using? Speaker: Dr. Stratos Idreos Computer Science at the Harvard School of Engineering and Applied Sciences Harvard University
9:00-9:30	Speech Title: Cybersecurity Education and Workforce Development at the National Science Foundation Speaker: Dr. Corby Hovis Lead Program Director Division of Undergraduate Education National Science Foundation
9:30-9:50	Performance Comparison of Big-Data Technologies in Locating Intersections in Satellite Ground Tracks Khoa Doan, Amidu Oloso, Kwo-Sen Kuo and Thomas Clune
9:50-10:10	New Quality Function Deployment Integrated Methodology for Design of Big Data E-Government System in Egypt Ahmed Attia and Hesham Mahmoud
10:10-10:30	Semantic Enabled Social-Collaborative Research Framework for Proteomics Domain Nazmul Hussain and Hai Wang
10:30-10:40	<b>Coffee Break</b>
10:40-11:00	Customer Churn Prediction, Segmentation and Fraud Detection in Telecommunication Industry Ahsan Rehman and Abbas Raza Ali
11:00-11:20	Connectors, Mavens, Salesmen and More: An Actor-Based Online Social Network Analysis Method Using Tensed Predicate Logic Joshua White and Jeanna Matthews
11:20-11:40	Robust High-dimensional Data Stream Clustering in Arbitrary Subspaces Katelyn Gao
11:40-12:00	Content Sharing in Global Organization: A Cross-Country Perspective Amit Pariyar, Yohei Murakami, Donghui Lin and Toru Ishida
12:00-13:30	<b>Lunch &amp; Poster Session</b>
	<b>Afternoon Session</b>
	<b>Chair: Dr. Gil Alterovitz</b>
13:30-13:50	Incremental Wrapper based Gene Selection with Markov Blanket Aiguo Wang, Ning An, Jing Yang, Lian Li and Gil Alterovitz.



<b>13:50-14:10</b>	Pseudonymization for Secondary Use of Cloud Based Electronic Health Records Liangyu Xu, Armin B. Cremers and Tobias Wilken
<b>14:10-14:30</b>	An Innovative Genetic Programming Framework in modelling a real Epileptic Seizure detection system Arpit Bhardwaj, Aruna Tiwari, Ramesh Krishna Maddula and Vishaal Varma Madaka
<b>14:30-14:50</b>	Towards a Cognitive Metric using Normalized Transfer Entropy Md. Hedayetul Islam Shovon, D Nanda Nandagopal, Vijayalakshmi Ramasamy, Jia Tina Du and Bernadine Cocks
<b>14:50-15:10</b>	Music Data Analysis: A State-of-the-art Survey Shubhanshu Gupta
<b>15:10-15:30</b>	A New Objective Automatic Computational Framework for Evaluating and Visualizing the Results of Binhang Yuan, David Khechoyan and Ron Goldman.
<b>15:30-15:40</b>	<b>Coffee Break</b>
<b>15:40-16:00</b>	CoGAPS: Stochastic Nonnegative Matrix Factorization by Bayesian Markov Chain Monte Carlo Wai-Shing Lee, Conor Kelton, John Stansfield, Ondrej Maxian, Alexander Favorov, Elana Fertig and Michael Ochs
<b>16:00-16:20</b>	Implementation of a Mobile Health System for Monitoring ECG signals Nassim Amour, Ahmad Hersi, Naif Alajlan, Yakoub Bazi and Haikel Alhichri.
<b>16:20-16:40</b>	Do Offline Social Effects Drive Interactions in Citation Networks? Huda Alhazmi and Swapna Gokhale
<b>16:40-17:00</b>	<b>Conference Closing Remarks</b>

## Speech Title: The Hidden Trade in Personal Medical Data

Dr. Adam Tanner  
Fellow of Institute for Quantitative Social Science  
Harvard University



**Biographical Sketch:** Dr. Adam Tanner is a fellow at Harvard's Institute for Quantitative Social Science. He has been at Harvard since the 2011-12 academic years when he was a Nieman fellow. He recently completed “What Stays in Vegas: The World of Personal Data – Lifeblood of Big Business – and the End of Privacy as We Know It,” a book to be published Sept. 2nd by Public Affairs (Perseus Group). From 1995 until 2011 Tanner worked for Reuters, including as bureau chief for the Balkans and in San Francisco. In other postings he served in Germany, Russia and Washington D.C. He was part of the Reuters team cited in 2012 as a Pulitzer finalist in international reporting. He has appeared on the BBC and National Public Radio, and writes for Forbes, Worth and other magazines. The 2014 winner of the NYU Arthur L. Carter Journalism Institute’s Reporting Award, Tanner is now researching the business of medical data. In his latest work, he is looking at the pharmaceutical data/information marketplace and the complex trade in personal medical details generated from doctor visits, hospital stays and other medical treatments. He speaks from time to time at Harvard and other university campuses and is giving the final keynote speech at the June 2014 Patients Privacy Rights conference in Washington D.C. He also gives lectures on international Voyages to Antiquity cruises.

**Abstract:** When you fill a prescription, visit the doctor or hospital, or get a test done, that information is often sold to outside commercial companies as part of what is called longitudinal patient data. Deeply intimate but anonymized, the trade in such data has increased dramatically in the last 15 years. How does so much of our intimate medical data end up in commercial circulation without our knowledge, and should we worry about the privacy implications?

## Speech Title: Privacy Tools for Sharing Research Data

Dr. Salil Vadhan

Computer Science at the Harvard School of Engineering and Applied Sciences  
Harvard University



**Biographical Sketch:** Salil Vadhan is the Vicky Joseph Professor of Computer Science and Applied Mathematics in the Harvard University School of Engineering and Applied Sciences, and the Director of the Harvard Center for Research on Computation and Society. He received his PhD in Applied Mathematics from MIT in 1999, and was an NSF Postdoctoral Fellow at MIT and the Institute for Advanced Study before joining the Harvard faculty in 2001. He is a recipient of a Simons Investigator Award, a Godel Prize, a Guggenheim Fellowship, a Phi Beta Kappa Award for Excellence in Teaching, and the ACM Doctoral Dissertation Award.

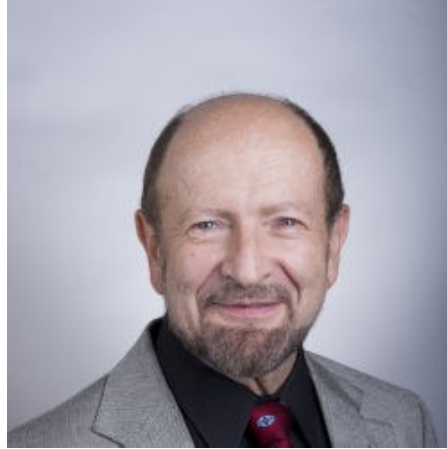
Vadhan's research area is theoretical computer science, specifically computational complexity, cryptography, and differential privacy. He is the Lead PI on an NSF frontier project "Privacy Tools for Sharing Research Data" (<http://privacytools.seas.harvard.edu/>).

**Abstract:** "Privacy Tools for Sharing Research Data" a large, multidisciplinary research project we have at Harvard to help enable the collection, analysis, and sharing of personal data for research in social science and other fields while providing privacy for the data subjects. Bringing together computer science, social science, statistics, and law, the project seeks to refine definitions and measures of privacy and data utility, and design an array of technological, legal, and policy tools for dealing with sensitive data.

In the talk, I will discuss the goals of the project, our progress to date, and challenges we see going forward. I will focus on two tools we're developing: a Data Tags tool that aims to help researchers navigate the complex array of legal requirements, contractual agreements, and best practices that govern how they should handle private-sensitive datasets; and a Differential Privacy tool that will enable making statistical information about privacy-sensitive datasets widely available when access to raw data needs to be constrained.

## Speech Title: Role of Computational Modeling in Big Data Analytics

Speaker: Dr. Misha Pavel  
College of Computer & Information Science  
Bouve College of Health Sciences  
Northeastern University



**Abstract:** The process of transforming healthcare to be more efficient, evidence-based and patient-centered will require the development of technological solutions in addition to solving many implementation, organizational and social issues, e.g. changing incentives. Recent efforts in digitization and interconnection of healthcare systems resulting in the availability of large amounts of data combined with a number of initial successes triggered an avalanche of conjectures that Big Data “will transform the way we think about data, data analysis and for that matter scientific inquiry”. These suppositions include the notions that there is no longer the need for hypotheses and statistical hypothesis testing since the “facts” will emerge from machine learning and data-mining of large data sets where  $N=ALL$ .

In this presentation I will identify a subset of potential consequences of these purely data-driven approaches and note that, in fact, every algorithm is based on a particular, often implicit, mathematical model. I will then describe a hierarchy of mathematical modeling approaches that may need to be evoked in order to render Big Data useful in optimizing practical decision making as well as advancing our theoretical understanding. In many domains, these models characterize processes at multiple scales combining data-driven and mechanistic approaches. I will illustrate aspects of this multiscale modeling on examples from an emerging discipline of behavioral informatics as applied to proactive and preventive healthcare.

**Speech Title: Soon Everyone Will Be a “Data Scientist” or a “Data Explorer.”  
What Data Systems Will They Be Using?**

Speaker: Dr. Stratos Idreos

Computer Science at the Harvard School of Engineering and Applied Sciences  
Harvard University



**Biographical Sketch:** Dr. Stratos Idreos is an assistant professor of Computer Science at the Harvard School of Engineering and Applied Sciences and leads DASlab, the Data Systems Laboratory at Harvard SEAS. Stratos works in the area of data management with emphasis on designing data systems for the big data era. Stratos obtained his Ph.D. from University of Amsterdam in the Netherlands. Before joining Harvard he spent 3 years as a tenure-track Scientific Staff Member with the Dutch National Research Center for Mathematics and Computer Science and held research internship and visiting scholar positions with Microsoft Research, Redmond USA, with EPFL, Switzerland and with IBM Research in Almaden USA. For his doctoral work on Database Cracking, Stratos won the 2011 ACM SIGMOD Jim Gray Doctoral Dissertation award which recognizes the best thesis internationally in the field of data management. In addition, he won the 2011 ERCIM Cor Baayen award as "most promising European young researcher in computer science and applied mathematics" from the European Research Council on Informatics and Mathematics. In 2010 he was awarded the IBM zEnterprise System Recognition Award by IBM Research, while in 2011 he also won the Challenges and Visions best paper award in the 2011 International Conference on Very Large Databases.

**Abstract:** How far away are we from a future where a data management system sits in the critical path of everything we do? Already today we need to go through a data system in order to do several basic tasks, e.g., to pay at the grocery store, to book a flight, to find out where our friends are and even to get coffee. Businesses and sciences are increasingly recognizing the value of storing and analyzing vast amounts of data. Other than the expected path towards an exploding number of data-driven businesses and scientific scenarios in the next few years, in this talk we also envision a future where data becomes readily available and its power can be harnessed by everyone. What both scenarios have in common is a need for new kinds of data systems which are tailored for data exploration, which are easy to use, and which can quickly absorb and adjust to new data and access patterns on-the-fly. In this talk, we will discuss this vision as well as recent and ongoing advances towards data systems which are tailored for data exploration.

## Speech Title: Open Language Grid - Towards a Global Language Service Infrastructure

Speaker: Dr. Toru Ishida  
Department of Social Informatics  
Kyoto University



**Biographical Sketch:** Dr. Toru Ishida has been a professor of Kyoto University since 1993. His academic background includes visiting scientist/professor positions at Columbia University, Technische Universität München, Université Pierre et Marie Curie, University of Maryland, Shanghai Jiao Tong University, Tsinghua University, Xinjiang University and Hong Kong Baptist University. He is a fellow of IEEE, a vice president of IEICE, and a member of the Science Council of Japan. He is a co-founder of the Department of Social Informatics, Kyoto University, and recently organized the Kyoto University Design School. His research interest lies with Autonomous Agents and Multi-Agent Systems and modeling collaboration within human societies. He contributed to create AAMAS/ICMAS/PRIMA conferences on Autonomous Agents and Multi-Agent Systems. His projects include Community Computing, Digital City Kyoto, Intercultural Collaboration Experiments, and the Language Grid.

**Abstract:** Intercultural collaboration, a subfield of social informatics, is a goal-directed group activity based on multilingual communication. We set its research target on collaboration, instead of communication, because we can clearly identify research issues when the goal of group activities is defined. To overcome language barriers and facilitate intercultural collaboration, we designed the Language Grid, which allows users to freely create language services from existing language resources and combine them to develop new services to develop their own communication environment. This talk explains the design concept and service architecture of the Language Grid together with our activity called YMC-Viet: a youth mediated communication project in Vietnam, where Japanese agricultural experts transfer knowledge to Vietnamese farmers in high illiteracy regions. By integrating various services registered to the Language Grid, we realized a communication channel between experts and farmers via children to bridge the significant communication gaps including language, knowledge, culture, and distance.



## Speech Title: How SQL Can Take Us from Hadoop to Energy Independence

Speakers: Ravi Chiruvolu (CEO of Slingshot)

Michael Hoskins (CTO of Actian)



**Biographical Sketch:** Ravi Chiruvolu is a CEO of Slingshot, Ravi was President of Noribachi (LED manufacturer), Managing General Partner for Charter Ventures (\$400m fund), and General Partner at Alta Partners (\$500m fund). Ravi held operating positions at Motorola, Ameritech, Peapod, and NASA Goddard and Johnson Flight Centers. He began his career in the Energy Practice at McKinsey & Company. Ravi holds a BS and MS in Mechanical Engineering from MIT (Robotics and Control Systems with minor in economics) and a MBA from the Harvard Business School. He is a First Lieutenant (Retired) in the United States Army, Ordinance Corps. Ravi is a tennis player and aspiring scratch golfer (more aspiring than scratch). His proudest accomplishment is coaching his 10 year old daughter's basketball team to a winning record.



**Biographical Sketch:** Michael Hoskins is a CTO who directs Actian's technology innovation strategies and evangelizes game-changing trends in big data, analytics, Hadoop and cloud to give insight into Accelerating Big Data 2.0™. Mike, a Distinguished and Centennial Alumnus of Ohio's Bowling Green State University, is a respected technology thought leader who has been featured in TechCrunch, Forbes.com, Datanami, The Register



and Scobleizer. Mike has been a featured speaker at events worldwide, including Strata NY + Hadoop World, keynoting at DeployCon, the “Open Standards and Cloud Computing” panel at the Annual Conference on Knowledge Discovery and Data Mining, the “Scaling the Database in the Cloud” panel at Structure, and the “Many Faces of Map Reduce - Hadoop and Beyond” panel at Structure Big Data. Mike received the AITP Austin chapter's Information Technologist of the Year Award for his leadership in developing Actian DataFlow, a highly parallelized framework to leverage multicore.

**Abstract:** All the technologies for energy independence have been present for some time. This session will explore how the only missing component -- the ability to analyze the economic, geographic, political and other factors that play into the cost of going solar -- is now attainable in Hadoop using SQL. Our current energy path has led us to reliance on foreign oil, a compromised economy and out-of-control carbon emissions. We've been hearing since the 1970s how solar will solve this, and by all rights it should -- every 15 seconds the sun gives the Earth enough energy to power humanity for a day. Why in 2014 is the solar home still a novelty? This session will explore how the thousands of diverse datasets which underlie the factors that play into the individual decision to go solar can be integrated and analyzed in Hadoop to develop a 5-year household roadmap to clean energy. Despite the complexity of the data behind the problem, Ravi Chiruvolu of Slingshot Power and Mike Hoskins, CTO of Actian, believe that Big Data will provide the answer which has eluded us for decades. They will discuss how data from the IofT, Google Earth, local government, NASA, drones, social media, and other sources can be understood using the common language of business analysis -- SQL -- to determine what is stopping households from going solar.

## Speech Title: Big Data versus Big Insight - A Case Study from Call Center Optimization



Dr. Bernhard Suhm  
Director Professional Services, AVOKE Call Center Analytics  
Raytheon BBN Technologies

**Biographical Sketch:** Bernhard Suhm joined the AVOKE Analytics team in 2006 to establish the professional services group. As Director of Professional Services he manages the delivery of data analysis and consulting services to the subscribers of the AVOKE end-to-end call analytics solution. Previously, as a Senior Scientist at BBN Technologies Bernhard co-developed many elements of the AVOKE Analytics solution. Bernhard has over 20 years of experience working with speech recognition, voice user interfaces, and data-driven methods to optimize contact center satisfaction and efficiency. He and has co-authored several patents, book chapters, and published papers on these topics. Prior to joining AVOKE, Bernhard was a Senior Consultant with the Enterprise Integration Group, a leading developer of voice user interfaces, and a Research Programmer with the Interactive Systems Laboratories at Carnegie Mellon University. Bernhard received a PhD in Computer Science specializing in speech user interfaces from Karlsruhe University in Germany.

**Abstract:** In recent years, larger amounts of data have become available than ever before. While most research in “big data” is focused on developing efficient data processing methods, equally important to gain novel insight from these large data sets are meaningful metrics and the ability to drill down to root cause within data sets. In this talk we describe how we accomplished this goal in our call center optimization work. - The metrics that matter in the call center environment are operational cost and caller experience. Agent labor cost represent 70% of the operational cost. Once we identify issues in how calls are handled, we estimate the frequency relative to the total inbound call volume, and measure the time agents spend on specific parts of the call handling process. The product of frequency and time spent per instance yields a simple model of issue impact on average agents call handling time (AHT). - Prior research suggests that effort is a better indicator of repeat business and overall customer experience than some commonly used survey-based measures of customer satisfaction, such as Net Promoter Score (NPS). To measure the impact on caller experience we developed a benchmark of 22 automated metrics of caller effort, such as time spent in the menu system, whether the caller was transferred to a second agent, or previously attempted to resolve their issue on a website. These effort-related metrics relate to both the difficulty of obtaining and the operational cost of delivering customer service, and thus represent a great criterion for call center optimization. - Having access to the complete caller experience, the complete call from dialing to hang up, has been invaluable to empower us to drill to root cause and deliver actionable recommendations. The complete experience is difficult to get by, especially when data is unstructured and fragmented like in many Big Data settings; however, the complete experience for a reasonably small subset of the data is sufficient for most problems of practical relevance. The talk will illustrate our methodology for mining

end-to-end customer service calls with case studies from our consulting practice, and demonstrate the cloud-based call analytics solution that captures anywhere from thousands to millions of live calls.

**Speech Title: SciDB: For scientists who want to do science, not computer science**



Marilyn Matz (CEO and co-founder of Paradigm4)



Alex Poliakov (Solutions Architect for Paradigm 4)

**Biographical Sketch:** Marilyn Matz is CEO and co-founder of Paradigm4, a big data analytics company accelerating data-driven discovery. She previously co-founded Cognex Corporation; is on the board of directors of LogMeIn, the technical advisory board of the Accelerated Cure Project, and co-chair of the Massachusetts Technology Leadership Council's Big Data cluster.

Alex has over a decade of experience in developing distributed database internals. As Solutions Architect for Paradigm 4, he helps researchers in scientific and commercial applications make optimal use of SciDB for their most challenging analytics.

**Abstract:** Breakthrough research increasingly depends on platforms that support data curation, exploration, and scalable complex analytics by multiple users. With the emergence of the Internet of Everything in the commercial and industrial worlds and with the advances in device and instrument technologies in the science world, there is an urgent need for scientists to be able to work easily with extremely large and diverse data sets without requiring supercomputers or extensive RAM. Neither conventional relational database management systems nor Hadoop-based systems satisfy these demands.

SciDB is an innovative computational database from MIT database researcher Mike Stonebraker. Its multidimensional array data model is well matched for scientific data: genomic measurements, sensor output, images, graphs and time series. It has been designed to meet the demands of rapid, interactive exploratory data analysis while guaranteeing data integrity and reproducible results in a multi-user, shared data environment. We'll present how SciDB has been used for dark matter physics, climatology, and precision medicine research.

## Speech Title: Harnessing the Power of the Multi Stakeholder Networks



Anne C. Bader  
Principal of Bader Resources, LLC

**Biographical Sketch:** Anne C. Bader has twenty five years' experience as a nongovernmental executive and consultant creating practical initiatives and networks that build local capacity and sustainable security. Her particular emphasis is on creating multinational, multi-stakeholder networks in peace, security, human rights and governance for clients in governments, business and education. A citizen of Canada and the United States, she is Principal of Bader Resources, LLC, an international consultancy in public affairs. In 2011, she founded the International Cybersecurity Dialogue to promote dialogue between policy makers and security technologists. She also serves as a Senior Associate Fellow of the Institute of Statecraft, London. Previously, Mrs. Bader served as a Senior Research Fellow and Director, Advanced Research Assessment Group, The Defense Academy of The United Kingdom, UK. Executive Vice President, The Fund for Peace; Vice President of the Atlantic Council of the United States; Trinity College and Business Executives for National Security, US. Her memberships include the National Press Club, Women in International Security and the Board of the Hungarian American Coalition. She received a NATO award for creating their TRUST strategic communications campaign for demilitarization and defense reform in the Former Soviet Union and was named an Honorary Citizen of Hungary for her capacity building and NATO membership work.

**Abstract:** Cyber technology is a tool that provides linkages to all aspects of earth's activities. Its power is unquestioned. Its capacity is not fully understood. Its infrastructure supports commerce, energy, finance, security, transport and communication. While governments and business continue to respond to cyber threats to privacy, property and infrastructure, we see increasing support for a true multi stakeholder approach to maintaining the freedom and innovation that this technology provides. Our report includes case studies of how small world networks of leaders, practitioners and theoreticians –agile, flexible and responsive - are becoming a realistic strategy to overcome the dysfunctionality of a fractured global operational landscape. ASE is one example.

## Operational Structure: Integrated but Independent



“In the long term, this infrastructure is what we need to preserve if we are going to ensure that the Internet remains innovative, free and open for the benefits of all users. It is this multi-stakeholder approach to Internet governance that ensures that no one organization can effectively control the Internet.”

## Speech Title: New Challenges for Big Data: Streaming Sensor Data for eHealth Interventions



Dr. Holly Brugge Jimison

Professor

College of Computer and Information Sciences  
and College of Health Sciences  
Northeastern University

**Biographical Sketch:** Holly Brugge Jimison is a Professor in the College of Computer and Information Sciences and the College of Health Sciences at Northeastern University in Boston, Massachusetts. Her research involves developing design principles for technology to enable patients to be more active and engaged in their health care. She currently serves as Director of Northeastern’s Consortium on Technology for Proactive Care with research projects on tailoring health interventions for older adults based on feedback from in-home monitoring, as well as computer modeling projects for assessing cognitive states using embedded metrics within adaptive computer games. Dr. Jimison recently served as a Technology Advisor and Health Scientist with the Office of Behavioral & Social Sciences Research at the National Institutes of Health. Her NIH work focused on defining new research directions for shared medical decision making and on “big data” initiatives related to developing clinically relevant behavioral markers based on the monitoring of health behaviors in the home and environment.

**Abstract:** There is an increasing focus on changing healthcare from being reactive and clinic- or hospital-based to being proactive and continuous, with an emphasis on interventions that make use of home monitoring and information/communications technology to facilitate scalable approaches for delivering care to the home. New developments in sensors, mobile apps and wireless devices have provided us with opportunities to track health behaviors. These new types of data streams present many new challenges to “Big Data” modelers. The issues go beyond just thinking about volume of data and how to summarize or store it. For the behavioral monitoring in the home and environment there are now additional issues of 1) how to model context, bias and noise from signals derived from opportunistic low-cost sensors; 2) how to infer activities and behaviors from multiple sources, often with differing sampling rates and accuracies; and 3) how to manage privacy and security of sensitive data . These new types of measures allow us to provide tailored health interventions with just-in-time feedback and support. These new monitoring techniques offer great promise for both reducing the cost of care and improving quality.



## Speech Title: Risk, Resilience, Blue Mountains Fire Ecology – Big Data and First Stop



CAPT (Assoc. Prof.) Dr. Simon Reay Atkinson RANR  
Faculty of Engineering and Information Technologies  
The University of Sydney

**Biographical Sketch:** Associate Professor, Doctor, Captain Simon Reay Atkinson RANR research into Complex Systems, the Cyber, Lodestone Programme, Synthetic Ecologies and Life Systems, Entanglement Theory, dynamic social networks / mapping (DSN/M), Leadership (command) & Management (and control), Network Law, Information Capture and Knowledge Exchange (ICKE), Knowledge Enterprise Economies (KEEs), Versatile Modular Systems (VMS™) and the Global Local Political Sûreté Economy (GPÊ) including Defence Economics, has applications as diverse as improving resilience in fire ecologies, managing risk, instabilities and uncertainties, delivering humanitarian relief and advising how we might recover from global recession and put humanity back in the loop.

Simon is a Captain in the RAN Reserves, a Fellow of the IET, he took the aircraft carrier HMS OCEAN from build to full operational status as its first Systems Engineer. He is an Associate at the Centre for International Security Studies; Director of the Masters of Project Leadership program with the Complex Systems Research Group at the University of Sydney and; held senior positions as UK Chief of Defence Staff Liaison Officer to the Australian Department of Defence, Capability Development Group; was Senior Systems Adviser to UK MoD during the Strategic Defence and Security Review; Cambridge University, Hudson Naval PhD Research Fellow; co-Lead General Petraeus' Command and Control and Knowledge Management Cell, Joint Strategic Assessment Team; Senior Research Fellow / Asymmetries Director, Advanced Research Assessment Group, UK Defence Academy.

**Abstract:** This panel presentation will examine Risk and Resilience based upon research into the October 2013 Blue Mountains Fires in New South Wales, Australia. It will examine the Fire Ecology of the Blue Mountains and how, based upon the examination of 'big data', the fire seasons may be changing. It will introduce the concept of a Synthetic Ecology from which a Fire Ecology may be derived. It will present on findings from an ecological fire risk register devised to enable an improved understanding of the management of the three phases of Prevent; Engage and Recover. It will introduce the concept of resilience and then look at Big Data requirements for creating



/ enabling resilience – including dynamic social mapping in disasters. This research led to white frequency management and to the development of First Stop by Random Hacks of Kindness (RHoK) for better information capture and knowledge exchange (ICKE) during a disaster – so improving both recovery and resilience in a community.

**Speech Title: Biomedical Big Data Analytics for Patient-Centric and Outcome-Driven Health Care**



Dr. May D. Wang  
Georgia Institute of Technology and Emory University

**Biographical Sketch:** Dr. May D. Wang is an Associate Professor in the Joint Department of Biomedical Engineering, School of Electrical and Computer Engineering, Winship Institute, IBB and IPaT at Georgia Institute of Technology and Emory University, USA. She is a Kavli Fellow, a Georgia Research Alliance Distinguished Cancer Scholar, Biocomputing and Bioinformatics Core Director in Emory-Georgia-Tech Center of Cancer Nanotechnology, and Co-Director of Georgia-Tech Center for Bio-Imaging Mass Spectrometry.

Prof. Wang's research is in Biomedical Big Data analytics with a focus on Biomedical and Health Informatics (**BHI**) for Personalized and Predictive Health. Her research includes high throughput NGS and -omic data mining to identify clinical biomarkers, bionanoinformatics, pathological imaging informatics to assist clinical diagnosis, critical and chronic care health informatics for evidence-based decision making, and predictive systems modeling to improve health outcome. Prof. Wang published 160+ peer-reviewed articles in **BHI** as the corresponding/co-corresponding author in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *Proceedings of The IEEE*, *IEEE Transactions on Information Technology in Biomedicine*, *Journal of American Medical Informatics Association*, *IEEE Journal of Biomedical and Health Informatics*, *Nature Protocols*, *Proceedings of National Academy of Sciences*, *Annual Review of Medicine*, *Circulation Genetics*, *Briefings in Bioinformatics*, *BMC Bioinformatics*, *Journal of Pathology Informatics*, *BMC Medical Imaging*, *Annals of BME*, *Trends in Biotechnology*, and *Nanomedicine* etc. She has led RNA-data analysis investigation in sequencing project within FDA-led MAQC Consortium. In addition, Dr. Wang has devoted to training young generation of data scientists & engineers, and was awarded Georgia-Tech's Outstanding Faculty Mentor for Undergraduate Research in 2005.

Currently, Prof. Wang serves as the Senior Editor for IEEE Journal of Biomedical and Health Informatics (J-BHI), an Associate Editor for IEEE Transactions on Biomedical Engineering (TBME), and an Emerging Area Editor for Proceedings of National Academy of Science (PNAS). She is an IEEE-EMBS 2014-2015 Distinguished Lecturer, and IEEE EMBS Biomedical and Health Informatics Technical Committee Chair.

**Abstract:** Rapid advancements in biotechnologies such as -omic (genomics, proteomics, metabolomics, lipidomics etc.), next generation sequencing, bio-nanotechnologies, molecular imaging, and mobile sensors etc. accelerate the data explosion in biomedicine and health wellness. Multiple nations around the world have been seeking novel and effective ways to make use of “big data” for evidence-based, outcome-driven, and affordable 5P (Patient-centric, Predictive, Preventive, Personalized,

and Precise) healthcare. I have developed multi-modal and multi-scale (i.e. molecular, cellular, whole body, individual, and population) biomedical data analytics research program on discovery, development, and delivery for health care. Specifically, we have performed translational bioinformatics in biomarker discovery for personalized care; imaging informatics in histopathology for clinical diagnosis decision support; bionanoinformatics for minimally-invasive image-guided surgery; critical care informatics for real-time evidence-based decision making; chronic care informatics for patient-centric health; and healthcare workflow optimization for effectiveness. Our research has been supported by NIH, NSF, Georgia Research Alliance, Georgia Cancer Coalition, Emory-Georgia Tech Cancer Nanotechnology Center, Children's Health Care of Atlanta, Atlanta Clinical and Translational Science Institute, Microsoft Research and HP.

In this short talk, I will summarize a few key points in Biomedical Big Data Analytics: (1) major challenges in biomedical and health informatics pipeline that consist of data quality control, information feature extraction, advanced knowledge modeling, decision making, and proper action taking through feedback; (2) progresses and opportunities in data integrity and integration, case-based reasoning for individualized care, and streaming data analytics for real-time decision support; and (3) patient-centric educational intervention, community-based crowd sourcing, and Biomedical Data Analytics MOOC development that educate healthcare stakeholders (i.e. patients, physicians, payers, and hospitals) and train young data scientists and engineers.

**Speech Title: Cybersecurity Education and Workforce Development  
at the National Science Foundation**



Dr. Corby Hovis  
Lead Program Director  
Division of Undergraduate Education  
National Science Foundation

**Biographical Sketch:** Corby Hovis is a senior program director at the National Science Foundation (NSF) in Arlington, Virginia, where he oversees the NSF-wide Research Experiences for Undergraduates (REU) program, which provides structured research experiences to over 5,000 college and university students each year. He also co-manages several grant programs in NSF's Directorate for Education and Human Resources, including those that support cybersecurity education and workforce development.

Before coming to NSF, Dr. Hovis served on the faculty of the honors college at Valparaiso University in Valparaiso, Indiana, and, at the same time, as science editor and associate producer at Encyclopædia Britannica and Britannica.com in Chicago. He earned his graduate degrees (Ph.D., M.S., M.A.) from Cornell University and his undergraduate degree from Wake Forest University. During his tenure at NSF, he also spent a year as an American Council on Education (ACE) Fellow in the Office of the President at The Ohio State University.

**Abstract:** The National Science Foundation (NSF) offers an array of grant programs that support the improvement of cybersecurity education at all levels and the preparation of the next generation of professionals in computing and security. This talk will provide an overview of the history and scope of these opportunities. The following programs and others will be discussed.

The *CyberCorps: Scholarship for Service (SFS)* program provides full scholarships and stipends for students in information assurance and computer security in exchange for their service as cybersecurity professionals in a Federal Government agency following graduation. The program also supports projects focusing on curriculum development, program development, professional development for faculty, outreach, and other activities that improve educational quality and increase the production of cybersecurity professionals in colleges and universities.

The *Advanced Technological Education (ATE)* program provides merit-based grants to improve the education of technicians for a range of high-tech fields that drive the U.S. economy, including cybersecurity and digital forensics. The program focuses on community colleges and encourages partnerships between colleges, universities, K-12 schools, and employers to create technician education programs that are relevant to current needs in business, industry, and government. Funded projects commonly undertake curriculum development, professional development for college faculty and secondary school teachers, and the building of career pathways from high schools to two-year colleges and from two-year colleges to four-year institutions.

The *Secure and Trustworthy Cyberspace (SaTC)* program supports a broad spectrum of research to improve the resilience of individual hosts, networked systems, hardware, software, applications, and critical infrastructure from malicious cyber-attacks while preserving privacy and promoting usability. In addition, the program supports projects that leverage results from basic research in cybersecurity and research on student learning to address the challenge of expanding existing educational opportunities and resources in cybersecurity.

The *Research Experiences for Undergraduates (REU)* program funds institutions to host cohorts of usually 8-10 students who conduct closely mentored, independent research projects in a wide range of STEM fields. Several REU Sites focus on student research in cybersecurity.

---

## Poster Session

December 16, 2014, 12:00-13:30

---

- **Big Data Application in Healthcare**  
Sathish Kumar Pandian Anbalagan
- **Comparison of Classifier Accuracy Between Normally distributed OFI to OFI with Different Inputs on Categorical Data**  
Raveendra Babu B and Lakshmi Sreenivasa Reddy
- **Toward Big Data-based Collaborative Decision Making (CDM) in Civil Aviation**  
Zhijun Wu
- **A Novel Approach for Designing and Making Jute Cards**  
Miftahur Rahman and Jarib Rahman
- **Supervized Community Detection for Big Data and Large-Scale Complex Networks**  
Pravin Chopade, Justin Zhan, and Marwan Bikdash
- **Identifiability of Vehicle Tollgate Records: The Milan Tollgate Dataset**  
Nick Manfredi, Darakhshan Mir, and Claire Schlenker
- **Comprehensive Branch Elimination for GPU Accelerated Biomedical Applications**  
Lucas Vespa and Alex Bauman
- **Big Data Issues with Streaming Sensor Data from the Home and Environment**  
Xuan “Sean” Li, Misha Pavel, and Holly Jimison
- **Windows Installer Security**  
C. Kadluba, M. Mulazzani, L. Zechner, S. Neuner, and E. Weippl
- **Drowning in Opinions: Extracting the Pearls**  
Bradley Meyer and Justin Zhan
- **Data Deanonymisation: A Framework for Analysis**  
Dalal Al-Azizy
- **Social Cohesion Radicalism and Its Impact on a City’s Resilience**  
Faisal Saman
- **Cyber Attack Resilience**  
Chun To Lau

- **A Resilient Sydney: A Preparatory Cultural Measure for the Future of Public Transport**  
James Carlton
- **Are You Saying What I Think You're Feeling?- Detecting Psycho-social Dimensions through Socio-linguistics**  
Peggy Wu, Christopher Miller, Tammy Ott, Sonja Schmer-Galunder, and Jeff Rye
- **Node Degree and Edge Clustering Correlation for Community Detection in Big Data and Large-Scale Networks**  
Haysam Selim, Pravin Chopade, and Justin Zhan
- **Discovering Hidden Networks Based On Twitter -Texts**  
Ahmad Albu-Shamah and Justin Zhan
- **Digital Image Authentication in Social Media**  
Balkirat Kaur, Malcolm Blow, and Justin Zhan
- **Uncovering the culture code of states in U.S. with use of Amazon.com online customer reviews**  
Wonsang Lee and So Young Sohn
- **A Two-Dimensional Screening Method for Multi Criteria Reciprocal Selection in social Systems**  
Zeynab Bahrami Bidoni, Roy George, and Ahmad Makui
- **Towards A Recommender System for DataBridge**  
Yogeshwar Rao Bachupally and Justin Zhan

# ASE

Academy of Science and Engineering



## ASE@360: Open Scientific Digital Library

### Quick Facts

- Not for profit
- Open source platform
- No physical boundary
- Round the clock availability
- Robust discovery mechanism
  - Collaborative
- Financially sustainable
- Operational environment

## Join ASE@ [www.asesite.org](http://www.asesite.org)

- ASE K-12 Community
- ASE Student Community
- ASE Academic Clusters
  - ASE Industry Forum
- ASE Professional Hub
- ASE Research Pentagon

