

2014

**The Second ASE International Conference on Big
Data Science and Computing
The Sixth ASE International Conference on Social
Computing
The Third ASE International Conference on Cyber
Security**



**May 27 - May 31, 2014
Tresidder Memorial Union
Stanford University,
459 Lagunita Drive, Stanford,
CA, USA, 94305**

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

www.asesite.org



Table of Contents

Welcome Message from the Conference Steering Chair.....	Page 2
Conference Committee	Page 3-10
Workshop Committee	Page 11-20
Conference Program/Sessions/Papers	Page 21-30
Workshop Program/Sessions/Papers.....	Page 31-48
Speaker Biographical Sketches and Abstracts.....	Page 49-84
Poster Session P1 May 28, 2014 (Posters and Papers).....	Page 85-86
Poster Session P2 May 29, 2014 (Posters and Papers).....	Page 87-88
Poster Session P3 May 30, 2014 (Posters and Papers).....	Page 89
Exhibitors Directory.....	Page 90

Welcome Message from the Conference Steering Chair



On behalf of the Academy of Science and Engineering (ASE), welcome to the Second ASE International Conference on Big Data Science and Computing, the Sixth ASE International Conference on Social Computing and the Third ASE International Conference on Cyber Security. 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 science, cyber security, and social computing. These ASE conferences are leading international forum for big data, social computing, cyber security 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 science, cyber security, and social 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, cyber security, and social 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, cyber security, and social 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 Stanford University facilities in making this event a reality.

We look forward to the conferences.

Sincerely,

Justin Zhan
Conference Steering Chair

The Second ASE International Conference on Big Data Science and Computing

Conference Committee

Conference Chairs

- Chaitanya Baru, University of California-San Diego
- Paul Kim, Stanford University
- Kai Hwang, University of Southern California
- L.W. Chang, Academy of Science and Engineering
- Merce Crosas, Harvard University
- Howard Lander, RENCI/UNC-Chapel Hill
- Arcot Rajasekar, University of North Carolina-Chapel Hill
- Stanley Ahalt, RENCI/UNC-Chapel Hill
- Tom Carsey, University of North Carolina-Chapel Hill
- Justin Zhan, North Carolina A&T State University
- Srinivas Aluru, Georgia Institute of Technology
- Yixin Chen, Washington University in St. Louis
- Longbing Cao, University of Technology, Sydney/Australia

Program Committee

- Evrim Acar, University of Copenhagen, Denmark
- Rajendra Akerkar, Western Norway Research Institute, Norway
- Leman Akoglu, Stony Brook University, USA
- Plamen Angelov, Lancaster University, UK
- Renato Assuncao, Federal University of Minas Gerais, Brazil
- Guy Boy, Florida Institute of Technology, USA
- Ulf Brefeld, TU Darmstadt, Germany
- David Camacho, Autonomous University of Madrid, Spain
- Sharma Chakravarthy, The University of Texas at Arlington, USA
- Varun Chandola, State University of New York at Buffalo, USA
- Nitesh Chawla, University of Notre Dame, USA
- Yixin Chen, Washington University in St. Louis, USA
- Jinchuan Chen, Renmin University of China, China
- Shu-Ching Chen, Florida International University, USA
- Gregorio Convertino, Informatica Corporation, USA
- Martin Farach-Colton, Rutgers University, USA
- Yun Raymond Fu, Northeastern University, USA
- Brian Gallagher, Lawrence Livermore National Laboratory, USA
- Fosca Giannotti, ISTI- CNR-Pisa, Italy
- Robert Gillen, Oak Ridge National Laboratory, USA
- Ruairi Glynn, University College London, UK
- Yuhong Guo, Temple University, USA

- Haibo He, University of Rhode Island, USA
- Xiaolin Hu, Georgia State University, USA
- Jun Luke Huan, University of Kansas, USA
- Joshua Huang, Shenzhen Institute of Advanced Technology, China
- Akihiro Inokuchi, Kwansei Gakuin University, Japan
- Ruoming Jin, Kent State University, USA
- Rob Johnson, Stony Brook University, USA
- Anna Jordanous, King's College London, UK
- George Karypis, University of Minnesota, USA
- Xiaolin (Andy) Li, University of Florida, USA
- Jiuyong Li, University of South Australia, Australia
- Xue Li, The University of Queensland, Australia
- Xiaoli Li, A*STAR, Singapore
- Tao Li, Florida International University, USA
- Shou-de Lin, National Taiwan University, Taiwan
- R. Ryan McCune, University of Notre Dame, USA
- Muhammad Asif Naeem, Auckland University of Technology, Auckland
- Hung Son Nguyen, Warsaw University, Poland
- Yonghong Peng, University of Bradford UK
- Victor Sheng, University of Central Arkansas, USA
- Karsten Steinhäuser, University of Minnesota, USA
- Masashi Sugiyama, Tokyo Institute of Technology, Japan
- Shan Suthaharan, UC-Berkeley, USA
- Krishnaprasad Thirunarayan, Wright State University, USA
- Vincent Tseng, National Cheng Kung University, Taiwan
- Eli Upfal, Brown University, USA
- Raju Vatsavai, Oak Ridge National Laboratory, USA
- Haifeng Wang, BaiDu, China
- Ted Willke, Intel Corporation, USA
- Raymond Wong, Hong Kong University of Science and Technology, China
- Junjie Wu, Beihang University, USA
- Chunsheng Yang, National Research Council of Canada, Canada
- Mi-Yen Yeh, Academia Sinica, Taiwan
- Min-Ling Zhang, Southeast University, China
- Yu Zhang, Hong Kong Baptist University, China
- Kun Zhang, Max Planck Institute for Intelligent Systems, Germany
- Xing Zhao, FICO, USA
- Ying Zhao, Tsinghua University, China
- Bin Zhou, University of Maryland, USA
- Xingquan Hill Zhu, Florida Atlantic University, USA
- Jun Zhu, Tsinghua University, China
- Jaroslaw Zola, Rutgers University, USA



The Sixth ASE International Conference on Social Computing

Conference Committee

Conference Chairs

- T. Y. Lin, San Jose State University, USA
- Kai Hwang, University of Southern California, USA
- Paul Kim, Stanford University, USA
- L.W. Chang, Academy of Science and Engineering, USA
- Justin Zhan, North Carolina A&T State University, USA
- Longbing Cao, University of Technology, Sydney, Australia
- Shujia Zhou, Northrop Grumman Information Systems, USA

International Advisory Board

- Kevin Almeroth, University of California, Santa Barbara, USA
- Adrian David Cheok, National University of Singapore, Singapore
- Noshir Contractor, Northwestern University, USA
- Irfan Essa, George Institute of Technology, USA
- David Lazer, Northeastern/Harvard University, USA
- Dongman Lee, KAIST, Korea
- Ramesh Jain, University of California, Irvine, USA
- Tom Malone, Massachusetts Institute of Technology, USA
- Kenji Mase, Nagoya University, Japan
- Anton Nijholt, University of Twente, The Netherlands
- Sandy Pentland, Massachusetts Institute of Technology, USA
- Whitman Richards, Massachusetts Institute of Technology, USA
- Bernt Schiele, TU Darmstadt, Germany
- Matthew Turk, University of California, Santa Barbara, USA
- Alessandro Vespignani, Indiana University, USA
- Alessandro Vinciarelli, University of Glasgow, UK, and Idiap Research Institute, Switzerland
- Wenjun Zhang, Shanghai Jiaotong University, USA

Program Committee

- Aimeur Esma, University of Montreal, Canada
- Michael Fire, Ben-Gurion University, Israel
- Rushed Kanawati, LIPN, France
- Masaki Aida, Tokyo Metropolitan University, Japan
- Oya Aran, IDIAP Research Institute, Switzerland
- Anneli Avatare Nou, SICS Swedish ICT, Sweden
- Guido Barbian, University of Leuphana, Germany
- Philipp Berger, University of Potsdam, Germany

- Nadia Berthouze, University College London, UK
- Siddhartha Bhattacharyya, University of Illinois at Chicago, USA
- Nick Campbell, University of Dublin, UK
- Angelo Cangelosi, University of Plymouth, UK
- Aleksandra Cerekovic, Zagreb University, Croatia
- Charalampos Chelms, University of Southern California, USA
- Mohamed Chetouani, Pierre and Marie Curie University, France
- Marco Cristani, University of Verona, Italy
- Kerstin Dautenhahn, The University of Hertfordshire UK
- Yves-Alexandre de Montjoye, MIT Media Lab, USA
- Thang Dinh, Virginia Commonwealth University, USA
- Chamari Edirisinghe, National University of Singapore, Singapore
- Elod EGYED-ZSIGMOND, INSA-Lyon, France
- Thomas Erickson, IBM T. J. Watson Research Center, USA
- Enrique Frias-Martinez, Telefonica Research, Madrid, Spain
- Wai-Tat Fu, University of Illinois at Urbana-Champaign, USA
- David Garcia, ETH Zurich, Switzerland
- Daniel Gayo-Avello, University of Oviedo, Spain
- Rachel Greenstadt, Drexel University, USA
- Christopher Griffin, Penn State University
- Georg Groh, TUM, Germany
- Lu Han, Google, USA
- Cynthia Hood, Illinois Institute of Technology, USA
- Jun Hu, Eindhoven University of Technology, The Netherlands
- Henric Johnson, Blekinge Institute of Technology, Sweden
- Joemon Jose, University of Glasgow, UK
- Krzysztof Juszczyszyn, Wroclaw University of Technology, Poland
- Carolin Kaiser, University of Erlangen-Nuremberg, Germany
- Jugal Kalita, University of Colorado, USA
- Jihie Kim, USC ISI, USA
- John Korah, The University of Texas at El Paso, USA
- Luis Lamb, Federal University of Rio Grande, Brazil
- Dongman Lee, KAIST, South Korea
- Bruno Lepri, FBK, Italy
- Mark Levene, Birkbeck University of London, UK
- Malik Magdon-Ismail, Rensselaer Polytechnic Institute, USA
- Lorenzo Magnani, University of Pavia, Italy
- Matthias Mehl, University of Arizona, USA
- Ronaldo Menezes, Florida Institute of Technology, USA
- Carlo Francesco Morabito, University of Mediterranean, Italy
- Vittorio Murino, IIT Foundation, Italy
- Mirco Musolesi, University of Birmingham, UK

- H.J.A. (Rieks) op den Akker, University of Twente , The Netherlands
- Scott Piao, Lancaster University, UK
- Viktor Prasanna, University of Southern California, USA
- Hamid Rabiee, Purdue University, USA
- Vaskar Raychoudhury, Indian Institute of Technology Roorkee, India
- Antonio Sanfilippo, Pacific Northwest National Lab, USA
- Eunice Santos, University of Texas at EI Pasco, USA
- Eugene Santos, Dartmouth College, USA
- Man-Kwan Shan, National Chengchi University, Taiwan
- Mohamed Shehab, University of North Carolina at Charlotte, USA
- Gita Reese Sukthankar, University of Central Florida, USA
- Jerzy Surma, University of Massachusetts, USA
- My Thai, University of Florida, USA
- Dan Thomsen, Smart Information Flow Technologies, USA
- Jan Treur, Vrije University, the Netherlands
- Matthew Turk, University of California-Santa Barbara, USA
- William Wallace, Rensselaer Polytechnic Institute, USA
- Jon Whittle, Lancaster University, UK
- Dinghao Wu, Penn State University, USA
- John Yen, Pen State University, USA



The Third ASE International Conference on Cyber Security

Conference Committee

Conference Chairs

- Kai Hwang, University of Southern California
- Paul Kim, Stanford University
- T. Y. Lin, San Jose State University
- L.W. Chang, US-Patent and Trademark Office
- Justin Zhan, North Carolina A&T State University
- Longbing Cao, University of Technology, Sydney/Australia

Program Committee

- Riham Abdel Moneim, The American University in Cairo, Egypt
- Robert Abercrombie, Oak Ridge National Laboratory, USA
- Claudio Agostino Ardagna, The University of Milann, Italy
- T. Scott Ankrum, The MITRE Corporation, USA
- Yan Bai, University of Washington, USA
- Samik Basu, Iowa State University, USA
- Kevin Bauer, University of Waterloo, Canada
- Amiya Bhattacharya, Arizona State University, USA
- Hasan Cam, US Army Research Laboratory, USA
- Jianneng Cao, Purdue University, USA
- Marco Carvalho, Florida Institute of Technology, USA
- Songqing Chen, George Mason University, USA
- Saumya Debray, University of Arizona, USA
- Tassos Dimitriou, Kuwait University, Kuwait
- Lilian Edwards, Strathclyde University, Glasgow
- Manuel Egele, Northeastern University, USA
- Khalil el-khatib, University of Ontario Institute of Technology, Canada
- Lia Emanuel, University of Bath, UK
- Stefan Fenz, TU Vienna, Austria
- Suzanne Fox Buchele, Ashesi University, Ghana
- Gabriela Gheorghe, University of Luxembourg, Luxembourg
- Kartik Gopalan, Binghamton University, USA
- Julius Holderer, University of Freiburg, Germany
- Wei Jiang, Missouri University of Science and Technology, USA
- James Joshi, University of Pittsburgh, USA
- Jan Jürjens, TU Dortmund & Fraunhofer ISST, Germany
- Hiroaki Kikuchi, Meiji university, Japan
- Ralf Kuesters, University of Trier, Germany
- Nicolai Kuntze, Fraunhofer SIT, Germany

- Kwok-Yan Lam, National University of Singapore, Singapore
- Michael Lewis, Chevron, USA
- Zhenkai Liang, National University of Singapore, Singapore
- Alex Malin, Los Alamos National Laboratory, USA
- Shivakant Mishra, University of Colorado, USA
- Kegomoditswe Molotsi, Deloitte, USA
- Thomas Morris, Mississippi State University, USA
- Ravi Mukkamala, Old Dominion University, USA
- Simin Nadjm-Tehrani, Linkoping University, Sweden
- Cyril Onwubiko, Research Series Ltd, UK
- Zachary Oster, University of Wisconsin, USA
- Andriy Panchenko, University of Luxembourg, Luxembourg
- Stefano Paraboschi, University of Bergamo, Italy
- Andreas Pashalidis, Katholieke University Leuven, Belgium
- Celia Ghedini Ralha, University of Brasilia, UK
- Srin Ramaswamy, ABB Corporate Research, India
- Indrakshi Ray, Colorado State University, USA
- Kui Ren, Illinois Institute of Technology, USA
- William Robertson, Northeastern University, USA
- Yaoping Ruan, IBM TJ Watson Research Center, USA
- Ravi Sandhu, Univ of Texas at San Antonio, USA
- Ashutosh Saxena, Infosys Labs, Infosys Limited, USA
- Nitesh Saxena, University of Alabama, USA
- Walt Scacchi, University of California, Irvine, USA
- Robert Schaefer, MIT Haystack Observatory, USA
- Kent Seamons, Brigham Young University, USA
- Jean-Pierre Seifert, Deutsche Telekom AG, Berlin
- Pravin Shankar, Loopt, USA
- Kaushal Solanki, Mayachitra Inc, USA
- Makoto Takizawa, Seikei University, Japan
- Vicenc Torra, IIIA-CSIC, Catalonia, Spain
- Margus Veanes, Microsoft Research USA
- Thijs Veugen, Delft University of Technology, The Netherlands
- Bárbara Vieira, Radboud University, Nijmegen
- Xinyuan Frank Wang, George Mason University, USA
- Lingyu Wang, Concordia University, Canada
- Paul Watters, University of Ballarat, Australia
- Edgar Weippl, Security Business Austria, Austria
- David Whelihan, MIT Lincoln Laboratory, USA
- Mary-Anne Williams, Stanford University, USA
- Ted Willke, Intel, USA
- Dinghao Wu, Penn State University, USA

- Attila Yavuz, Robert Bosch Research and Technology Center, USA
- Xun Yi, Victoria University, Australia
- Michael Yip, University of Southampton, USA
- Ting Yu, North Carolina State University, USA
- Sotirios Ziavras, New Jersey Institute of Technology, USA
- Urko Zurutuza, University of Mondragon, Spain

International Workshop on Big Data Analytics for Predictive Organization and Big Transformations

Workshop Organizing Committee

Pravin Chopade, Justin Zhan

ILAB, North Carolina A&T State University, Greensboro, NC-27411, USA

Program Committee Chairs

- Jalal Mapar, U.S. DHS S&T Directorate, Washington, DC, USA
- Una-May O'Reilly, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Program Committee Members

- Rajeev Agrawal, North Carolina A&T State University, NC, Greensboro, USA
- Marwan Bikdash, North Carolina A&T State University, Greensboro, NC, USA
- Gabriela F. Ciocarlie, SRI International, Menlo Park, California, USA
- Andrew Collins, Old Dominion University, Suffolk, VA, USA
- Robert M. Connors, Raytheon, Tewksbury, Massachusetts, USA
- Kenneth Crowther, MITRE Corporation, McLean, VA, USA
- Dukka KC, North Carolina A&T State University, Greensboro, NC, USA
- Vinaya Kelkar, North Carolina A&T State University, Greensboro, NC, USA
- Kishore Mane, intiGrow, Duluth, GA, USA
- George Markowsky, The University of MAINE, Orono, ME, USA
- Nader Mehravari, Carnegie Mellon University, Pittsburgh, PA, USA
- Christie Nelson, CCICADA, Rutgers University, Piscataway, New Jersey, USA
- William M. Pottenger, CCICADA, DHS S&T, Rutgers University, & Intuidex, Inc, USA
- Rajendra Raj, Rochester Institute of Technology, Rochester, NY, USA
- Carol Romanowski, Rochester Institute of Technology, Rochester, NY, USA
- Sharad Sharma, Bowie State University, Bowie, MD, USA

International Workshop on Distributed Storage Systems and Coding for Big Data

Workshop Organizing Committee

General Chairs

- S. Y. Robert Li, The Chinese University of Hong Kong, Hong Kong

Co-Chairs

- Hui Li, Peking University, China
- Kenneth W. Shum, Institute of Network Coding, the Chinese University of Hong Kong, Hong Kong

Program Committee Members

- Ming Xiao, Royal Institute of Technology, Sweden
- Minghua Chen, The Chinese University of Hong Kong, Hong Kong
- Shujiang Zhao, Huawei Technology, China
- Patrick P. C. Lee, the Chinese University of Hong Kong, Hong Kong
- Ke Chu, ZTE Corporation, China
- Kenneth W. Shum, Institute of Network Coding, CUHK, Hong Kong
- Xiao MA, Sun Yat-sen University, China
- WaiHo MOW, Univ. of Science and Technology, Hong Kong
- Yuesheng Zhu, Peking University, China
- Manish K Gupta, Dhirubhai Ambani Institute of Info.& Comm. Tech.(DA-IICT), India
- Salim El Rouayheb, Illinois Institute of Technology, USA
- Xin WANG, Fudan University, China
- Jianshe MA, Tsinghua Univ., China
- Yinlong XU, University of Science and Technology of China
- Qiang Cao, Huazhong University of Science and Technology
- Hui Li, Peking University, China



International Workshop on Enabling Science from Big Image Data

Workshop Organizing Committee

Peter Bajcsy

Computer Scientist
Information Technology Laboratory (ITL), NIST,
100 Bureau Drive MS 8970
Gaithersburg, MD 20899-8970
Office Phone: 301-975-2958
Fax: 301-975-6097
Email: peter DOT Bajcsy AT nist DOT gov
URL: <http://www.nist.gov/itl/ssd/is/bajcsy.cfm>

Yousef Khalidi

Distinguished Engineer at Microsoft
Microsoft Corporation
One Microsoft Way
Redmond, Washington 98052
United States
Email: yousefkhalidi AT Comcast DOT net

Hanchuan Peng

Associate Investigator
Allen Institute for Brain Science
551 N 34th St
Seattle, WA 98103
Phone: (206) 548-7000
Email: hanchuanp AT alleninstitute DOT org

Kirk D Borne, Professor

George Mason University,
School of Physics, Astronomy, and Computational Sciences
4400 University Drive MS 6A2
Fairfax 22030
United States
Phone: +1 703 993 8402
Fax: +1 703 993 1269
Email 1: kborne AT gmu DOT edu

Technical Program Committee Members

- Parvez Ahammad, Janelia Farm Research Campus, Howard Hughes Medical Institute (HHMI)
- Peter Bajcsy, National Institute of Standards and Technology (NIST)
- Tomasz Bednarz, The Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia
- Kirk Borne, George Mason University (GMU)
- Mary Brady, National Institute of Standards and Technology (NIST)
- Douglas Burke, Harvard-Smithsonian Center for Astrophysics
- Michael Cho, University of Illinois at Chicago (UIC)
- Amit Chourasia, San Diego Supercomputing Center (SDSC)
- Lincoln Cobb, U.S. Department of Transportation
- Mark Conrad, National Archives and Records Administration (NARA)
- Maric Dragan, National Institute of Health (NIH), National Institute of Neurological Disorders and Stroke (NINDS)
- Daniel Duffy, Science and Exploration Directorate, National Aeronautics and Space Administration (NASA)
- David Forsyth, University of Illinois at Urbana-Champaign (UIUC)
- Natalie Gluck, Institute for Defense Analyses (IDA)
- David Gutman, Emory University
- Michael Hawrylycz, Allen Institute for Brain Science
- Daniel Hoepfner Lieber, Institute for Brain Development
- Xiaoming Huo, National Science Foundation (NSF)
- Yousef Khalidi, Microsoft Corporation
- Jason Kinser, George Mason University (GMU)
- Marcin Kocielek, Technical University of Lodz
- Sang-Chul Lee, Inha University, South Korea
- Richard Marciano, University of North Carolina-Chapel Hill
- Kenton McHenry, National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign (UIUC)
- Hanchuan Peng, Allen Institute for Brain Science
- Anne Plant, National Institute of Standards and Technology (NIST)
- Stephen Plaza, Janelia Farm Research Campus, Howard Hughes Medical Institute (HHMI)
- John Henry Scott, National Institute of Standards and Technology (NIST)
- Lior Shamir, Michigan Technological University (MTU)
- Amitabh Varshney, University of Maryland at College Park (UMD)
- Seung-Chul Yoon, U.S. Department of Agriculture (USDA)

International Workshop on Social Computing for Urban Intelligence and Smarter Lives (SMARTLIFE 2014)

Workshop Organizing Committee

General Chairs

- Dr. Fernando Koch, Director R&D, SAMSUNG Research Institute
- Sergio Borger, Senior Manager, IBM Research – Brazil
- Dr. Siome Goldenstein, Institute of Computing, University of Campinas, Brazil

Steering Committee

- Prof. Dr. Jean Paul Jacobs, Research Emeritus, IBM Research – Almaden
- Dr. Camille Crittenden, Deputy Director, CITRIS @ Berkeley University (TBC)

Program Committee Members

- Christian Guttman (IBM Research, Australia)
- Carlos Cardonha (IBM Research, Brazil)
- Kent Steer (IBM Research, Australia)
- Jan Richter (IBM Research, Australia)
- Fabio Piva (SAMSUNG Research Institute, Brazil)
- Andrew Koster (SAMSUNG Research Institute, Brazil)
- Philippe Pasquier (Simon Fraser University, Canada)
- Frank Dignum (Utrecht University, The Netherlands)
- Jim Black (Nossal Institute for Global Health, Australia)
- Gabriela Celani (UNICAMP, Brazil)
- Guido Araujo (UNICAMP, Brazil)
- Ana Bazzan (UFRGS, Brazil)
- Rafael Bordini (PUC-RS, Brazil)
- Dinei Florencio (Microsoft Research Redmond, EUA)
- Mauro Barni (University of Siena, Italy)

International Workshop on Cloud Computing (CLOUDCOM 2014)

Workshop Organizing Committee

Workshop Chairs

- ANTONIO MUÑOZ, University of Málaga, Spain
- ERNESTO DAMIANI, University of Milan, Italy

Program Committee Members

- ALCARAZ-CALERO, JOSE M., Hewlett-Packard, UK
- ANTON DEL PINO, PABLO, U. of Málaga, Spain
- BOYD, COLIN, Queensland U. of Tech., Australia
- COLLBERG, CHRISTIAN, Arizona University, USA
- CUELLAR, JORGE, Siemens, Germany
- DAVIDS, CAROL, Illinois Institute of Technology, USA
- DUSIT NIYATO, Nanyang Technological U., Singapore
- ENDICOTT-POPOVSKY, BARBARA, U. of Washington, USA
- FERNANDEZ, EDUARDO B., Florida Atlantic U., USA
- GIORGINI, PAOLO, University of Trento, Italy
- GRAWROCK, DAVID, Intel, USA
- GÜRGENS, SIGRID, Fraunhofer SIT, Germany
- JÜRJENS, JAN, TU of Dortmund, Germany
- KIYOMOTO, SHINSAKU, KDDI R&D Labs, Japan
- LAMBRINOUDAKIS, COSTAS, U. of Piraeus, Greece
- LAURENCE T. YANG, St Francis Xavier University, Canada
- LEVI, ALBERT, Sabanci University, Turkey
- LOEVENICH, DANIEL, Bundesamt für Sicherheit in der Informationstechnik (BSI), Germany
- LOSAVIO, MICHAEL, U. of Kentucky, USA
- LOTZ, VOLKMAR, SAP AG, France
- MAÑA, ANTONIO, University of Malaga, Spain
- MARTINELLI, FABIO, CNR-IIT, Italy
- MARTINEZ-PEREZ, GREGORIO, U. of Murcia, Spain
- MICHELE BEZZI, SAP, France
- NADARAJAM, R., PSG College of Technology, India
- POSEGGA, JOACHIM, U. of Passau, Germany
- PRESENZA, DOMENICO, Engineering, Italy
- QUISQUATER, JEAN-JACQUES, U. Catholique De Louvain, Belgium
- RAY, INDRAKSHI, Colorado State University, USA

- RUDOLPH, CARSTEN, Fraunhofer SIT, Germany
- SABETTA, ANTONINO, SAP, France
- SORIA-RODRIGUEZ, PEDRO, ATOS R&D, Spain
- SKIANIS CHARALABOS, University of Aegean, Greece
- SPANOUDAKIS, GEORGE, City University, UK
- WASHIZAKI, HIRONORI, Waseda University, Japan
- WESPI, ANDREAS, IBM, Switzerland
- YOSHIOKA, NOBUKAZU, Nat. I. of Informatics, Japan
- ZULKERNINE, MOHAMMAD, Queen's U., Canada

International Workshop on Exploiting Big Data in Commerce and Finance

Workshop Organizing Committee

Program Committee Chairs

- Vivian Lee, Fujitsu Laboratories of Europe Limited, UK
- Aisha Naseer, Fujitsu Laboratories of Europe Limited, UK
- Vit Novacek, INSIGHT @ NUI Galway (formerly known as DERI), Ireland

Program Committee Members

- Takahide Matsutsuka (Fujitsu Laboratories Limited Europe, UK)
- Masatomo Goto (Fujitsu Laboratories Limited Europe, UK)
- Eric E Cohen (PricewaterhouseCoopers LLP, US)
- Nobuyuki Igata (Fujitsu Laboratories Limited, Japan)
- Andreas Harth (Karlsruher Institut für Technologie, Germany)
- Stavros Isaiadis (Bank Of America, UK)
- Roger Munday (Fujitsu Laboratories Limited Europe, UK)
- Sean O’Riain (Insight Galway, Ireland)
- Gianluca Garbellotto (IPHIX, US)
- Kecheng Liu (Henley Business School, Reading University, UK)
- Pierre-Yves Vandenbussche (Fujitsu Limited, Ireland)
- Keiichi Nakata (Henley Business School, Reading University, UK)
- Armin Haller (Australian National University, Australia)
- Maciej Zaremba (Citi Innovation Lab, Ireland)
- Laurentiu Vasiliu (Peracton Ltd., Ireland)
- Peter Haase (Fluid Operations, Germany)
- Oscar Corcho (Universidad Politécnica de Madrid, Spain)
- Saeed Roohani (Bryant University, US)

International Workshop on Engineering Cyber Security and Resilience

Workshop Organizing Committee

Conference Chair

Rudolph, Carsten, Fraunhofer SIT, Germany

Technical Program Chairs

Security Engineering:

Maña, Antonio, University of Malaga, Spain

Risk Modelling:

Ross, Ron, NIST, US

Technical Solutions:

Manz, David, PNNL, US

Publicity Chair

Endicott-Popovsky, Barbara, University of Washington

Technical Program Co-Chair

Ruiz, Jose Francisco, University of Malaga, Spain

Kuntze, Nicolai, Fraunhofer SIT, Germany

Technology and Industrial Relations Chair

McEvelley, Michael, The MITRE Corporation, US

Leivesley, Sally, Newrisk Limited, UK

Program Committee Members

- Ardagna, Claudio, Università degli Studi di Milano, Italy
- Arjona, Marcos, University of Malaga, Spain
- Badger, Lee, NIST, US
- Blundo, Carlo, University of Salerno, Italy
- Breuer, Peter, University of Birmingham, UK
- Calloni, Ben, Lockheed Martin, US
- Cuellar, Jorge, Siemens AG, Germany
- Di Pietro, Roberto, Università di Roma, Italy
- Fernandez, Eduardo, Florida Atlantic University, USA
- Frincke, Deborah, NSA/CSS, US
- García Rosado, David, Universidad de Castilla la Mancha, Spain
- Gürgens, Sigrid, Fraunhofer SIT, Germany
- Jones, Kevin, City University London, U. K.
- Hamid, Brahim, University of Toulouse, France
- Hild, Daryl, The MITRE Corporation, US
- Holtzman, Danny, The MITRE Corporation, US, US
- Karnouskos, Stamatis, SAP, Germany
- Katsikas, Sokratis, University of Piraeus, Greece
- Kemmerich, Thomas, Norwegian Information Security laboratory, Norway
- Lambrinoudakis, Costas, University of Piraeus, Greece
- Laurent, Maryline, Institut Mines-Télécom, Télécom Sud Paris, France
- Martinelli, Fabio, CNR-IIT, Italy
- Martinez Perez, Gregorio, University of Murcia, Spain
- Mouratidis, Haris, University of East London, UK
- Naccache, David, ENS, France
- Oren, Janet C., NSA, US
- Pearson, Siani, HP Labs, UK
- Quisquater, Jean-Jacques, Université Catholique De Louvain, Belgium
- Rass, Stefan, Alpen-Adria Universität Klagenfurt, Austria
- Savas, Erkay, Sabanci University, Turkey
- Stouffer, Keith, NIST, US
- Turner, Glenda, The MITRE Corporation, US, US
- Voas, Jeff, NIST, US
- Zisman, Andrea, City University of London, UK

May 27, 2014 (Tuesday)		
7:00-17:00	Registration: Tresidder Memorial Union, Stanford University Address: 459 Lagunita Drive, Stanford, CA, 94305	
8:30-17:30	International Workshop on Enabling Science From Big Image Data (Room: Oak Lounge East)	International Workshop on Exploiting Big Data in Commerce and Finance (Room: Cypress Room North)
8:30-17:30	International Workshop on Big Data Analytics for Predictive Organization and Big Transformations International Workshop on Distributed Storage Systems and Coding for Big Data (Room: Oak Lounge West)	International Workshop on Social Computing and Urban Intelligence and Smarter Lives International Workshop on Cloud Security (Room: Cypress Room South)
May 28, 2014 (Wednesday)		
7:30 -17:00	Registration	
8:00-8:20	Open Remarks	
	Morning Session Chair: Dr. Chaitanya Baru San Diego Supercomputer Center	
8:20 -8:50	Speech Title: Future of Data Intensive Applications Dr. Milind Bhandarkar Chief Scientist Pivotal, Inc.	
8:50-9:20	Speech Title: System U: Computational Discovery of Personality Traits from Social Media to Deliver Hyper-Personalized Experience Speaker: Dr. Michelle Zhou Senior Research Manager IBM Research–Almaden	
9:20-9:50	Speech Title: A Building Code for Building Code Dr. Carl Landwehr Lead Research Scientist George Washington University	
9:50 -10:10	Speech Title: Insight into the CIO World of Cybercrime and Cloud Computing Speaker: Chris Scotte Vice President Stealth Software Inc.	
10:10-10:30	Speech Title: Big Data in the Finder and Aladdin Video Programs Dr. Jill D. Crisman Program Manager IARPA	
10:30-10:45	Coffee Break	
10:45-11:05	Speech Title: Advancing Big Data Analytics and Data Science Through Measurements, Evaluations and Challenge Problems Dr. Ashit Talukder Chief, Information Access Division National Institute of Standards and Technology	
11:05-11:30	Speech Title: Generate Comprehensive Threat Reports and Security Scores for Android Apps Speaker: Dr. Wei Yan Chief Executive Officer VisualThreat Inc.	
11:30-12:00	Speech Title: Why Data Science? Dr. Stanley C. Ahalt,	

	<p>Director, Renaissance Computing Institute (RENCI) Professor of Computer Science University of North Carolina at Chapel Hill</p>	
12:00-12:30	<p>Speech Title: Healthcare and Data Science Vinod Khosla Partner of Khosla Ventures</p>	
12:30-13:30	<p>Lunch Break</p>	
	<p>Afternoon Session Room: Oak Lounge West Chair: Dr. Thomas Hinke NASA Advanced Supercomputing Division</p>	<p>Afternoon Session Room: Oak Lounge East Chair: Dr. Carl Landwehr George Washington University</p>
13:30-13:50	<p>Calculating Edit Distance for Large Sets of String Pairs using MapReduce [PDF] Shagun Jhaver (University of Texas at Dallas), Latifur Khan (University of Texas at Dallas), Bhavani Thuraisingham (University of Texas)</p>	<p>Topic and Sentiment Analysis on OSNs: a Case Study of Advertising Strategies on Twitter [PDF] Shana Dacres (QML), Hamed Haddadi (QML), Matthew Purver (QML)</p>
13:50-14:10	<p>New Insights into Individual Activity Spaces using Crowd-Sourced Big Data [PDF] Nick Malleson (University of Leeds), Mark Birkin (University of Leeds)</p>	<p>Detecting Adverse Drug Reactions using a Sentiment Classification Framework [PDF] Hashim Sharif (LUMS), Ahmed Abbasi (University of Virginia), Fareed Zaffar (LUMS), David Zimbra (Santa Clara University)</p>
14:10-14:30	<p>Exploring Netflow Data using Hadoop [PDF] Xiaofeng Zhou (University of Florida), Milenko Petrovic (Florida Institute for Human and Machine Cognition), Tom Eskridge (Florida Institute for Human and Machine Cognition), Marco Carvalho (Florida Institute for Human and Machine Cognition), Xi Tao (University of Florida)</p>	<p>Competence Modeling in Twitter: Mapping Theory to Practice [PDF] Byungkyu Kang (UC Santa Barbara), John O'Donovan (UC Santa Barbara), Tobias Hollerer (UC Santa Barbara)</p>
14:30-14:50	<p>Securing Virtual Service Generation on the Network: Adapting Digital Rights Management to Cloud-Delivered Media [PDF] Alexandra Mikityuk (TU Berlin), Oliver Friedrich (Telekom Innovation Laboratories)</p>	<p>Demographic breakdown of Twitter users: An analysis based on names [PDF] Huseyin Oktay (UMass), Aykut Firat (Crimson Hexagon), Zeynep Ertem (Texas A&M)</p>

14:50-15:10	<p>Mining a large-scale EHR with machine learning methods to predict all-cause 30-day unplanned readmissions [PDF]</p> <p>Haijun Zhai (Cincinnati Children's Hospital Medical Center), Srikant Iyer (Cincinnati Children's Hospital Medical Center), Yizhao Ni (Cincinnati Children's Hospital Medical Center), Todd Lingren (Cincinnati Children's Hospital Medical Center), Eric Kirkendall (Cincinnati Children's Hospital Medical Center), Huaxiu Tang (Cincinnati Children's Hospital Medical Center), Qi Li (Cincinnati Children's Hospital Medical Center), Imre Solti (Cincinnati Children's Hospital Medical Center)</p>	<p>Predicting New Collaborations in Academic Citation Networks of IEEE and ACM Conferences [PDF]</p> <p>Irfan A. Shah (National University of Sciences and Technology, Islamabad), Muhammad Ilyas (National University of Sciences and Technology, Islamabad), Mamoon Raja (National University of Sciences and Technology, Islamabad), Saad Saleh (SEECS, NUST, Islamabad), Muhammad Khan (National University of Sciences and Technology, Islamabad), Ali Qamar (National University of Sciences and Technology, Islamabad), Muhammad Shafiq (College of Engineering, Michigan State University, East Lansing, MI), Alex Liu (College of Engineering, Michigan State University, East Lansing), Hayder Radha (College of Engineering, Michigan State University, East Lansing, MI)</p>
15:10-15:30	<p>An Analysis of the Virtual Machine Migration Incurred Security Problems in the Cloud [PDF]</p> <p>Beulah Navamani (University of Colorado Colorado Springs), Chuan Yue (University of Colorado Colorado Springs), Xiaobo Zhou (University of Colorado Colorado Springs), Edward Chow (University of Colorado Colorado Springs)</p>	<p>Real Time Closeness and Betweenness Centrality Calculations on Streaming Network Data [PDF]</p> <p>Wei Wei (Carnegie Mellon University), Kathleen Carley (Carnegie Mellon University)</p>
15:30-15:50	<p>Automatic Labeling for Entity Extraction in Cyber Security [PDF]</p> <p>Robert Bridges (Oak Ridge National Laboratory), Corinne Jones (Oak Ridge National Laboratory), Michael Iannacone (Oak Ridge National Laboratory), Kelly Testa (Oak Ridge National Laboratory), John Goodall (Oak Ridge National Laboratory)</p>	<p>What Drives the Growth of YouTube? Measuring and Analyzing the Evolution Dynamics of YouTube Video Uploads [PDF]</p> <p>Golshan Golnari (University of Minnesota), Yanhua Li (HUAWEI Noah's Ark Lab), Zhi-Li Zhang (University of Minnesota)</p>
15:50-16:10	<p>An Engineering Process to Address Security Challenges in Cloud Computing [PDF]</p> <p>Marcos Arjona (University of Malaga), Rajesh Harjani (University of Malaga), Anto Muñoz (University of Malaga), Antonio Maña (University of Malaga)</p>	<p>Playful Crowdsourcing of Archival Metadata through Social Networks [PDF]</p> <p>Dimitris Paraschakis (Malmö University), Marie Gustafsson Friberger (Malmö University)</p>

16:10-16:30	Seeding Trusts and Tolerance in Cyber Communities [PDF] Simon Reay Akinson (University of Sydney), Seyedamir Seyedamir Tavakoli Taba (University of Sydney), David Walker (University of Sydney), Liaquat Hossain (The University of Hong Kong)	Using Determinantal Point Processes for Clustering with Application to Automatically Generating and Drawing xkcd Movie Narrative Charts [PDF] Apoorv Agarwal(Columbia University), Sarthak Dash(Columbia University), Sriramkumar Balasubramanian(Columbia University), Jiehan Zheng(Columbia University)
16:30-18:30	Poster Session P1	Invited Session: Security in Big Data Room: Oak Lounge East Chair: Dr. TY Lin Panelist: Dr. Tom Hinke, Dr. Kai Hwang, Dr. Car Landwehr, Mr. Eric Leighninger , Dr. TY Lin, Dr. Sylvia Osborn, Dr. Bhavani Thuraisingham, Dr. Felix Wu
May 29, 2014 (Thursday)		
7:30-17:00	Registration	
	Morning Session Chair Dr. Stanley Ahalt Renaissance Computing Institute (RENCI) University of North Carolina at Chapel Hill	
8:20-9:50	Speech Title: Enabling Cloud Analytics for Big-Data Security and Intelligence Dr. Kai Hwang Professor University of Southern California	
8:50-9:20	Speech Title: Delivering on the Promise of Big Data Arvind Parthasarathi President YarcData Cray, Inc.	
9:20-9:50	Speech Title: The Big Deal About Big Data: a Sociobiological Perspective Speaker: Rebecca D. Costa American Sociobiologist and Author	
9:50-10:30	Panel: The Cautionary Side of Big Data Moderator: American Sociobiologist and Author, Rebecca D. Costa Panelists: Steve Beier, IBM Big Data Program Director Guy Kawasaki, author, entrepreneur and business advisor Forrest Melton, Senior Research Scientist NASA Ames Research Center, Cooperative for Research in Earth Science & Technology	
10:30-10:45	Coffee Break	
10:45-11:05	Speech Title: Scaling R to Big Data Science Dr. Mario Inchiosa Revolution Abnalytics, Inc.	

11:05-11:25	Speech Title: Big Data and The Customer Decision Journey Matt Hertig Co-Founder Alight Analytics	
11:25-11:45	Speech Title: Simplified Data Parsing and Ingestion with DFDL Stephen Lawrence Software Engineer Tresys Technology, Inc.	
11:45-12:05	Speech Title: Exploiting Parallelism and Scalability for Big Data in Medicine Dr. Akash Singh CEO, Neuron Networks	
11:05-12:30	Speech Title: Big Data and Medicine: from Genome to Populations Speaker: Dr. Russ Altman Kenneth Fong Professor of Bioengineering, Genetics, Medicine and Computer Science Stanford University	
12:30-13:30	Lunch Break	
	Afternoon Session Room: Oak Lounge West Chair: Michael Shoffner Renaissance Computing Institute (RENCI) University of North Carolina at Chapel Hill	Afternoon Session Room: Oak Lounge East Chair: Dr. Akash Singh Neuron Networks
13:30-13:50	Mining Urban Performance: Scale-Independent Classification of Cities Based on Individual Economic Transactions [PDF] S. Sobolevsky (SENSEable city lab, MIT), Izabela Sitko (Department of Geoinformatics Z\GIS, University of Salzburg), Sebastian Grauwin (SENSEable city lab, MIT), Remi Tachet des Combes (SENSEable city lab, MIT), Bartosz Hawelka (Department of Geoinformatics Z\GIS, University of Salzburg), Juan Murillo Arias (New Technologies, BBVA), Carlo Ratti (SENSEable city lab, MIT)	To Call, or To Tweet? Understanding 3-1-1 Citizen Complaint Behaviors [PDF] Vanessa Frias-Martinez (University of Maryland), Abson Sae-tang (Ecole Polytechnique Federal de Lausanne EPFL), Enrique Frias-Martinez (Telefonica Research)
13:50-14:10	A Layered Locality Sensitive Hashing based Sequence Similarity Search Algorithm for Web Sessions [PDF] Angana Chakraborty (Indian Statistical Institute), Sanghamitra Bandyopadhyay (Indian Statistical Institute)	Framing Strategies in Facebook Discussion Groups [PDF] Keith C. Wang (University of California, Davis), Huan-Kai Peng (Carnegie Mellon University), S. Felix Wu (University of California, Davis)
14:10-14:30	Demand-Driven Asset Reutilization Analytics Abbas Raza Ali (IBM) [PDF]	To Post or Not to Post: The Effects of Persuasive Cues and Group Targeting Mechanisms on Posting Behavior [PDF] Bruce Ferwerda (Johannes Kepler University), Markus Schedl (Johannes Kepler University), Marko Tkalcic (Johannes Kepler University)

14:30-14:50	Scalable Distributed Virtual Data Structures [PDF] Sushil Jajodia (George Mason University Fairfax, Virginia, USA), Witold Litwin (Dauphine University Paris, France), Thomas Schwarz (Universidad Católica del Uruguay), SJ(Uruguay University)	Will this Celebrity Tweet Go Viral? An Investigation of Retweets [PDF] Lisa Gandy (Central Michigan University), Libby Hemphill (Illinois Institute of Technology)
14:50-15:10	Anomaly Detection in Real-Valued Multidimensional Time Series [PDF] Michael Jones (MERL), Daniel Nikovski (MERL), Makoto Imamura (Mitsubishi Electric), Takahisa Hirata (Mitsubishi Electric)	Analysis and Modeling of Lowest Unique Bid Auctions [PDF] Tanmoy Chakraborty (Indian Institute of Technology - Kharagpur, India), Vihar Tammana (Microsoft Corporation, Redmond, USA), Niloy Ganguly (Indian Institute of Technology - Kharagpur, India), Animesh Mukherjee (Dept. of CSE, IIT Kharagpur)
15:10-15:30	A Look Inside the Nodes: Cohesive Attribute Micro-Clustering [PDF] Bradley Meyer (North Carolina A&T State University)	Classifying News Authors Based on their Authoritativeness [PDF] Riham Mansour (Microsoft Research), Tamer Deif (Microsoft research), Antoine Morcos (Microsoft Research)
15:30-15:50	Classification Based IP Geolocation Approach to Locate Data in the Cloud Datacenters [PDF] Biswajit Biswal, Sachin Shetty, Tamera Rogers(Tennessee State Univeristy)	Topic Mining from Heterogeneous Web Sources [PDF] Rumi Ghosh (Robert Bosch LLC), Sitaram Asur (HP)
15:50-16:10	Curbing the Spread of Topological Worms - A Defense Topology for the P2P Cyberspace [PDF] Narasimha Shashidhar (Sam Houston State University), Lei Chen (Sam Houston State University)	Patterns of Social Media Conversations Using Second Screens [PDF] Partha Mukherjee (Pennsylvania State University), Jian-Syuan Wong (Pennsylvania State University), Bernard Jansen (Pennsylvania State University)
16:10-16:30	Community Detection in Large Scale Big Data Networks [PDF] Pravin Chopade, Justin Zhan (North Carolina A&T State University)	Site-Constrained Privacy Options for Users in Social Networks through Stackelberg Games [PDF] Christopher Griffin (Pennsylvania State University), Anna Squicciarini (Pennsylvania State University), Sarah Rajtmajer (Pennsylvania State University), Matthew Tentilucci (Pennsylvania State University), Shen Li (Pennsylvania State University)
16:30-18:30	Poster Session P2	Exhibitor Demo Room: Oak Lounge East 17:00-17:30 VisualThreat Inc. 17:30-18:30 Bay Storage Technology

May 30, 2014 (Friday)		
7:30-17:00	Registration	
	Morning Session Chair: Howard Lander Renaissance Computing Institute (RENCI) University of North Carolina at Chapel Hill	
8:30-9:00	Speech Title: On Content, Discussions, Opinions, and Deliberative Participation over Social Media Systems Speaker: Dr. S. Felix Wu Professor University of California-Davis	
9:00-9:30	Speech Title: Cloud-Centric Assured Information Sharing Dr. Bhavani Thuraisingham Louis A. Beecherl, Jr. Distinguished Professor of Computer Science Executive Director of the Cyber Security Research and Education Institute The University of Texas at Dallas	
9:30-9:50	Speech Title: Twitter Analytics for Insider Trading Fraud Detection System [PDF] Dr. Ketty Gann, John Day, Shujia Zhou Northrop Grumman Information Systems	
9:50-10:10	Speech Title: Augmenting Text Analysis: Text Insights and Social Media [PDF] Dr. Seth Howell, Dr. Ajith Warriar, Dr. Damir Herman Senior Data Scientists Ayasdi, Inc.	
10:10-10:30	Speech Title: Big Data and Semantic Web Meet Applied Ontology Dr. Ram D. Sriram Chief, Software and Systems Division National Institute of Standards and Technology	
10:30-10:45	Coffee Break	
10:45-11:15	Speech Title: Beyond Map Reduce: The Next Generation of Big Data Analytics Speaker: Stephen Turner HAMR by ET International, Inc.	
11:15-12:15	Panel Discussion on “Big Data in the Social Sciences” Moderators: Dr. Tom Carsey and Dr. Ashok Krishnamurthy University of North Carolina at Chapel Hill Panelists: Justin Grimmer, Stanford University Jon Crabtree, University of North Carolina and Odum Institute Hye-Chung Kum, Texas A&M University Michael Shoffner, University of North Carolina and RENCi	
12:15-13:30	Lunch Break	
	Afternoon Session Room: Oak Lounge West Chair: Jim Peek Bay Storage Technology	Afternoon Session Room: Oak Lounge East Chair: Dr. Felix Wu University of California-Davis
13:30 -13:50	An Argumentation-Based Framework to Address the Attribution Problem in Cyber-Warfare [PDF] Paulo Shakarian (U.S. Military Academy),	Modeling Interactions in Web Forums [PDF] Weifeng Li (University of Arizona), Ahmed Abbasi (University of Virginia), Shiyu Hu (University of Arizona), Victor Benjamin (University of Arizona),

	Gerardo Simari (Oxford Univ.), Geoffrey Moores (U.S. Military Academy), Simon Parsons (Univ. Liverpool), Marcelo Falappa (Universidad Nacional del Sur)	Hsinchun Chen (University of Arizona)
13:50-14:10	CyberPhySecLab: A testbed for modeling, detecting and responding to security attacks on Cyber Physical Systems [PDF] Sachin Shetty (Tennessee State University), Tayo Adedokun (Tennessee State University), L.H. Keel (Tennessee State University)	Political Social Networks Reveal Strong Party Loyalty in Brazil and Weak Regionalism [PDF] Marcos Oliveira (Florida Institute of Technology), Carmelo Bastos-Filho (University of Pernambuco), Ronaldo Menezes (Florida Institute of Technology)
14:10-14:30	Cyber-Security Via Computing With Words [PDF] John Rickard (Distributed Infinity, Inc.), Allen Ott (Distributed Infinity, Inc.)	A Combined Classification-Clustering Framework for Identifying Disruptive Events [PDF] Nasser Alsaedi (Cardiff University), Pete Burnap (Cardiff University), Omer Rana (Cardiff University)
14:30-14:50	Epidemic Node Replica Detection in Sensor Networks - Securing the Sensor Network Cyberspace against Replica Attacks [PDF] Narasimha Shashidhar (Sam Houston State University), Chadi Kari (Bridgewater State University), Rakesh Verma (University of Houston)	Data Matters: Reflection on User Defined Social Prioritization [PDF] Juwel Rana (Telenor Research), Kåre Synnes (Luleå University of Technology), Johan Kristiansson (Luleå University of Technology)
14:50-15:10	Biocryptographic Secure Socket Layer (BSSL) [PDF] Abdullah Albahdal (UCCS), Terrance Boulton (UCCS)	Leveraging Community Detection for Accurate Trust Prediction [PDF] Ghazaleh Beigi (Sharif University of Technology), Mahdi Jalili (Sharif University of Technology), Hamidreza Alvari (University of Central Florida), Gita Sukthankar (Department of EECS, University of Central Florida)
15:10-15:30	Spectral Approach for Information Diffusion [PDF] Kiyotaka Ide (National Defense Academy), Akira Namatame (National Defense Academy of Japan), Loganathan Ponnambalam (Computing Sciences, Institute of High Performance Computing, A*STAR, Singapore), Fu Xiuju (Computing Sciences, Institute of High Performance Computing, A*STAR, Singapore), Rick Goh Siow Mong (Computing Sciences, Institute of High Performance Computing, A*STAR, Singapore)	Opinion dynamics in correlated time-varying social networks [PDF] Suman Kalyan Maity (Dept. of CSE, IIT Kharagpur), T. Venkata Manoj (Dept. of CSE, IIT Kharagpur), Animesh Mukherjee (Dept. of CSE, IIT Kharagpur)
15:30-15:50	A Policy Consumption Architecture that enables Dynamic and Fine Policy Management [PDF] Hiroyuki Sato (University of Tokyo), Shigeaki Tanimoto (Chiba Institute of Technology), Atsushi Kanai (Hosei University)	A Visibility-based Model for Link Prediction in Social Media [PDF] Linhong Zhu (Information Sciences Institute), Kristina Lerman (Information Sciences Institute)

15:50-16:10	Learning Temporal User Profiles of Web Browsing Behavior [PDF] Myriam Abramson (Naval Research Laboratory)	EMOTEX: Detecting Emotions in Twitter Messages [PDF] Maryam Hasan (WPI), Elke Rundensteiner (WPI), Emmanuel Agu (WPI)
16:10-16:30	A Simple and More Versatile Authentication Path Computation Algorithm for Binary Hash Tree [PDF] Shuanghe Peng (Beijing Jiaotong University), Yingjie Qin (Beijing Jiaotong University), Zhige Chen (Beijing Jiaotong University)	SecP2PSIP: A Distributed Overlay Architecture for Secure P2PSIP [PDF] Rasib Khan (University of Alabama at Birmingham), Ragib Hasan (University of Alabama at Birmingham)
16:30-18:30	Poster Session P3	Exhibitor Demo Room: Oak Lounge East
May 31, 2014 (Saturday)		
7:30-17:00	Registration Morning Session Room: Oak Lounge West Chair: Dr. Mario Inchiosa Revolution Analytics, Inc.	
8:15-8:45	Speech Title: Randomized Matrix Algorithms and Large-scale Scientific Data Analysis Speaker: Dr. Michael Mahoney ICSI and UC Berkeley	
8:45-9:05	Discovering Community Structure in Dynamic Social Networks using the Correlation Density Rank [PDF] Zeynab Bahrami Bidoni (Clark Atlanta University), Roy George (Clark Atlanta University)	
9:05-9:25	Using Interactions in the Quantification of Media Bias [PDF] Diogo Pacheco (Florida Institute of Technology), Dillon Rose (Florida Institute of Technology), Fernando Lima-Neto (University of Pernambuco), Ronaldo Menezes (Florida Institute of Technology)	
9:25-9:45	Information Relaxation is Ultradiffusive [PDF] Rumi Ghosh (Robert Bosch LLC), Bernardo Huberman (HP)	
9:45-10:05	ASH: Scalable Mining of Collective Behaviors in Social Media using Riemannian Geometry [PDF] Huan-Kai Peng (Carnegie Mellon University), Radu Marculescu (Carnegie Mellon University)	

10:05-10:30	More Efficient Tagging Systems with Tag Seeding [PDF] Vikas Kumar (University of Minnesota), Daniel Kluver (University of Minnesota, Twin Cities), Loren Terveen (University of Minnesota, Twin Cities), John Riedl (University of Minnesota, Twin Cities)
10:30-10:45	Coffee Break
10:45-11:05	Feature Selection in Massive Game Log Analysis Using K-L Divergence [PDF] Y. Dora Cai (Univ. of Illinois), Bettina Riedl (LMU), M. Scott Poole (Univ. of Illinois)
11:05-12:00	Tutorial Title: Deep Learning for Natural Language Processing Presenter: Richard Socher Department of Computer Science Stanford University
12:00-13:00	Lunch Break
	Afternoon Session Room: Oak Lounge West Chair: Dr. Wei Yan VisualThreat
13:00-14:00	Tutorial Title: Tools for Scaling Genomics Analysis Dr. Uri Laserson Data Scientist Cloudera, Inc.
14:00-15:00	Tutorial Title: Toward Integrated Security in Cloud Based Data and Compute Infrastructures Michael Shoffner Research Scientist Renaissance Computing Institute University of North Carolina at Chapel Hill
15:00-16:30	Tutorial Title: Behavioral Data Mining and Social Network Analysis in Massive Online Games Dr. Muhammad Aurangzeb Ahmad and Jaideep Srivastava Research Scientist and Professor Department of Computer Science University of Minnesota
16:30-18:00	Tutorial Title: Revolution R Enterprise Dr. Mario Inchiosa and Joseph Rickert US Chief Scientists Revolution Analytics, Inc.
9:00-17:00	International Workshop on Engineering Cyber Security and Resilience (Room: Oak Lounge East)

WORKSHOP PROGRAM

International Workshop on Big Data Analytics for Predictive Organization and Big Transformations

&

International Workshop on Distributed Storage Systems and Coding for Big Data


Registration and Workshop Venue:


Room: Oak Lounge West

Tresidder Memorial Union, Stanford University

459 Lagunita Drive, Stanford, CA, 94305


Workshop Agenda:

Tuesday, May 27, 2014		
8:00 am	Registration	
9:00 am to 9:45am	<p>Keynote Speech Title: From JSON Logs to Latent Variable Models: Knowledge Mining Massive Open Online Courses Data</p> <p>Abstract: Millions of students around the world are taking courses online offered by universities. Known as Massive Open Online Courses (MOOCs), these provide data pertaining to hundreds of thousands of students as they navigate through the web based platform, submit assignments and participate in forums. The data comes with a promise of enabling us to find patterns that could help MOOC instructors teach more effectively and improve student engagement. That data also presents us with a unique opportunity to learn about how students learn and help us improve on campus instruction as well. However, in a single course there are hundreds of millions of click stream events and thousands of discussion threads. The first course offered by MITx had</p>	 <p>Dr. Kalyan Veeramachaneni Research Scientist and Leader AnyScale Learning For All (ALFA) Group, Massachusetts Institute of Technology (MIT) Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA</p>

	<p>approximately 200 million click stream events, ~10 million assignment submissions and around 90000 forum posts. In this talk, I will present how we are harnessing this data and enabling MOOC data science at scale. I will focus on a number of predictive models we are building to be able to predict student stopout. On our way to solving this problem, we built a number of platforms and tools to enable data science at scale. I will also present a summary of those innovative platforms.</p>	
9:45 am to 10:30 pm	<p>Keynote Speech Title: Big Data Challenges in the Biosciences</p>	
	<p>Abstract: In just a little over a decade, the cost of sequencing a complex organism such as the human dwindled from the \$100 million range to sub \$1000 range. This rapid decline is brought about by the advent of a number of high-throughput sequencing technologies, collectively known as next generation sequencing. Their usage has become ubiquitous, enabling single investigators with limited budget to carry out what could only be accomplished by a network of major sequencing centers just a decade ago. This is leading to an explosive growth in the number of organisms sequenced, and in the number of individuals sequenced in search of important genetic variations. Next-gen sequencers enable diverse applications, each requiring its own class of supporting algorithms. This talk will highlight some of the bigdata challenges arising from these developments in the context of microbial communities, agricultural biotechnology, and human health.</p>	<p>Dr. Srinivas Aluru Professor, School of Computational Science and Engineering College of Computing, Georgia Institute of Technology, GA, USA</p>

10:30 am to 10:45 am Coffee break

10:45 am to 11:30 am	Invited Talk Title: Codes with Locality for Data Storage	 <p>Dr. Parikshit Gopalan, Researcher, Microsoft Research, Silicon Valley, USA</p>
	<p>Abstract: In erasure codes used for big data storage, it is important to ensure that error patterns which occur frequently (such as single disk failure) can be corrected efficiently. This has motivated the study of codes with good locality, where any data symbol can be reconstructed using a few other codeword symbols. In this talk, we will summarize what we know about rate distance tradeoffs for such codes. The main challenge in constructing such codes is to maximize their reliability while keeping the field size small. We will survey the few known result and numerous open problems in this area.</p> <p>* Joint work with Cheng Huang, Bob Jenkins, Huseyin Simitci and Sergey Yekhanin (all from Microsoft).</p>	
11:30 am to 12:00 noon	Invited Talk Title: Data Recovery with Minimum Bandwidth and Disk I/O in Distributed Storage Systems	 <p>Dr. Hui Li, Professor, School of Electronic & Computer Engineering, Peking University, Shenzhen, P. R. China</p>
	<p>Abstract: Recently, erasure codes have been increasingly used in distributed storage systems to reduce the large storage overhead incurred by the widely used replication scheme. However, these codes require significantly high network bandwidth and disk I/O usage during recovery of missing data. In this talk, we present a novel storage code inspired by the recently proposed BASIC framework to reduce recovery bandwidth and disk I/O. We build a system called FASTER based on the novel code and evaluate our</p>	

	<p>prototype atop a HDFS cluster test bed with 20 nodes. The recovery bandwidth achieves minimum during recovery of both data block and parity block with FASTER. Another attractive result is that the recovery disk I/O achieves minimum during recovery of data block. With the reduction of recovery bandwidth and disk I/O, the degraded read throughput is boosted notably.</p>	
12:00 noon to 12:30 pm	<p>Invited Talk Title: A Polynomial Construction of Sector-disk Code Abstract: There are two types of disk errors in storage systems, namely, entire disk failures and disk sector failures. In conventional storage codes, a disk sector failure is considered as an entire disk failure, even though most of the remaining sectors in the disk remain intact. Sector-Disk (SD) code was proposed by Plank et al. to recover a mixture of these two types of failure patterns. Gopalan et al. gave a construction of maximally recoverable codes, which also have the SD property for single disk failure plus any number of sector failures. Another family of SD codes constructed by Blaum et al. can tolerate any number of disk failures and two erased sectors. These two families of SD codes are constructed by specifying the parity-check matrix. In this talk, we present a construction of SD codes which specifies the generator matrix using bi-variate polynomials. Our new construction can repair any number of disk failures with up to three sector failures. *Joint work with Junyu Chen, M.E., CUHK.</p>	 <p>Dr. Kenneth Shum Associate Professor, Institute of Network Coding, The Chinese University of Hong Kong (CUHK), Hong Kong</p>
12:30 pm to 1:30 pm	Lunch break	
1:30 pm to 3:30 pm	Technical Sessions 1: On areas 1, 2 and 3 Session Chair: Dr. Markowsky, Co-Chair: Xing Fang In Praise of Small Data [PDF] Dr. George Markowsky (University of Maine)	

	<p>Agent Based Simulation Model for Data Visualization During Evacuation [PDF] Kola Ogunlana, Sharad Sharma (Bowie State University)</p>
	<p>Big Data Predictive Analytics for Proactive Semiconductor Equipment Maintenance: A Review [PDF] Sathyan Munirathinam(Micron Technology, Inc., Boise, USA) B. Ramadoss (National Institute of Technology, Trichy, India)</p>
	<p>Statistical Modeling and Scalable, Interactive Visualization of Large Scale Big Data Networks [PDF] Haysam Selim, Pravin Chopade, Justin Zhan (North Carolina A&T State University)</p>
3:30 pm to 4:00 pm Coffee break	
4:00 pm to 6.00 pm	<p>Technical Sessions 2: On areas 4, 5 and 6 Session Chair: Dr. George Markowsky, Co-Chair: Bradely Meyer</p>
	<p>An Efficient Algorithm for Mining Locally Frequent Itemsets [PDF] Fokrul Mazarbhuiya, Mohammed Husamuddin, Mohammed Shenify (Al-Baha University)</p>
	<p>Identification of Closest Safe Places and Exit Routes During Evacuation from GIS Data [PDF] Hamzeh Qabaja, Marwan Bikdash (Computational Science and Engineering, North Carolina A&T State University)</p>
	<p>Defining a New Paradigm for Data Protection in the World of Big Data Analytics [PDF] Alessandro Mantelero (Polytechnic University of Turin, Italy)</p>
	<p>The Bustle of Bioinformatics: Cloudy with a Chance for Big Data [PDF] Samuel Chapman, Scott Harrison, Marwan Bikdash, Dukka KC (North Carolina A&T State University)</p>

International Workshop on Enabling Science from Big Image Data

Registration and Workshop Venue:

Room: Oak Lounge East

Tresidder Memorial Union, Stanford University

459 Lagunita Drive, Stanford, CA, 94305

Workshop Agenda:

Morning Sessions

Time	Presentation
9:00 am	Opening remarks: Mary Brady & Peter Bajcsy, NIST
9:10 am	Peter Bajcsy, NIST To Measure or Not To Measure Terabyte-Sized Images [PDF]
9:30 am	Keynote speaker: Hanchuan Peng, Allen Institute for Brain Science Seeing More, Exploring More, and Knowing More: UltraFast 3D Visualization of Big Imaging Data and Human-Machine Interaction for High-Throughput Biology [PDF]
10:00 -10:15am	Break
10:15 am	Panel on Enabling Science from Big Image Data using Cross Cutting Infrastructure Chair: Peter Bajcsy David Gutman, William Dunn, Lee A D Cooper, Emory University The Cancer Digital Slide Archive: A Web Based Resource Linking

	<p>Pathology, Radiology, and Genomics [PDF]</p> <p>Tomasz Bednarz, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia</p> <p>Cloud-based image analysis, visualization and big data analytics [PDF]</p> <p>Amit Chourasia, San Diego Supercomputer Center (SDSC)</p> <p>Ubiquitous sharing of results via SeedMe platform [PDF]</p> <p>Joaquin Correa, Lawrence Berkeley National Laboratory</p> <p>Big Data Bioimaging: Advances in Analysis, Integration, and Dissemination [PDF]</p>
--	--

11:50 –1:00 pm Lunch break

Afternoon Sessions

Time	Presentation
1:00 pm	<p>Panel on Enabling Science from Big Image Data in Cell Microscopy and Medical Imaging</p> <p>Chair: Hanchuan Peng</p> <p>Amit Paul, Michael Cho, UIC</p> <p>Biometric Multi-Content Analytics to Predict Stem Cell Fate [PDF]</p> <p>Manfred Auer, Lawrence Berkeley National Lab</p> <p>Terabyte-Scale Understanding of Disease Mechanisms and Microbial Ecology: High-Resolution Ultrastructural 3D Imaging of Cells, Tissues and Biofilms</p> <p>Maric Dragan, NIH</p> <p>Deciphering the complex systems biology of tissue remodeling after ischemic brain injury using multiplex fluorescence immunohistology and big image data</p>

	quantitative analysis [PDF]
2:15 – 2:30pm	Break
2:30 pm	<p>Panel on Enabling Science from Big Image Data in Transportation, Civil Engineering, and Astronomy</p> <p>Chair: Mary Brady</p> <p>Pujitha Gunaratne, Toyota Technical Center, Ann Arbor, MI</p> <p>Mohan Trivedi, University of California at San Diego (UCSD)</p> <p>Challenges in Big Driving Data Analytics [PDF]</p> <p>David Lattanzi, George Mason University</p> <p>Robotic Bridge Inspection: A Needle in a Haystack Problem</p> <p>Phil Marshall, Stanford University</p> <p>Big astronomical image data</p>
3:45 – 4:00pm	Break
4:00 pm	<p>Panel on Enabling Science from Big Image Data in Archival and Genomics</p> <p>Chair: Peter Bajcsy</p> <p>Kenton McHenry, National Center for Supercomputing Applications (NCSA), UIUC</p> <p>Applied Computer Vision and Big Compute Towards Indexing Digitized Collections [PDF]</p> <p>Charles Law, Kitware</p> <p>Archiving Image Data for Big Science: Applications from Computer Vision to Neuroscience</p>

	<p>Marc Salit, NIST/Stanford (The Advances in Biomedical Measurement Science (ABMS) program)</p> <p>Big challenges in genome-scale measurement assurance</p>
5:15 pm	<p>Closing remarks</p> <p>Peter Bajcsy, NIST</p>
5:30 pm	<p>Poster Sessions</p> <p>1. Polymer Discovery [PDF]</p> <p>Thomas Goddard, Konstantin Litovskiy, Nathan Nichols-Roy, Matthew Reed, Igor Shvartser, Nicholas Smith, David Zeppa, Julia Rice, Hans Horn, Amanda Engler, Linda Werner Baskin School of Engineering, UC Santa Cruz & IBM Almaden Research Center, San Jose, CA</p> <p>2. The Cancer Digital Slide Archive: A Web-based Resource Linking Pathology, Radiology, and Genomics [PDF]</p> <p>William D Dunn Jr., Lee A.D. Cooper, David A Gutman,</p> <p>Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA</p> <p>3. Biometric Multi-Content Analytics to Predict Stem Cell Fate [PDF]</p> <p>Amit Paul and Michael Cho,</p> <p>Department of Bioengineering, University of Illinois, Chicago, IL, USA</p> <p>4. Cloud-Based Image Analysis and Processing Toolbox [PDF]</p> <p>Tomasz Bednarz, Yulia Arzhaeva, Piotr Szul, Alex Khassapov, Neil Burdett, Dadong Wang, Shiping Chen, Darren Thompson, Ryan Lagerstrom, Changming Sun, Pascal ValloIon, Luke Domanski, Timur Gureyev, John Taylor,</p> <p>The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia</p>

5. Applied Computer Vision and Big Compute Towards Indexing Digitized Collections [\[PDF\]](#)

Liana Diesendruck, Rob Kooper, Luigi Marini, Kenton McHenry

National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

6. Big Data Bioimaging: Advances in Analysis, Integration, and Dissemination [\[PDF\]](#)

Joaquin Correa, David Skinner, Manfred Auer

Lawrence Berkeley National Laboratory

National Energy Research Scientific Computing Center

7. To Measure or Not To Measure Terabyte-Sized Images [\[PDF\]](#)

Peter Bajcsy, Antoine Vandecreme, Jing Gao, Joe Chalfoun, Julien Amelot, Mike Majurski, Mary Brady, Kiran Bhadriraju, Michael

Halter, John Elliott, Anne Plant, Nicholas W. M.Ritchie, John Henry J. Scott, Dale E. Newbury (NIST)

8. Background Intensity Correction for Terabyte-Sized Time-Lapse Images [\[PDF\]](#)

Joe Chalfoun, Mike Majurski, Kiran Bhadriraju , Steve Lund, Peter Bajcsy , Mary Brady

Informational Technology Laboratory and Material Measurement Laboratory, NIST

9. Analysis and Visualization of Terabyte Image Data [\[PDF\]](#)

Charles Law, Dhanannjay Deo

Kitware, Clifton Park, NY, USA



International Workshop on Social Computing for Urban Intelligence and Smarter Lives (SMARTLIFE 2014)

&

International Workshop on Cloud Computing (CLOUDCOM 2014)

Registration and Workshop Venue:

Room: Cypress Room South

Tresidder Memorial Union, Stanford University

459 Lagunita Drive, Stanford, CA, 94305

Workshop Agenda:

09:00 – 09:30	KEYNOTE – Prof. Dr. Camille Crittenden, Deputy Director CITRIS, University of California, Berkeley
09:30 – 10:30	TUTORIAL – "On Social Apps and Intelligent Mobile Services", Fernando Koch, SAMSUNG Research Institute, Brazil
10:30 – 10:45	Break
10:45 – 10:50	SMARTLIFE 2014 Opening Session
10:50 – 11:05	"Geo-fencing based Disaster Management Service" [PDF] Piotr Szczytowski, Technisch Universitat Darmstadt, Germany
11:05 – 11:20	"Caring for My Neighborhood: a platform for public oversight" [PDF] Gisele Craveiro and Andrés Martano, University of Sao Paulo (USP), Brazil
11:20 – 11:35	"Inferring Social Intelligence from Open Communication Interfaces" [PDF] Kelly Shigeno et al, IBM Research – Brazil
11:35 – 11:50	"Towards a Ubiquitous Service-Oriented Architecture for Urban Sensing" [PDF] Carlos Rolim et al – Federal University of Rio Grande do Sul (UFRGS), Brazil
11.50 – 12:05	"The Sensing Society: New Modulations and Empowerments" John Fitzgerald, Jens Zinn, The University of Melbourne, Australia
12:05 – 13:00	Lunch Break

13:00-13:30	CLOUDCOM 2014 Opening Session: Antonio Muñoz
13:30-14:00	Security Meets LDAP Multi-Master Replication [PDF] :Thomas Bauereiss, Dieter Hutter, Stefan Gohmann and Alexander Klaeser(German Research Center for Artificial Intelligence)
14:00-14:30	Certified service-based coalitions support for development of secure cloud applications [PDF]: Javier Espinar, Marcos Arjona and Antonio Maña(University of Malaga)
14:30-15:00	CloViS : a Secured Cloud Virtualized Storage [PDF]: Olivier Lafontaine, Jacques Jorda and Abdelaziz Mzoughi
15:00-15:30	Towards a Certified-by-design approach for Cloud Services Development [PDF]: Rajesh Harjani and Antonio Maña(University of Malaga)
15:30-17:30	Round table discussion

The City and the Citizen: A New Urban Renaissance

Keynote, SMARTLIFE Workshop, May 27, 2014

Prof. Dr. Camille Crittenden
Deputy Director CITRIS, University of California, Berkeley

Abstract: Increasing trends in urbanization around the world have given rise to the heuristic of the “city” in academic, philanthropic, and commercial domains. University programs have proliferated in urban or metropolitan studies, city and regional planning, research in the built environment, and the like. The Rockefeller Foundation is investing in 100 “Resilient Cities,” and the “Future Cities Lab” in San Francisco and others develop design projects in urban settings. “Smart cities” is the title of a recent book and the topic of countless articles and conferences. Corporate programs at IBM, Cisco and elsewhere also use the concept of “smart cities,” while governments around the world are pouring money into public-private partnerships to develop data-gathering and analytics applications that will benefit transportation, energy, and the environment.

The vibrant interest in cities and urban technology seems fresh and innovative, but the framework of the city as crucible of creativity, democracy, and commerce harks back to the Renaissance. Advances in arts and sciences, exploration and mapping, democratic governance and independent business models flourished then and now, many of which were facilitated by new technology. As Renaissance humanists recognized—and as this workshop title suggests—technological advances must be matched by the human, *social* component in order to realize their full benefit.

This talk will identify four scales or range of parameters to consider in the current technology-enabled urban context, then examine a few specific use cases and their effect along these scales. How does technology and social computing mediate or exacerbate the poles of privacy and security, individual vs. collective interests, grassroots vs. hierarchical organization, and concerns of poverty and wealth

International Workshop on Exploiting Big Data in Commerce and Finance

Registration and Workshop Venue:

Room: Cypress Room North
Tresidder Memorial Union, Stanford University
459 Lagunita Drive, Stanford, CA, 94305

Workshop Agenda:

Tuesday 27 th May 2014 (0830 – 1600)		
0830 – 0900	Welcome Aisha Naseer (Fujitsu Laboratories of Europe Limited)	
0900 – 0930	Keynote Speech: Big and Open Data Technologies for Financial Reporting	Masatomo Goto (Fujitsu Labs of Europe Limited)
0930 – 1010	Panel Speech 1: Tame, Mash and Merge: Using Semantics to Tame Big Data, Mash with Open Data and Merge with Private Data for Sustainable Decision Making	Gianluca Garbellotto (IPHIX) Eric Cohen (PricewaterhouseCoopers LLP)
1015 – 1035	Panel Speech 2: Distributed Processing of Large Scale Financial ‘big-data’ from Desperate Sources using an Event Bus Architecture	Fuad Rahman, Ari Mitra (APURBA Technologies)
1040 – 1100	Panel Speech 3: Cryptography for Big Data	Arnab Roy (Fujitsu Labs of America)
1100 – 1200	Discussion and Q/A session	
1200 – 1300	Lunch Break	
1300 – 1330	Paper Session 1: LOD metadata harvester: enabling better decision making [PDF]	Vivian Lee, Aisha Naseer, Terunobu Kume (Fujitsu Laboratories of Europe Limited, Fujitsu Laboratories Limited)
1330 – 1345	Break	
1345 – 1415	Paper Session 2: A novel crowd-sourcing technique for big data harvesting on social media [PDF]	Georgios Xydopoulos, Hiran Basnayake, Panos Louvieris, Lampros Stergioulas (Brunel University)
1415 – 1430	Break	
1430 – 1500	Poster Session: Framework for Identifying like-minded Communities for Measuring Satisfaction Utilizing Big Data Analytics	Ajay Kumar, Ravi Shankar, Roma Mitra Debnath, Jaideep Prabhu (Indian Institute of Technology Delhi Indian Institute of Public Administration, University of Cambridge)
1500	Workshop Closing Remarks	

<p>ECSaR 2014</p> <p>31 May 2014</p>	<p>International Workshop on Engineering Cyber Security and Resilience</p> <p>Registration and Workshop Venue: Room: Cypress Room North Tresidder Memorial Union, Stanford University 459 Lagunita Drive, Stanford, CA, 94305</p>
<p>Workshop Agenda:</p>	
09:00	<p>Welcome (Carsten Rudolph, Fraunhofer SIT)</p>
09:15	<p>Keynote Title – Industrial Control System (ICS) Cyber Security – State of the State Joe Weiss, Realtime ACS</p> <p>ICSs monitor and control critical industrial infrastructures including electric power and water. ICSs consist of human-machine interfaces (HMIs) which are generally Windows and field devices which utilize proprietary real time operating systems and have minimal if any security. There have already been more than 325 actual ICS cyber incidents with impacts including major electric outages, equipment failures, and even deaths. This presentation will provide a state-of-the state on ICS cyber security including what makes ICSs different than IT, recent case histories, and selected industry efforts. Case histories will address what is publicly known, similarities and differences with selected ICS cyber incidents, and identify types of cyber forensics and technologies needed for ICS cyber security.</p>
10:00	<p>Session 1</p> <p>Enhancing Intrusion Detection in Substation Networks [PDF] Hagen Lauer and Otis Alexander, Nicolai Kuntze, Michael Jager</p> <p>Trusted Computing Concepts for Resilient Embedded Networks [PDF] Martin Schramm and Andreas Grzempa</p>
11:00	<p>Coffee Break</p>
11:15	<p>Panel discussion <i>The role of trust in IoT and ICS</i> David Mattes, Asguard Networks David Manz, PNNL Steve Venema, Boeing Ron Ross, Nist</p>
12:30	<p>Lunch</p>

13:30	Session 2 Security Patterns for Local Assurance in Cloud Applications [PDF] Marcos Arjona, Jose Fran. Ruiz and Antonio Maña(University of Malaga) Discrete Mathematical Approaches to Graph-Based Traffic Analysis [PDF] Cliff Joslyn, Wendy Cowley, Emilie Hogan and Bryan Olsen(Pacific Northwest National Library) Security concept for gateway integrity protection in a smart grid environment [PDF] Kai-Oliver Detken, Carl-Heinz Genzel, Richard Sethmann and Olav Hoffmann
15:00	Coffee Break
15:15	Discussion: Developing an ECSaR Position Statement on Engineering Cyber Security Moderation: David Manz and Carsten Rudolph
17:00	End of workshop
19:30	Dinner (to be decided)

Keynote Speaker

Joseph Weiss

Joseph Weiss is an industry expert on control systems and electronic security of control systems, with more than 35 years of experience in the energy industry. Mr. Weiss spent more than 14 years at the Electric Power Research Institute (EPRI) where he led a variety of programs including the Nuclear Plant Instrumentation and Diagnostics Program, the Fossil Plant Instrumentation & Controls Program, the Y2K Embedded Systems Program and, the cyber security for digital control systems. As Technical Manager, Enterprise Infrastructure Security (EIS) Program, he provided technical and outreach leadership for the energy industry's critical infrastructure protection (CIP) program. He was responsible for developing many utility industry security primers and implementation guidelines. He was also the EPRI Exploratory Research lead on instrumentation, controls, and communications. Mr. Weiss serves as a member of numerous organizations related to control system security. These include the North American Electric Reliability Corporation (NERC) Control Systems Security Working Group (CSSWG), the International Electrotechnical Commission (IEC) Technical Committee (TC) 57 Working Group 15 - Data and Communication Security, the Process Controls Security Requirements Forum, CIGRÉ WG D2.22 - Treatment of Information Security for Electric Power Utilities (EPU), and other industry working groups. He served as the Task Force Lead for review of information security impacts on IEEE standards. He is also a Director on ISA's Standards and Practices Board. He has provided oral and written testimony to three House subcommittees, one Senate Committee, and a formal statement for the record to another House Committee. He has also responded to

numerous Government Accountability Office (GAO) information requests on cyber security and Smart Grid issues. He is also an invited speaker at many industry and vendor user group security conferences, has chaired numerous panel sessions on control system security, and is often quoted throughout the industry. He has published over 60 papers on instrumentation, controls, and diagnostics including a chapter on cyber security for Electric Power Substations Engineering, Protecting Industrial Control Systems from Electronic Threats (ISBN 978-1-60650-197-9), and Cyber Security Policy Guidebook (ISBN: 978-1-118-02780-6). He supported MITRE and NIST in extending NIST SP800-53 to include control systems and the development of NIST SP800-82. He was tasked to write the White Paper on Industrial Control Systems Security for the Center for Strategic and International Studies Blue Ribbon Panel preparing cyber security recommendations for the Obama administration. Mr. Weiss has conducted SCADA, substation, plant control system, and water systems vulnerability and risk assessments and conducted short courses on control system security. He is a member of Transportation Safety Board Committee on Cyber Security for Mass Transit. He also established and chairs the annual Industrial Control System (ICS) Cyber Security Conference. Mr. Weiss has received numerous industry awards, including EPRI Presidents Award (2002) and is an ISA Fellow, Managing Director of ISA Fossil Plant Standards, ISA Nuclear Plant Standards, ISA Industrial Automation and Control System Security, a Ponemon Institute Fellow, and an IEEE Senior Member. He is a Voting Member of the TC65 TAG and a US Expert to TC65 WG10, Security for industrial process measurement and control – network and system security and IEC TC45A Nuclear Plant Cyber Security. He has two patents on instrumentation and control systems, is a registered professional engineer in the State of California, a Certified Information Security Manager (CISM) and Certified in Risk and Information Systems Control (CRISC).

Panel Members



Ron Ross

National Institute of Standards and Technology
Information Technology Laboratory
Computer Security Division
100 Bureau Drive (Mail Stop 8930)
Gaithersburg, MD 20899-8930
Office: (301) 975-5390
Mobile: (301) 651-5083

Ron Ross is a Fellow at the National Institute of Standards and Technology (NIST). His current focus areas include information security and risk management. Dr. Ross leads the Federal Information Security Management Act (FISMA) Implementation Project, which includes the development of security standards and guidelines for the federal government, contractors, and the United States critical information infrastructure. His recent publications include Federal Information Processing Standards (FIPS) Publication 199 (security categorization standard), FIPS Publication 200 (security requirements standard), NIST Special Publication (SP) 800-39 (risk management guideline), NIST SP 800-53 (security and privacy controls guideline), NIST SP 800-53A (security assessment

guideline), NIST SP 800-37 (security authorization guideline), and NIST SP 800-30 (risk assessment guideline). Dr. Ross is the principal architect of the Risk Management Framework and multi-tiered approach that provides a disciplined and structured methodology for integrating the suite of FISMA standards and guidelines into a comprehensive enterprise-wide security program. Dr. Ross also leads the Joint Task Force, an interagency partnership with the Department of Defense, the Office of the Director National Intelligence, and the Committee on National Security Systems) that developed the Unified Information Security Framework for the federal government.

In addition to his responsibilities at NIST, Dr. Ross supports the U.S. State Department in the international outreach program for information security and critical infrastructure protection. He has also served as a guest lecturer at universities and colleges including the Massachusetts Institute of Technology, Dartmouth College, Stanford University, and the Naval Postgraduate School. Dr. Ross previously served as the Director of the National Information Assurance Partnership, a joint activity of NIST and the National Security Agency. A graduate of the United States Military Academy at West Point, Dr. Ross served in many leadership and technical positions during his twenty-year career in the United States Army. While assigned to the National Security Agency, Dr. Ross received the Scientific Achievement Award for his work on an inter-agency national security project and was awarded the Defense Superior Service Medal upon his departure from the agency. Dr. Ross is a three-time recipient of the Federal 100 award for his leadership and technical contributions to critical information security projects affecting the federal government and is a recipient of the Department of Commerce Gold and Silver Medal Awards. Dr. Ross has been inducted into the Information Systems Security Association (ISSA) Hall of Fame and given its highest honor of ISSA Distinguished Fellow. Dr. Ross has also received many private sector information security awards and recognition including the Applied Computer Security Associates (ACSA) Distinguished Practitioner Award, the Vanguard Chairman's Award, the Symantec Cyber 7 Award, InformationWeek's Government CIO 50 Award, Best of GTRA Award, ISACA National Capital Area Conyers Award, SC Magazine's Cyber Security Luminaries, (ISC)² Inaugural Lynn F. McNulty Tribute Award, and the Top 10 Influencers in Government IT Security. During his military career, Dr. Ross served as a White House aide and as a senior technical advisor to the Department of the Army. Dr. Ross is a graduate of the Defense Systems Management College and holds Masters and Ph.D. degrees in Computer Science from the U.S. Naval Postgraduate School specializing in artificial intelligence and robotics.

David Mattes founded Asguard Networks, Inc., in 2012 and serves as its Chief Executive Officer. He serves as Director of Asguard Networks, Inc. Mr. Mattes creates products that address the challenge of managing connectivity and information security for Industrial Control Systems (ICS). Prior to Asguard Networks, he spent 13 years in Boeing's R&D organization. At Boeing, he focused on ICS security issues, particularly on the challenge of segmenting connectivity for ICS devices into private networks and securely connecting them to and through Boeing's Enterprise networks. Mr. Mattes served as co-creator and technical and implementation lead on an architecture that not only satisfied Boeing's InfoSec governance and security requirements, but also met the needs of the end users. In November of 2011, Mr. Mattes was nominated for the Boeing R&D Engineer of the Year award. He received an MSEE from the University of Washington and a BSEE from the University of New Mexico.

David Manz is currently a Senior Cyber Security Scientist in the National Security Directorate at the Pacific Northwest National Laboratory. He holds a B.S. in Computer and Information Science from the Robert D. Clark Honors College at the University of Oregon and a Ph.D. in computer science from the University of Idaho. David's work at PNNL includes enterprise resilience and cyber security, secure control system communication, and critical infrastructure security. Prior to his work at PNNL, David spent five years as a researcher on Group Key



Management Protocols for the Center for Secure and Dependable Systems at the University of Idaho (U of I). David also has considerable experience teaching undergraduate and graduate computer science courses at U of I, and as an adjunct faculty at WSU. David has co-authored numerous papers and presentations on cyber security, control systems, and cryptographic key management.

Speech Title: Future of Data Intensive Applications
Dr. Milind Bhandarkar
Chief Scientist
Pivotal, Inc.



Biographical Sketch: Milind Bhandarkar was the founding member of the team at Yahoo! that took Apache Hadoop from 20-node prototype to datacenter-scale production system, and has been contributing and working with Hadoop since version 0.1.0. He started the Yahoo! Grid solutions team focused on training, consulting, and supporting hundreds of new migrants to Hadoop. Parallel programming languages and paradigms has been his area of focus for over 20 years. He worked at the Center for Development of Advanced Computing (C-DAC), National Center for Supercomputing Applications (NCSA), Center for Simulation of Advanced Rockets, Siebel Systems, Pathscale Inc. (acquired by QLogic), Yahoo! and LinkedIn. Currently, Milind Bhandarkar is the Chief Scientist at Pivotal, a new EMC joint venture with VMware that includes the Greenplum, Pivotal Labs, SpringSource, Cloud Foundry and Cetas business lines. Pivotal is building big data infrastructure that can handle next-generation workloads. Milind Bhandarkar holds his Ph.D. degree in Computer Science from the University of Illinois at Urbana-Champaign.

Abstract: “Big Data” is a much-hyped term nowadays in Business Computing. However, the core concept of collaborative environments conducting experiments over large shared data repositories has existed for decades. In this talk, I will outline how recent advances in Cloud Computing, Big Data processing frameworks, and agile application development platforms enable Data Intensive Cloud Applications. I will provide a brief history of efforts in building scalable & adaptive run-time environments, and the role these runtime systems will play in new Cloud Applications. I will present a vision for cloud platforms for science, where data-intensive frameworks such as Apache Hadoop will play a key role.

**Speech Title: System U: Computational Discovery of Personality Traits from Social Media
to Deliver Hyper-Personalized Experience**

Speaker: Dr. Michelle Zhou
Senior Research Manager
IBM Research–Almaden



Biographical Sketch: Dr. Michelle Zhou is a Senior Research Manager at IBM Research – Almaden, where she manages the User Systems and Experience Research (USER) group. Zhou received a Ph.D. in Computer Science from Columbia University. Her expertise is in the interdisciplinary areas of intelligent user interaction, information visualization and smart visual analytics, and analytics-driven social computing and crowdsourcing. She has published over 70 peer-reviewed, refereed articles and filed over twenty patents in above areas. Zhou is active in several research communities, including Intelligent User Interfaces, Information Visualization and Visual Analytics, and Multimedia. She has co-organized/co-chaired conferences and workshops, and served on the technical program committees for key conferences in these areas. She is the general conference co-chair for ACM Recommender Systems 2014, was the general conference co-chair for ACM IUI 2007, and the technical program co-chair for ACM Multimedia 2009 and Intelligent User Interfaces 2010. Michelle has served on the editorial board of three journals, ACM Transactions on Multimedia Computing, Communications, Applications (TOMCCAP), ACM Transactions on Intelligent Systems and Technology (TIST), and ACM Transactions on Interactive Intelligent Systems. Zhou was named an ACM Distinguished Scientist in 2009.

Abstract: Hundreds of millions of people leave digital footprints in public (e.g., social media/social networking sites and review sites). We are developing System U, which uses psycholinguistic analytics to automatically derive one's personality traits from their digital footprints. Such traits uniquely characterize an individual's psychological, cognitive, and affective style and properties, and can then be used to make hyper-personalized recommendations to the individual and influence/intervene the actions of the individual. In this talk, I will give an overview of System U and describe how it automatically derives several types of personality traits from one's tweets, including human basic value (one's belief + motives) and fundamental needs (e.g., ideals vs. practical). Moreover, I will present a set of validation studies that assess how accurate the System U-derived traits are compared to "ground truth" and how these derived traits influence recommendations and people's behavior in the real world. I will use live demos and concrete examples, ranging from precision marketing to individualized customer care, to demonstrate the applications of System U and discuss research directions in this space.

Speech Title: A Building Code for Building Code

Dr. Carl Landwehr
Lead Research Scientist
George Washington University



Biographical Sketch: Dr. Carl Landwehr has been a leader in cyber security research and development since the late 1970s. Following an active [research career](#) at the Naval Research Laboratory, he led the National Science Foundation's efforts to establish and build national research programs in Trusted Computing, Cyber Trust, Trustworthy Computing and Secure and Trustworthy Cyberspace. He helped DARPA develop its programs in Information Assurance and served on its ISAT advisory group. At IARPA, he initiated programs to develop and apply concepts of information flow analysis, automated software analysis and private information retrieval that continue to produce results. Professionally, he served four years as editor-in-chief of *IEEE Security and Privacy Magazine*. He was an early chair of the IEEE Technical Committee on Security & Privacy and has chaired the IEEE Symposium on Security & Privacy. He was the founding chair of IFIP WG11.3 on Data and Application Security, and is an active member of IFIP WG10.4 on Dependability and Fault Tolerance. In the past two years he has served as an advisor on cybersecurity research and software engineering education programs to the governments of Australia and Israel, respectively, as well as two major US laboratories. He was made an IEEE Fellow for his contributions to cybersecurity and was in the first class of 11 individuals inducted into the National Cyber Security Hall of Fame in 2012. Other awards include the NSF Director's Award for Meritorious Service, IEEE's Distinguished Service Award, ACM SIGSAC's Outstanding Contribution Award. At present, He is an independent consultant and a Lead Research Scientist at the Cyber Security Policy and Research Institute at George Washington University, as well as a member of the NSF-funded Trustworthy Health and Wellness (THaW) research project, led by Dartmouth University and including Johns Hopkins University, University of Illinois, University of Michigan, and Vanderbilt.

Abstract: Cyberspace, though it has a physical reality of computers and communication channels, sensors and actuators, is in fact made real mostly by the programs that control those things. Today, systems of programs control most of our critical infrastructures. Metaphors are frequently used as a way to communicate to people what these programs are intended to do. Workers in cybersecurity have adopted many rich metaphors: Trojan Horse, virus, worm, firewall, and more. Difficulties arise when the metaphor blinds us to the underlying reality. The talk examines critically several common cybersecurity metaphors and proposes the adoption of a new (or at least

underutilized) one, that of a building code for critical infrastructure software, as a means of putting what we have learned in forty years of system development experience into practice.

Speech Title: Insight into the CIO World of Cybercrime and Cloud computing: the Cybercrime Business, What are Companies Protecting, and What Can be done to Protect the Integrity of Organizations.

Chris Scotte
Vice President Marketing and Business Development
Stealth Software



Biographical Sketch: Chris Scotte is the VP of Marketing and Business Development for Stealth Software. Stealth Software is an IT software company that has developed innovative security software that solves many of the problems of security, trust and compliance when organizations use the cloud. Stealth Software enables companies to use a range of collaboration platforms that optimize organizational performance in a cost-effective and secure manner. Stealth Software has won six awards from Gartner around data security on premise and in the cloud, and won the most promising company award by PWC Accelerator in 2012. The company has been asked by the European Commission to advise them regarding cloud security. Chris has a background in management consulting and education where he has taught and has conducted numerous research projects in South Korea, Australia, China and the Middle East. Chris has a Bachelor of Business (Human Resource Management) from Griffith University, a Master of Business (Marketing) from Queensland University of Technology and a Master of Management (Research) from Griffith University.

Abstract: Cybercrime is big business - both in the U.S. and globally. This talk will highlight the current situation with regards to what CIOs must deal with today, the information that organizations are protecting, and the difference between perceived threat and reality. A major focus of this speech will be about using the cloud and the major concerns organizations have when including the cloud as part of their IT environment. The research for this speech comprises an analysis of key industry white papers and feedback from Stealth Software customers.

Speech Title: Big Data in the Finder and Aladdin Video Programs

Dr. Jill D. Crisman
Program Manager
IARPA

Abstract: The Intelligence Advanced Research Projects Activity (IARPA) Incisive Analysis Office sponsors programs that help analysts make sense of massive data. This talk will focus on two such programs, Finder and Aladdin Video. Finder is developing technologies that can locate where in the world a query image or video was taken based on the query's content alone. The Aladdin Video program is developing technologies that can quickly search massive video collections for a user's events-of-interest. This talk overviews the goals and objectives of these programs, examines current results, and illustrates the size of the data involved.

Speech Title: Generate comprehensive threat reports and security scores for millions of android Apps using smart threat analytics

Dr. Wei Yan
VisualThreat Inc.



Biographical Sketch: Dr. Wei Yan is currently CEO of VisualThreat, a leading mobile security vendor. He worked as Security Solution Architect at major security vendors. He conducted cutting edge research in anti-malware in-the-cloud, stream-based security solutions for next-generation firewalls, mobile security, malware detection and classification, unpacking, botnets, data recovery, and intrusion detection systems. Dr. Yan is also an active referee and serves as Editorial Board member of several peer-reviewed professional journals and Technique Committee member of more than 50 international security-related conferences. He is the author of leading industry and academic conferences, such as Virus Bulletin 2009 and AVAR 2009, 2010, 2012, and SYSCAN 2012, 2013. He has a PhD in computer engineering from New Jersey Institute of Technology.

Abstract: Traditional mobile security solutions are severely limited in detecting new malware variants. They are very hard to provide real-time and accurate visibility of mobile threat status. In this talk, we have built a free online mobile security tool where mobile malware content are cross-indexed on the family base. It generates the most comprehensive threat report, which is able to detect each submitted sample which malware family it belongs to, with which other families to share similar codes, which other individual samples have the same or similar malicious activities as the submitted sample, where these information are located inside samples. All of the above queries can be processed in seconds to save tremendous tedious work for users and security professionals.

Speech Title: Advancing Big Data Analytics and Data Science through Measurements, Evaluations and Challenge Problems

Dr. Ashit Talukder

Chief, Information Access Division
National Institute of Standards and Technology



Biographical Sketch: Dr. Ashit Talukder leads and manages the Information Access Division in the Information Technology Laboratory at the National Institute of Standards and Technology (NIST). At NIST he leads and directs a division of over 100 researchers and staff in the areas of information access, data processing, and information search and retrieval for multimedia, biometrics, Big Data, visualization, image processing, computer vision, video analytics, speech processing, speech recognition, machine translation, human-computer interaction, human-factors and usability, and multimodal data for science, defense and other applications. The programs that he leads at IAD involves collaborations within NIST, and with many federal agencies including DARPA, IARPA, NSA, FBI, DoD, DoJ, DHS and others. He initiates and leads new research efforts in the above areas, and facilitates collaborations and partnerships between research labs, academia and industry. He received his Ph.D. in computer science from Carnegie Mellon University and a MS from Iowa State University. He was previously at the Jet Propulsion Laboratory, California Institute of Technology, a federally funded research and development center at NASA. He has served as a research faculty member at the University of Southern California (USC). His research background and expertise covers machine learning, search and retrieval, human language technologies, NLP, pattern recognition, image / signal processing, multimedia processing, video understanding, computer vision, data mining and analytics of massive datasets, Big Data, distributed control, biometrics, robotics, sensor networks and cyber-physical systems. He has led several projects and programs funded by DARPA, NSF, NIH, DHS, and other agencies. He has more than 65 journal and conference publications, and has served as a reviewer on numerous conferences and journals. He was a recipient of the Premium Award for Academic Excellence (1992), and several NASA Space Act Awards. He is an inventor of 2 patents. He is a member of the ACM and the IEEE Computer Society.

Speech Title: Why Data Science?

Dr. Stanley C. Ahalt

Director, Renaissance Computing Institute (RENCI)

Professor of Computer Science

University of North Carolina at Chapel Hill



Biographical Sketch: Stanley C. Ahalt, PhD, is director of the Renaissance Computing Institute (RENCI), professor of computer science at the University of North Carolina at Chapel Hill, and the head of the Biomedical Informatics Service for the North Carolina Translational and Clinical Sciences Institute (NC TraCS). As director of RENCI, he was instrumental in launching two major data science initiatives: The National Consortium for Data Science (NCDS), a public-private partnership to advance the field of data science and address big data challenges and opportunities in research and business; and the iRODS Consortium, an international group aimed at sustaining the popular integrated Rule-Oriented Data System (iRODS) as enterprise-quality software. Dr. Ahalt is PI on the Water Science Software Institute, a U.S. National Science Foundation project to prototype an integrated platform for sharing, using, and managing data across water science fields. Other leadership roles include past chair of the Coalition for Academic Scientific Computing, former co-chair of the Ohio Broadband Council, extramural member of the National Cancer Institute's Advanced Biomedical Computing Center's Oversight Committee, and member of the Council on Competitiveness High Performance Computing Advisory Committee. He has authored or co-authored more than 120 technical papers and been principal investigator or co-principal investigator on research [grants](#) totaling more than \$17 million. His recent research publications have focused on decoupling data through encryption. Before coming to RENCI in 2009, Dr. Ahalt was executive director of the Ohio Supercomputer Center from 2003 – 2009 and a professor in the department of electrical and computer engineering at The Ohio State University for 22 years. He holds a PhD in electrical and computer engineering from Clemson University and master's and bachelor's degrees in electrical engineering from Virginia Polytechnic Institute and State University.

Abstract: Data is has become the central driving force of key to insight, governance, science new discoveries in science, informed governance, insight into society, improved healthcare, and economic growth in the 21st century. Abundant data is a direct result of innovations including the Internet, faster computer processors, [cheap storage](#), and the proliferation of sensors. Further, abundant data has the potential to improve healthcare, [increase business](#) productivity, and enable scientific discovery. However, while data is abundant and everywhere, we do not

have an understanding of data at a fundamental level. We are still struggling to develop ways to value, manage and interpret data that grows every day. We believe that the reason for this is that data – as a field of research – has not been subjected to systematic scientific analysis. In this talk Stanley C. Ahalt, PhD, will argue that data has become so critical to innovation and discovery that we must approach it the same way we approach other scientific disciplines. We must develop a data science: a systematic study of the structure and behavior of physical and natural data through both observation and experimentation. Dr. Ahalt will discuss how we can begin to understand the economics and mathematics of data and what this means for data scientists. He will also discuss the need to develop methodologies for measuring the value of data, and organizations that have been founded to advance the field of data science, such as the U.S.-based National Consortium for Data Science.

Speech Title: Healthcare and Data Science

Vinod Khosla
Partner of Khosla Ventures

Abstract: Healthcare today is often really the “practice of medicine” rather than the “science of medicine”. Healthcare should be much more scientific and data-driven, but that’s hard for the average physician to pull off without technology, because of the increasing amount of data and research released every year. Technology makes up for human deficiencies and amplifies strengths. Automated healthcare will amplify physicians by arming them with more complete, synthesized, and up-to-date research data, all leading to better patient outcomes, and next-generation medicine will be the scientific arrival at diagnostic and treatment conclusions & real testing of what’s actually going on in your body. It will also be much more personalized than your physician can provide. Eventually, this shift in how healthcare is delivered will allow for less money to be spent on capital equipment, cutting health care costs. It will allow us to provide care to those who don’t have it now. And, it will prevent simple things from getting worse before being addressed. In this Q&A, Vinod will address the future of data-driven healthcare and how machine learning and automation will be used to enable the science of medicine.

Invited Session: Security in Big Data

Chair: Dr. TY Lin

Panelist: Dr. Tom Hinke, Dr. Kai Hwang, Dr. Carl Landwehr, Mr. Eric Leighninger, Dr. TY Lin, Dr. Sylvia Osborn, Dr. Bhavani Thuraisingham, Dr. Felix Wu



Dr. Tsau Young (T.Y.) Lin received his PhD in Mathematics from Yale University. He is a Professor of Computer Science at San Jose State University and a fellow in Berkeley Initiative in Soft Computing, University of California. He is the President of International Granular Computing Society. He is one of EIC of *International Journal of Granular Computing*, *Rough Sets and Intelligent Systems*. He has served on various roles in reputable international journals and conferences. His interests include data/text/web mining, data security and novel methodology, such as granular/fuzzy/Petri net/rough computing. He received (1) Best contribution awards from ICDM01 and (2) Best service award from IEEE/WIC/ACM WI-IAT2007 and (3) A pioneer award from GrC 2008.

Dr. Thomas H. Hinke has been at NASA Ames Research Center since 1999, where he is currently the government lead for IT security for NASA's largest supercomputer. Prior to joining NASA he was a tenured associated professor in the Computer Science Department at the University of Alabama in Huntsville, where he created the department's first computer security course and performed sponsored research in database inference. Previous experience at TRW and System Development Corporation involved research and development in secure databases, secure end-to-end encryption and other areas in the high-assurance computer security area. He holds a Bachelor's degree in Electrical Engineering from the University of California Berkeley, an MBA degree from Oklahoma City University, a Master's degree in Computer Science from UCLA, and a Ph.D. degree in Computer Science from the University of Southern California.

Dr. Kai Hwang is a Professor of EE/CS at the University of Southern California (USC). He is also an EMC-endowed visiting Chair Professor at Tsinghua University. He received the Ph.D. from University of California, Berkeley in 1972. He has published 8 books and over 220 scientific papers in computer architecture, parallel processing, distributed systems, cloud computing and network security. His books have been adopted worldwide and translated into Chinese, Korean, Spanish, and German languages. His works have been cited more than 12,000 times with an h-index of 48. His latest book: *Distributed and Cloud Computing* was published by Kaufmann in 2012. Dr. Hwang was recognized with an IEEE Fellow in 1986. He received the very-first 2004 CFC Outstanding Achievement Award, the IPDPS-2011 Founder's Award, and the Lifetime Achievement Award from the IEEE Cloud2012 for his pioneering work in parallel computing and distributed systems. He has served as the founding Editor-in-Chief of the *Journal of Parallel and Distributed Computing* for 28 years. He has produced numerous Ph.D students at Purdue University and at USC. Five of his former Ph.D students are elected IEEE Fellows and IBM Fellow. He has delivered three dozens of keynote addresses on advanced computing systems and cutting-edge information technologies in major IEEE/ACM Conferences. Hwang has performed advisory, consulting and collaborative work for IBM, Intel, MIT Lincoln Lab, JPL at Caltech, ETL in Japan, Academia Sinica in China, GMD in Germany, and INRIA in France.

Dr. Carl Landwehr has been a leader in cyber security research and development since the late 1970s. Following an active [research career](#) at the Naval Research Laboratory, he led the National Science Foundation's efforts to establish and build national research programs in Trusted Computing, Cyber Trust, Trustworthy Computing and Secure and Trustworthy Cyberspace. He helped DARPA develop its programs in Information Assurance and served on its ISAT advisory group. At IARPA, he initiated programs to develop and apply concepts of information flow analysis, automated software analysis and private information retrieval that continue to produce results. Professionally, he served four years as editor-in-chief of *IEEE Security and Privacy Magazine*. He was an early chair of the IEEE Technical Committee on Security & Privacy and has chaired the IEEE Symposium on Security & Privacy. He was the founding chair of IFIP WG11.3 on Data and Application Security, and is an active member of IFIP WG10.4 on Dependability and Fault Tolerance. In the past two years he has served as an advisor on cybersecurity research and software engineering education programs to the governments of Australia and Israel, respectively, as well as two major US laboratories. He was made an IEEE Fellow for his contributions to cybersecurity and was in the first class of 11 individuals inducted into the National Cyber Security Hall of Fame in 2012. Other awards include the NSF Director's Award for Meritorious Service, IEEE's Distinguished Service Award, ACM SIGSAC's Outstanding Contribution Award. At present, He is an independent consultant and a Lead Research Scientist at the Cyber Security Policy and Research Institute at George Washington University, as well as

a member of the NSF-funded Trustworthy Health and Wellness (THaW) research project, led by Dartmouth University and including Johns Hopkins University, University of Illinois, University of Michigan, and Vanderbilt.

Mr. Eric Leighninger has been working in the information security field for nearly 30 years. He has been involved in projects for government and commercial organizations ranging from security architecture, trusted software development, cryptographic technologies, and identity management systems to risk, audit and compliance assessments. He retired in 2010 as chief security architect for a major insurance company and previously was director of security research for a major consulting firm. His recent activities include conducting ISO 27001/27002 evaluations and other risk and compliance reviews. He is completing his CISM certification and is certified as a QSA for PCI compliance assessments. Mr. Leighninger is a U.S. Navy veteran and holds the MA degree in mathematics from Arizona State University and the MDiv degree from Princeton Theological Seminary.



Dr. Sylvia Osborn received her PhD in Computer Science from the University of Waterloo. Since 1977, she has been a faculty member in the Computer Science Department at the University of Western Ontario in London, Ontario, Canada. She is the author of numerous research papers, starting in the database field in dependency theory, and object-oriented databases. More recently she has been active in research into role-based access control including comparison of access control models, administration of access control, delegation. Recently, she has been focusing on the integration of privacy issues with access control, and how the consideration of privacy of individuals' data does or does not differ from access control.

Dr. Bhavani Thuraisingham is the Louis A. Beecherl, Jr. Distinguished Professor of Computer Science and the Executive Director of the Cyber Security Research and Education Institute (CSI) at The University of Texas at Dallas. She is an elected Fellow of IEEE, the AAAS, the British Computer Society, and the SPDS (Society for Design and Process Science). She received several prestigious award including IEEE Computer Society's 1997 Technical Achievement Award for "outstanding and innovative contributions to secure data management", the 2010 ACM SIGSAC (Association for Computing Machinery, Special Interest Group on Security, Audit and Control) Outstanding Contributions Award for "seminal research contributions and leadership in data and applications security for over 25 years" and the SDPS Transformative Achievement Medal for her contributions to interdisciplinary research. She has unique experience working in commercial industry, research laboratory, US government and academia and her 30+ year career includes research and development, technology transfer, product development, program management, and consulting for the federal government. Her work has resulted in 100+ journal articles, 200+ conference papers, 100+ keynote and invited talks, six US patents (two pending) and twelve books.

Dr. S. Felix Wu has been doing "experimental" system research, i.e., building prototype systems to justify and validate novel architectural concepts. Since 1995, he and his students/postdocs have built many experimental systems in the areas of fault tolerant network, IPSec/VPN security policy, attack source tracing, wireless network security, intrusion detection and response, visual information analytics, and, more recently, future Internet design. An article titled "Networking: Four ways to reinvent the Internet" published in Nature 463 (February 3rd, 2010, by Katharine Gammon) provided a brief but very nice cover about his primary thought on a Social-network-based

future Internet architecture. During the past seven years, he has been pretending (and hoping) to know a little bit more about humanity science so he can claim that he is working on multidisciplinary research. He strongly believes that thoroughly considering the factor of human relationships is necessary for any IT innovation. Therefore, his primary research objective, before he retires, is to help and contribute to the information technology advancement that would truly help our human society. As an initial step, he recently released the SINCERE (Social Interactive Networking and Conversation Entropy Ranking Engine, sponsored by NSF) search engine under <http://www.sincere.se>, which is trying to help our Internet society to discover “interesting/unusual” discussions. Felix received his BS from Tunghai University, Taiwan, in 1985, both MS and PhD from Columbia University in 1989 and 1995, all in Computer Science. He has about 120+ academic publications, which means that he should probably focus much more on the depth and quality. He is currently a Professor with the Computer Science department at UC Davis.

Speech Title: Enabling Cloud Analytics for Big-Data Security and Intelligence

Dr. Kai Hwang

Professor

University of Southern California



Biographical Sketch: Kai Hwang is a Professor of EE/CS at the University of Southern California (USC). He is also an EMC-endowed visiting Chair Professor at Tsinghua University. He received the Ph.D. from University of California, Berkeley in 1972. He has published 8 books and over 220 scientific papers in computer architecture, parallel processing, distributed systems, cloud computing and network security. His books have been adopted worldwide and translated into Chinese, Korean, Spanish, and German languages. His works have been cited more than 12,000 times with an h-index of 48. His latest book: *Distributed and Cloud Computing* was published by Kaufmann in 2012. Dr. Hwang was recognized with an IEEE Fellow in 1986. He received the very-first 2004 CFC Outstanding Achievement Award, the IPDPS-2011 Founder’s Award, and the Lifetime Achievement Award from the IEEE Cloud2012 for his pioneering work in parallel computing and distributed systems. He has served as the founding Editor-in-Chief of the *Journal of Parallel and Distributed Computing* for 28 years. He has produced numerous Ph.D students at Purdue University and at USC. Five of his former Ph.D students are elected IEEE Fellows and IBM Fellow. He has delivered three dozens of keynote addresses on advanced computing systems and cutting-edge information technologies in major IEEE/ACM Conferences. Hwang has performed advisory,

consulting and collaborative work for IBM, Intel, MIT Lincoln Lab, JPL at Caltech, ETL in Japan, Academia Sinica in China, GMD in Germany, and INRIA in France. He can be reached via Email: kaihwang@usc.edu.

Abstract: In this talk, Dr. Hwang addresses the growing interest in big-data science surrounding the use of cloud analytics, social networks and *Internet of things* (IoT). He will assess critical issues to upgrade big-data analysis, privacy and cloud security. The purpose is to achieve enhanced ubiquity, mobility, security, scalability and *quality of service* (QoS) of clouds and highly-visited social networks or datacenters. In particular, he will evaluate the widespread use of clouds over massive datasets generated by e-business, social networks, sensors, RFID, GPS, etc. His talk reveals major R/D challenges and presents new approaches to preserving data privacy, assuring cloud security, and enhancing cyber intelligence. To remove the security and trust barriers in bare-metal or virtual clouds, he examines the top-10 security and privacy issues released by Cloud Security Alliance in 2012. Some new approaches and hidden opportunities are discussed towards the building of a trusted and intelligent cloud computing environment over both structured and unstructured big datasets. Finally, he compares the security and capability in BYOD (*Bring Your Own Devices*) solutions with those offered by the new BYOC (*Bring Your Own Clouds*) approach for inter-cloud (mashup) applications.

Speech Title: The Big Deal About Big Data: a Sociobiological Perspective

Speaker: Rebecca D. Costa

American Sociobiologist and Author



Biographical Sketch: Rebecca D. Costa is a sociobiologist and author who offers an evolutionary explanation for current events and emerging trends. A new voice in the mold of Alvin Toffler, Thomas Friedman and Malcolm Gladwell, Costa attributes modern consternation – from terrorism, crime on Wall Street, epidemic obesity and upheaval in the Middle East – to genetic imperatives. Retiring from a career in Silicon Valley, Costa spent six years researching and writing *The Watchman's Rattle: A Radical New Theory of Collapse*. The success of the book in 25 countries led to a weekly nationally syndicated radio news program called *The Costa Report*. Costa is presently represented by Kneerim, Williams & Bloom, the Scott Meredith Literary Agency, the American Program Bureau, VoiceAmerica, the Genesis Communications Network and Global American Broadcasting Network.

Panel: The Cautionary Side of Big Data

Moderator: American Sociobiologist and Author, Rebecca D. Costa

Panelists: Steve Beier, IBM Big Data Program Director

Guy Kawasaki, author, entrepreneur and business advisor

Forrest Melton, Senior Research Scientist NASA Ames Research Center, Cooperative for Research in Earth Science & Technology

Steve Beier



Biographical Sketch: Steve is an experienced corporate entrepreneur with over 18 years in the computer industry. He is a program director of big data at IBM software group. He currently works for IBM Software Group focused on their Big Data product portfolio. IBM's core Big Data portfolio includes InfoSphere BigInsights, InfoSphere Streams and the acquisition products from Netezza and Vivisimo. He works closely with customers and partners in the field who are looking to design and build applications that solve complex problems by utilizing IBM's Big Data products.

Steve has also spent years dealing with the issues of Sensor Information Management at massive scale. He was one of the IBM "co-founders" who led the creation of IBM's Radio Frequency IDentification and Supply Chain Traceability software products and related global standards. Steve received his BS from the University of California, San Diego (UCSD), studied Management Science and Engineering (MS&E) at Stanford University and has an MS in Technology Commercialization from the University of Texas at Austin, Red McCombs School of Business.

Guy Kawasaki



Biographical Sketch: Guy Kawasaki is a special advisor to the Motorola business unit of Google. He is also the author of *APE*, *What the Plus!*, *Enchantment*, and nine other books. Previously, he was the chief evangelist of Apple. Kawasaki has a BA from Stanford University and an MBA from UCLA as well as an honorary doctorate from Babson College.

Forrest Melton



Biographical Sketch: Forrest Melton is a Senior Research Scientist in the Division of Science and Environmental Policy at California State University, Monterey Bay, and the NASA Ames Cooperative for Research in Earth Science and Technology (*ARC-CREST*). Since 2003, he has worked in the Ecological Forecasting Lab at NASA Ames Research Center on the development of modeling and data assimilation frameworks including the Terrestrial Observation and Prediction System (TOPS) and the NASA Earth Exchange (NEX). His research interests include ecosystem and carbon cycle modeling, and applications of satellite data, ecosystem models, and high performance computing to improve management of natural resources. Forrest holds B.S. and M.S. degrees in Earth Systems Science from Stanford University, and he has authored numerous papers and book chapters on applications of remote sensing. He is the recipient of honor awards from NASA for his contributions to TOPS and NEX, and has been recognized for his work on applications of satellite data for water management with awards from the California Department of Water Resources and the Federal Labs Consortium for Technology Transfer.

Speech Title: Scaling R to Big Data Science

Dr. Mario Inchiosa

Revolution Abnalytics, Inc.



Biographical Sketch: Mario Inchiosa's passion for data science and high performance computing drives his work at Revolution Analytics, where he focuses on delivering parallelized, scalable advanced analytics integrated with the R language. Previously, Mario served as Analytics Architect in IBM's Big Data organization and US Chief Scientist in Netezza Labs, advancing Hadoop and SQL-based parallel data analytic platforms. Their success led to Netezza's acquisition by IBM. Mario also served as US Chief Science Officer at NuTech Solutions, a computer science consultancy specializing in simulation, optimization, and data mining, and Senior Scientist at BiosGroup, a complexity science spin-off of the Santa Fe Institute. Dr. Inchiosa holds Bachelors, Masters, and PhD degrees from Harvard University. He has been awarded four patents and has published over 30 research papers, earning Publication of the Year and Open Literature Publication Excellence awards.

Abstract: In this talk, I will describe current challenges and issues in big data analytics -- the talent gap caused by tool chains requiring myriad skills, lack of reusability across platforms, the cost and delay of writing advanced parallelized analytics from scratch -- and how Revolution Analytics is helping to address these challenges via Revolution R Enterprise (RRE). RRE is our commercial distribution of open source R that includes Intel Math Kernel Libraries, select open source extension packages, Revolution's ScaleR functions for scalable, cross-platform advanced analytics, and Revolution's DeployR Web Services framework. We have recently enhanced RRE to support in-Hadoop and in-Teradata analytics with big data implementations of advanced analytics and machine learning algorithms for exploratory data analysis, regression, classification, prediction, and unsupervised learning. We will examine how RRE implements these technical advances, making big data analytics more powerful and accessible.

Speech Title: Big Data and The Customer Decision Journey

Matt Hertig
Co-Founder
Alight Analytics



Matt Hertig is the Co-Founder of Alight Analytics, an independent marketing analytics firm based in Kansas City, Missouri. Under his leadership, Alight has developed an industry-leading marketing analytics platform, ChannelMix, which has established Alight Analytics at the cutting edge of multi-channel marketing analytics. Prior to founding Alight Analytics, Matt built an extensive background in CRM, database marketing, web strategy and business intelligence by serving as an executive leading multi-million dollar database marketing systems and multi-channel web strategies at companies such as AMC Theatres, American Century Investments and Payless ShoeSource.

Abstract: Understanding the paths customers take to making a decision has never been more important. The marketing channel interactions customers utilize in their journey to make decisions represent a Big Data gold mine of information for organizations to maximize customer acquisition and retention efforts. In this session, learn how to establish your customers' journey, determine what data is needed and how to leverage this Big Data strategy to maximize marketing and sales resources.

Speech Title: Simplified Data Parsing and Ingestion with DFDL

Stephen Lawrence
Software Engineer
Tresys Technology, Inc.

Biographical Sketch: Stephen Lawrence has worked as a software engineer at Tresys Technology since 2007, while contributing to the open-source Daffodil project as a core maintainer for almost two years. He works alongside Michael Beckerle, the co-chair of the DFDL Working Group, to develop Daffodil and improve the DFDL specification. Outside of Daffodil, he focuses on computer security applications, including file inspection and sanitization, Security Enhanced Linux (SELinux), and cross domain solutions.

Abstract: Until now, no open standard has been developed with the capability to describe and normalize a wide variety of data formats, including both textual and binary. The Data Format Description Language (DFDL), an upcoming open standard and part of the Open Grid Forum, is a modeling language for describing general text and

binary data using a subset of XML Schema augmented with data format annotations. A DFDL processor reads a schema and data in its native format and normalizes it into an instance of an information set or an XML document, which can also be converted back to its native format. The DFDL information set promotes seamless integration with common XML utilities (e.g. XProc, XSLT) for data ingestion, normalization, processing, and analysis regardless of the format of the native data. DFDL is quickly growing in maturity. Two processors exist, one created by IBM written in both Java and C and includes graphical tools for modeling, parsing, and debugging DFDL schemas. The other processor, Daffodil, is an open-source project written in Scala, with a design focused on speed and correctness. Additionally, multiple commercial and scientific formats have been successfully modeled with DFDL, ranging from simple formats like comma separated values and network packet capture to more complicated formats like health care and electronic payments. With these successes in data modeling, the nearing standardization of the DFDL specification, and the two DFDL implementations making great strides, DFDL is becoming a promising tool that will ease data parsing, ingestion, processing, and analysis.

Speech Title: Delivering on the Promise of Big Data

Speaker: Arvind Parthasarathi
President of YarcData, Cray, Inc.



Biographical Sketch: Arvind Parthasarathi serves as President of YarcData, a Cray company, focused on delivering business solutions for the growing Big Data market. Mr. Parthasarathi joined Cray from Informatica Corporation where he most recently served as Senior Vice President and General Manager of the company's Master Data Management business unit. In this role, he led a global team delivering business-focused data management solutions for life sciences, financial services, retail, manufacturing, healthcare and government. Previously, he served as Vice President of Product Management for Informatica's data quality products, and before that was Senior Director responsible for data migration and data integration solutions. Prior to joining Informatica in 2005, Mr. Parthasarathi was Director of Product Management at i2 Technologies where he was responsible for a number of data-centric supply chain applications including RFID and Event Management. He started his career in product management at Oracle Corporation. Mr. Parthasarathi holds a bachelor's degree in computer science from the Indian Institute of Technology and a master's degree in computer science from the Massachusetts Institute of

Technology. He also serves on the Board of Trustees of the Center for Excellence in Education, a non-profit organization dedicated to nurturing high school and college students into careers in science and technology.

Speech Title: Exploiting Parallelism and Scalability for Big Data in Medicine

Speaker: Dr. Akash Singh
CEO, Neuron Networks



Biographical Sketch: Dr. Akash Singh is CEO, Neuron Networks, previously Senior Architecture Advisor at IBM Corporation, California. He has received numerous honors for his work; he is a Chartered Scientist of Science Council UK, BCS Fellow, senior member of IEEE, member of AACR, AAAS, and AAAI. He is a recognized scientific leader and pioneer in the areas of artificial intelligence, brain imaging understanding, pattern theory, computer vision, medical imaging, computational anatomy and computational neuroscience. He created a new area of Cognitive computing algorithms for Computational Complexity and Big Data problems unsolved for decades especially in Medical Sciences. He led research projects on Big Data, Cloud Computing, Supercomputing, Cognitive Radio and High Performance computing and Brain imaging platforms. He is invited as Keynote plenary speaker and tutorial lecture worldwide; published more than 550 papers, 20 books on Artificial Intelligence, Brain Imaging, Cancer Imaging, Big Data and Bigdata in Medicine and patents. Dr Singh participates in organizing conferences on Computational Intelligence, Artificial Intelligence, serves as Chair for the IEEE Conferences and many other worldwide conferences. He serves on the Editorial Board of over two hundred professional journals.

Abstract: This research talk is for the purpose of developing the parallel computing platform and elastic scalable for the year 2020 and especially to run Big Data platform. The demand for increased parallelism in computing systems is partially due to the need for high performance, highly reactive systems that interact with other environments (audio/video systems, control systems, networked applications, etc. One of the major goals of concurrent computing systems is to support heterogeneity. New heterogeneous architectures continue to provide increases in achievable performance, but programming these devices to reach maximum performance levels is not

straightforward. All computing systems, from mobile to supercomputers, are becoming heterogeneous parallel computers using both multi-core CPUs and many-thread GPUs for higher power efficiency and computation throughput. While the computing community is racing to build tools and libraries to ease the use of these heterogeneous parallel computing systems, effective and confident use of these systems will always require knowledge about the low-level programming interfaces in these systems. **(1). Increased Processing power** Data volumes are growing exponentially. There are many reasons for this growth, including the creation of nearly all data today in digital form, a proliferation of sensors, and new data sources such as high-resolution imagery and video. The collection, management, and analysis of data is a fast-growing concern of XPS research. Automated analysis techniques such as data mining and machine learning facilitate the transformation of data into knowledge, and of knowledge into action. Every Federal agency needs to have a big data strategy. **(2). More security** Engineering large software systems to ensure that they are secure (behaving as expected in the presence of an adversary) and trustworthy (behaving as expected in the absence of an adversary) remains a daunting challenge. The growing complexity of the systems we are building and our increasing societal reliance upon them outpace our ability to reason about them, and to engineer them to be secure and trustworthy. **(3). More Privacy** As XPS increasingly pervades daily life, systems are storing and processing a greater volume and diversity of private information about individuals. Privacy is a critical issue in all societal applications of XPS most obviously in areas such as healthcare and electronic commerce, but also in areas such as energy, transportation, and education. Privacy challenges do not and must not require us to forgo the benefits of XPS in addressing national priorities. Rather, we need a practical science of privacy protection, based on fundamental advances in XPS, to provide us with tools we can use to reconcile privacy with progress. **Broader Impacts:** Advances in Parallelism and Scalability research are crucial to achieving our major national and global priorities in energy and transportation, education and life-long learning, healthcare, and national and homeland security. It accelerates the pace of discovery in nearly all other fields of research including progression of Alzheimer's disease, nature of combustion, size of the ozone hole, etc. In addition to addressing chronic conditions, it pursues advances in the innovative use of research to understand the basic biological and psychological mechanisms that underlie disease, transformational innovation in healthcare delivery. XPS research area also includes the interactions of people with computing systems and devices, interactions between XPS research and the physical world, such as in sensors, imaging, robotic and vision systems, and wearable and mobile devices; large-scale data capture, management and analysis; systems that protect personal privacy and sensitive confidential information, are robust in the face of malfunction, and stand up to cyber-attack; scalable systems and networking (i.e., systems and networks that can be either increased or decreased in complexity, size, generality, and cost); and software creation and evolution. HPCC is but one of many important areas of research, and America's prowess in HPCC is but one of many measures of our international competitiveness in research.

Speech Title: Big Data and Medicine: from genome to populations

Speaker: Dr. Russ Altman

Kenneth Fong Professor of Bioengineering, Genetics, Medicine and Computer Science
Stanford University



Biographical Sketch: Dr. Russ Biagio Altman is a professor of bioengineering, genetics, & medicine (and of computer science, by courtesy) and past chairman of the Bioengineering Department at Stanford University. His primary research interests are in the application of computing and informatics technologies to basic biological problems relevant to medicine. He is particularly interested in methods for understanding drug action at molecular, cellular, organism and population levels. His lab studies how human genetic variation impacts drug response (e.g. <http://www.pharmgkb.org/>). Other work focuses on the analysis of biological molecules to understand the action, interaction and adverse events of drugs (<http://features.stanford.edu/>). He leads one of seven NIH-supported National Centers for Biomedical Computation, focusing on physics-based simulation of biological structures (<http://simbios.stanford.edu/>). Dr. Altman holds an A.B. from Harvard College, and M.D. from Stanford Medical School, and a Ph.D. in Medical Information Sciences from Stanford. He received the U.S. Presidential Early Career Award for Scientists and Engineers and a National Science Foundation CAREER Award. He is a fellow of the American College of Physicians, the American College of Medical Informatics, and the American Institute of Medical and Biological Engineering. He is a member of the Institute of Medicine of the National Academies. He is a past-president, founding board member, and a Fellow of the International Society for Computational Biology, and the President-Elect of the American Society for Clinical Pharmacology & Therapeutics. He currently chairs the Science Board advising the FDA Commissioner. He is an organizer of the annual Pacific Symposium on Biocomputing (<http://psb.stanford.edu/>), and a founder of Personalis, Inc. He won the Stanford Medical School graduate teaching award in 2000.

Speech Title: On Content, Discussions, Opinions, and Deliberative Participation over Social Media Systems

Speaker: Dr. S. Felix Wu

Professor, University of California-Davis



Biographical Sketch: Prof. S. Felix Wu has been doing “experimental” system research, i.e., building prototype systems to justify and validate novel architectural concepts. Since 1995, he and his students/postdocs have built many experimental systems in the areas of fault tolerant network, IPSec/VPN security policy, attack source tracing, wireless network security, intrusion detection and response, visual information analytics, and, more recently, future Internet design. An article titled “Networking: Four ways to reinvent the Internet” published in Nature 463 (February 3rd, 2010, by Katharine Gammon) provided a brief but very nice cover about his primary thought on a Social-network-based future Internet architecture. During the past seven years, he has been pretending (and hoping) to know a little bit more about humanity science so he can claim that he is working on multidisciplinary research. He strongly believes that thoroughly considering the factor of human relationships is necessary for any IT innovation. Therefore, his primary research objective, before he retires, is to help and contribute to the information technology advancement that would truly help our human society. As an initial step, he recently released the SINCERE (Social Interactive Networking and Conversation Entropy Ranking Engine, sponsored by NSF) search engine under <http://www.sincere.se>, which is trying to help our Internet society to discover “interesting/unusual” discussions. Felix received his BS from Tunghai University, Taiwan, in 1985, both MS and PhD from Columbia University in 1989 and 1995, all in Computer Science. He has about 120+ academic publications. He is currently a Professor with the Computer Science department at UC Davis.

Abstract: Social Media is changing many different aspects of our lives. By participating in online discussions, people exchange opinions on various topics or contents, shape their stances, and gradually build their own characteristics. In this talk, we will present a framework for identifying online user characteristics and understanding the formation of user deliberation and bias in online newsgroups. Under the SINCERE.se (Social Interactive Networks: Conversation Entropy Ranking Engine), we have designed a dynamic user like graph model to recognize user deliberation and bias automatically in online newsgroups. We evaluated our identification results with linguistic features and implemented this model under SINCERE as a real-time service. By applying this model to large online newsgroups, we study the influence of early discussion context on the formation of user characteristics. Our conclusion is that the formation of user deliberation and bias is a product of situations, not simply dispositions: confronting disagreement in unfamiliar circumstances promotes more consideration of different opinions, while recurring conflict in familiar circumstances evokes close-minded behavior and bias. Based on this

observation, we leveraged a supervised learning model to predict user deliberation and bias at their early life-stage. Our results show that knowing only the first three months of users' interaction data generates an F1 accuracy level of around 70% in predicting user deliberation and bias in online newsgroups.

Speech Title: Cloud-Centric Assured Information Sharing

Dr. Bhavani Thuraisingham

Louis A. Beecherl, Jr. Distinguished Professor of Computer Science
Executive Director of the Cyber Security Research and Education Institute
The University of Texas at Dallas



Biographical Sketch: Dr. Bhavani Thuraisingham is the Louis A. Beecherl, Jr. Distinguished Professor of Computer Science and the Executive Director of the Cyber Security Research and Education Institute (CSI) at The University of Texas at Dallas. She is an elected Fellow of IEEE, the AAAS, the British Computer Society, and the SPDS (Society for Design and Process Science). She received several prestigious award including IEEE Computer Society's 1997 Technical Achievement Award for "outstanding and innovative contributions to secure data management", the 2010 ACM SIGSAC (Association for Computing Machinery, Special Interest Group on Security, Audit and Control) Outstanding Contributions Award for "seminal research contributions and leadership in data and applications security for over 25 years" and the SDPS Transformative Achievement Medal for her contributions to interdisciplinary research. She has unique experience working in commercial industry, research laboratory, US government and academia and her 30+ year career includes research and development, technology transfer, product development, program management, and consulting for the federal government. Her work has resulted in 100+ journal articles, 200+ conference papers, 100+ keynote and invited talks, six US patents (two pending) and twelve books.

Abstract: This presentation will describe our research and development efforts in assured cloud computing for the Air Force Office of Scientific Research. We have developed a secure cloud computing framework as well as multiple secure cloud query processing systems. Our framework uses Hadoop to store and retrieve large numbers of RDF triples by exploiting the cloud computing paradigm and we have developed a scheme to store RDF data in a Hadoop Distributed File System. We implemented XACML-based policy management and integrated it with our query processing strategies. For secure query processing with relational data we utilized the HIVE framework. More recently we have developed strategies for secure storage and query processing in a hybrid cloud. In particular, we have developed algorithms for query processing wherein user's local computing capability is

exploited alongside public cloud services to deliver an efficient and secure data management solution. We have also developed techniques for secure virtualization using the XEN hypervisor to host our cloud data managers as well as an RDF-based policy engine hosted on our cloud computing framework.

We have also developed demonstration systems with our European partners: Kings College, University of London and the University of Insubria Italy who are funded by EOARD (The European Office of Aerospace Research and Development). The first demonstration illustrates how information may be shared in our cloud, based on policies specified in XACML. In the second demonstration we have implemented a semantic web-based policy engine and will show how multiple social networks may share information on our cloud utilizing semantic web-based policies.

Speech Title: Twitter Analytics for Insider Trading Fraud Detection System

Dr. Ketty Gann

Northrop Grumman Information Systems



Abstract: Twitter analytics has been developed to process Twitter data at macro level for use in an insider trading detection system in order to establish normal trading patterns between daily stock price change and public sentiment. Two machine learning models, Support Vector Machine (SVM) and Decision Tree, are built based on annotated historical Twitter data and Stanford Sentiment140 Tweet corpus, respectively. The paper focuses on the discussions of polarized sentiment (positive and negative), comparison of SVM and Decision Tree models, Sentiment Key Performance Index (SKPI) and Daily Sentiment Index (DSI) and mood analysis. The results illustrate that Twitter SKPI and DSI are useful indexes to predict the future stock price movement on regular stock trading.

Biographical Sketch: Dr. Ketty Gann is a Senior Research Engineer at Northrop Grumman Corporation. She has 20+ years professional experience in researching, designing, analyzing system prototypes and operational systems. She is specializing in ontology engineering, metadata management and text analytics including natural language processing (NLP), machine translation (MT), information extraction (IE), sentiment analysis, information retrieval (IR) and knowledge management (KM). Ketty Gann is the technical lead for several large-scale and high profile defense contracts and a certified knowledge manager. Before joining Northrop Grumman, Ketty was an Associate Tech Fellow at Boeing company and Language Technology Manager at Booz Allen & Hamilton Inc. Ketty received her M.S. and Ph.D in computational linguistics from Georgetown University.

Speech Title: Augmenting Text Analysis: Text Insights and Social Media

Dr. Seth Howell, Dr. Ajith Warriar, Dr. Damir Herman

Senior Data Scientists

Ayasdi, Inc.



Abstract: How do you process social media information? How do you operationalize the data? Topological data analysis augments conventional approaches to understanding social media. Traditional analytics is labor intensive; more data can lead to more complexity. Analysts and Data Scientists are required to process and understand the data. Insights are derived if the analytic team has correctly guessed the correct questions to ask on the data set. If additional questions are needed, a new query is formulated, and the process restarts. Topological data analysis changes the process. A massive number of algorithms are applied to the data to create a data shape. This shape provides the analyst with insights on both the data and the queries that apply to the data set. For social media, the problem of using the data can be compounded because the processing of social media can result in a high dimensional data set. Phrases can be condensed, words counted, sentiment measured, and scoring systems established. Given this wealth of information, topological data analysis provides a shape of the data allowing identification of common themes, trends, or emerging ideas. Topological data analysis leverages high-dimensionality in a data set because all variables are represented in the shape of the data. More data leads to more refined insights. Sub-groups are found based on all the data characteristics instead of being discovered through a set of queries. A query-free approach can discover answers to important, but unknown, questions. Putting a shape to social media involves confronting questions on data ingestion, distance between themes, capturing themes, and discovering emergent concepts. This talk will present several use cases to show the application of the Ayasdi Platform to Twitter data across different contexts. We will look at an example of segmentation of groups using tweets from a political perspective, business, i.e. stock prices, and a military context. Each use case will cover the processing of social media, the resulting shape of the data, and the interpretation of the data shape.

Biographical Sketch: **Dr. Seth Howell** is a Senior Data Scientist at Ayasdi focusing on financial applications and text insights. He received a Ph.D. in Washington University in St. Louis in mathematics (Several Complex Variables). Dr. Howell then worked for the Department of Defense for the next eight and a half years. His primary assignments were as an Operations Research and Systems Analyst and Mathematical Statistician. **Dr. Howell** spent three years at the Joint Improvised Explosive Device Defeat Organization (JIEDDO) in both analytic and management positions. **Dr. Ajith Warriar** is a Senior Data Scientist at Ayasdi. He received his PhD in Neuroscience from the University of California, Davis and had performed

postdoctoral work in labs at the University of California, San Francisco and University of California, Berkeley. Dr. Warriar has specialties in bioinformatics, neuroscience, and algorithm development. As a neuroscientist, he was primarily interested in factors affecting synaptic transmission and plasticity processes in the retina, the striatum, and the hippocampus and had devised novel ways of studying these processes. At Ayasdi, Dr. Warriar has used his biological and programming expertise to develop pipelines to process and decode NGS and text data into meaningful insights. **Dr. Damir Herman** is a Senior Data Scientist at Ayasdi. His background is theoretical physics, computational biology and translational research in hematology and oncology. While Life Sciences and Healthcare have been his main focus at Ayasdi, Damir happily works on all other problems that come his way such as digital signal processing, transportation, equipment malfunctioning, manufacturing QC, financial fraud, customer segmentation, text and social media analysis.

Speech Title: Big Data and Semantic Web Meet Applied Ontology

Dr. Ram D. Sriram

Chief, Software and Systems Division
National Institute of Standards and Technology



Biographical Sketch: Dr. Ram D. Sriram is currently the chief of the Software and Systems Division, Information Technology Laboratory, at the National Institute of Standards and Technology. Before joining the Software and Systems Division, Sriram was the leader of the Design and Process group in the Manufacturing Systems Integration Division, Manufacturing Engineering Laboratory, where he conducted research on standards for interoperability of computer-aided design systems. He was also the manager of the Sustainable Manufacturing Program. Prior to joining NIST, he was on the engineering faculty (1986-1994) at the Massachusetts Institute of Technology (MIT) and was instrumental in setting up the Intelligent Engineering Systems Laboratory. At MIT, Sriram initiated the MIT-DICE project, which was one of the pioneering projects in collaborative engineering and documented in the book entitled *Distributed and Integrative Collaborative Engineering Design*, Sarven Publishers, 2002. Sriram has extensive experience in developing knowledge-based expert systems, natural language interfaces, object-oriented software development, life-cycle product and process models, geometrical modelers, object-oriented databases for industrial applications, health care informatics, bioinformatics, and bioimaging. He has consulted for several leading corporations all over the world. His client list (during his tenure at MIT) include Boeing, GE, NTT Data (Japan), NASA, Xerox Corporation, United Technologies, IIC (Spain). Sriram has co-authored or authored nearly [250 papers, books, and reports, including several books](#). He has published in a wide range of journals in engineering, computer science, and health care fields. The papers he and his group wrote have won many recognitions and awards, including many best paper awards and most cited paper awards. In 1989, he

was awarded a Presidential Young Investigators Award from the National Science Foundation, U.S.A. In 2011, Sriram received the ASME Design Automation Award for "pioneering contributions to design automation, particularly in developing new technologies to support distributed and integrated collaborative design."

Sriram was a founding co-editor of the International Journal for AI in Engineering (1986). Sriram served on the Executive Committee of the ASME's Computers in Engineering division for six years, including as its chair. From 2012 onwards, Sriram has been serving as a co-chair of NITRD's Software Design and Productivity group. He and his co-chairs have been developing a strategic direction for the group, including organizing a workshop to understand the software related problems in manufacturing and health care sectors in September 2013. Sriram is a fellow of the American Society of Mechanical Engineers, a fellow of the American Association for the Advancement of Science, a senior member of the Institute of Electrical and Electronics Engineers, a member (life) of the Association for Computing Machinery, a member of the Association for the Advancement of Artificial Intelligence, and a member (life) of the Washington Academy of Sciences.

Abstract: Since the beginnings of the Semantic Web, ontologies have played key roles in the design and deployment of new semantic technologies. Yet over the years, the level of collaboration between the Semantic Web and Applied Ontology communities has been much less than expected. Within Big Data applications, ontologies appear to have had little impact. On the one hand, the Semantic Web, Linked Data, and Big Data communities can bring a wide array of real problems (such as performance and scalability challenges and the variety problem in Big Data) and technologies (automated reasoning tools) that can make use of ontologies. On the other hand, the Applied Ontology community can bring a large body of common reusable content (ontologies) and ontological analysis techniques. Identifying and overcoming ontology engineering bottlenecks is critical for all communities. The primary goal of the Ontology Summit

2014 – the 9th in a series – is to provide a platform and opportunity for building bridges between the Semantic Web, Linked Data, Big Data, and Applied Ontology communities. The Summit activities brought together insights and methods from these different communities, synthesize new insights, and disseminate knowledge across field boundaries.

The summit, launched on January 16, 2014, comprised of a series of events which spanned three-and-a-half (3.5) months. Activities included asynchronous virtual discourse (over our archived mailing lists and wikis) and weekly virtual panel discussions (which were held every Thursday) and culminated in a 2-day face-to-face workshop/symposium held on April 28 & 29, 2014 at NCO_NITRD, Arlington, Virginia.

The various organizers and committee members and other details of the summit, as well as the proceedings and the entire body of work developed over this Summit, can be accessed from the summit website:

<http://ontolog.cim3.net/OntologySummit/2014/>.

The discourse was structured in four technical tracks: 1) Common re-usable semantic content; 2) Making use of ontologies: tools, services, and techniques; 3) Overcoming ontology engineering bottlenecks; and 4) Tackling the variety problem in big data. Additionally, a one-weekend international Hackathon was also held. The Hackathon comprised of six (6) projects, with 47 participants from 11 countries, and explored hacking solutions that spanned both paradigm and technology gaps between the Big Data, Semantic Web, Applied Ontology domains. At the end of the face-to-face symposium a Communiqué which summarizes the key issues raised and discussed during the summit will be adopted and released. In this presentation we will provide a summary of the communiqué, along with other important insights discussed at the summit symposium. (Tentative outline of) The Communiqué is accessible via the URL:

http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2014_Communique.

Speech Title: Beyond Map Reduce: The Next Generation of Big Data Analytics

Speaker: Stephen Turner
Enterprise Consultant
HAMR by ET International, Inc.



Abstract: ET International (ETI) is introducing a way to keep the best aspects of MapReduce and its popular implementation, Apache Hadoop, while reducing the number of add-on tools needed to make it relevant for commercial application. ETI's novel multisource analytics product "HAMR" runs both batch and realtime streaming. It complements the current paradigm and accommodates the next generation of systems that will begin to render Hadoop MapReduce – as we know it – obsolete. Given that Moore's Law has an 18-month cycle time, it is urgent that information systems professionals budget and plan now for future generations. MapReduce does what it was intended to do: store and process large datasets in batches, inexpensively. However, many organizations that implemented Hadoop have experienced unexpected challenges, in part because they want to do more than MapReduce was designed to do. In response, an ecosystem emerged that continually introduces additional tools to help overcome some of those challenges. Unfortunately, this has only made implementations more complex and daunting, giving rise to the need for the simpler toolset offered by HAMR.

Biographical Sketch: Stephen Turner works with enterprises that are grappling with Big Data. Problems include data cleansing / master data management, contract data mining and fraud detection. He is leading the commercialization of ET International's Big Data analytics engine HAMR, which addresses these problems. ETI is a spinoff of the University of Delaware and got its start as a contractor on the U.S. Government's most secure high performance computing initiatives. HAMR processes huge volumes and variety of information to enable businesses to gain more insights on a near real-time basis. Turner has a MS in Information Systems Management and has recently published papers on Big Data, Security and Cloud Computing with AABRI's *Journal of Technology Research*. Turner's research on Predictive Coding and Information Governance was also published by the Pennsylvania Bar Institute in 2013. He is co-author of the book *NET Value: How the Digital Culture is Changing Your Value Proposition* and specializes in translating technology benefits into business value. He began his career in San Francisco, and has worked with clients such as Adobe, Yahoo, Disney, PeopleSoft, HP, McAfee and the world's first quantum computer maker, D-Wave.

Panel on Social Science and Big Data

Chair: Dr. Ashok Krishnamurthy

Deputy Director of RENCI

Renaissance Computing Institute

University of North Carolina, Chapel Hill



Biographical Sketch: Dr. Ashok Krishnamurthy became deputy director of RENCI, the Renaissance Computing Institute at the University of North Carolina, Chapel Hill, in February 2013. His work focuses on managing and enhancing research partnerships with faculty at UNC Chapel Hill, Duke University and North Carolina State University, building relationships between RENCI and Triangle area businesses, and leading efforts to bring new federal research funding to RENCI and its partner institutions. Previously, he was director of research and scientific development at the Ohio Supercomputer Center (OSC) and an associate professor in the computer and electrical engineering department at The Ohio State University. He also served as OSC's co-interim executive director from September 2009 to August 2012. Krishnamurthy played a crucial role in establishing OSC's successful industrial outreach initiative called Blue Collar Computing. The program targets small and medium-sized businesses that lack high performance computing resources, providing them with the training, expertise and advanced technology tools they need to enhance their companies' competitiveness. He also helped develop and deploy cyberinfrastructure that allows researchers to easily access and use computing and storage resources at OSC. Before his work at OSC and OSU, he served as the academic lead for the U.S. Department of Defense (DoD) High Performance Computing Modernization Program in the Integrated Modeling and Test area. He has designed and provided numerous training courses for DoD User Groups on all aspects of the MATLAB programming language. Krishnamurthy holds Ph.D. and master's degrees in electrical engineering from the University of Florida and a bachelor's degree in electrical engineering from the Indian Institute of Technology. His areas of expertise include high performance computing (HPC), cyberinfrastructure, data exploitation, HPC in industry, American competitiveness initiatives, signal and image processing and software development.

Speech Title: Randomized matrix algorithms and large-scale scientific data analysis

Speaker: Dr. Michael Mahoney

ICSI and UC Berkeley



Abstract: Matrix problems are ubiquitous in many large-scale scientific data analysis applications; and in recent years randomization has proved to be a valuable resource for the design of better algorithms for many of these problems. Depending on the situation, better might mean faster in worst-case theory, faster in high-quality numerical implementation, e.g., in RAM or in parallel and distributed environments, or more useful for downstream domain scientists. This talk will describe the theory underlying randomized algorithms for matrix problems such as least-squares regression and low-rank matrix approximation; and it will describe the use of these algorithms in large-scale scientific data analysis and numerical computing applications. Examples of the former include

the use of interpretable CUR matrix decompositions to extract informative markers from DNA single nucleotide polymorphism data as well as informative wavelength regions in astronomical galaxy spectra data; and examples of the latter include a randomized algorithm that beats Lapack on dense overconstrained least-squares problems for data in RAM, and a randomized algorithm to solve the least absolute deviations problem on a terabyte of distributed data.

Biographical Sketch: Michael Mahoney is at the University of California at Berkeley in the Department of Statistics and at the International Computer Science Institute. He works on algorithmic and statistical aspects of modern large-scale data analysis. Much of his recent research has focused on large-scale machine learning, including randomized matrix algorithms and randomized numerical linear algebra, geometric network analysis tools for structure extraction in large informatics graphs, scalable implicit regularization methods, and applications in genetics, astronomy, medical imaging, social network analysis, and internet data analysis. He received his PhD from Yale University with a dissertation in computational statistical mechanics, and he has worked and taught at Yale University in the mathematics department, at Yahoo Research, and at Stanford University in the mathematics department. Among other things, he is on the national advisory committee of the Statistical and Applied Mathematical Sciences Institute (SAMSI), he was on the National Research Council's Committee on the Analysis of Massive Data, he runs the biennial MMDS Workshops on Algorithms for Modern Massive Data Sets, and he spent fall 2013 at UC Berkeley co-organizing the Simons Foundation's program on the Theoretical Foundations of Big Data Analysis.

Tutorial Title: Deep Learning for Natural Language Processing

Presenter: Richard Socher
Department of Computer Science
Stanford University



Biographical Sketch: Richard Socher is at Stanford working with Chris Manning and Andrew Ng. His research interests are machine learning for NLP and vision. He is interested in developing new deep learning models that learn useful features, capture compositional structure in multiple modalities and perform well across different tasks. He was awarded the 2011 Yahoo! Key Scientific Challenges Award, the Distinguished Application Paper Award at ICML 2011, a Microsoft Research PhD Fellowship in 2012 and a 2013 "Magic Grant" from the Brown Institute for Media Innovation.

Abstract: Machine learning is everywhere in today's NLP, but by and large machine learning amounts to numerical optimization of weights for human designed representations and features. The goal of deep learning is to explore how computers can take advantage of data to develop features and representations appropriate for complex interpretation tasks. This tutorial aims to cover the basic motivation, ideas, models and learning algorithms in deep learning for natural language processing. Recently, these methods have been shown to perform very well on various NLP tasks such as language modeling, POS tagging, named entity recognition, sentiment analysis and paraphrase detection, among others. The most attractive quality of these techniques is that they can perform well without any external hand-designed resources or time-intensive feature engineering. Despite these advantages, many researchers in NLP are not familiar with these methods. Our focus is on insight and understanding, using graphical illustrations and simple, intuitive derivations. The goal of the tutorial is to make the inner workings of these techniques transparent, intuitive and their results interpretable, rather than black boxes labeled "magic here". The first part of the tutorial presents the basics of neural networks, neural word vectors, several simple models based on local windows and the math and algorithms of training via backpropagation. In this section applications include language modeling and POS tagging. In the second section we present recursive neural networks which can learn structured tree outputs as well as vector representations for phrases and sentences. We cover both equations as well as applications. We show how training can be achieved by a modified version of the backpropagation algorithm introduced before. These modifications allow the algorithm to work on tree structures. Applications include sentiment analysis and paraphrase detection. We also draw connections to recent work in semantic compositionality in vector spaces. The principle goal, again, is to make these methods appear intuitive

and interpretable rather than mathematically confusing. By this point in the tutorial, the audience members should have a clear understanding of how to build a deep learning system for word-, sentence- and document-level tasks. The last part of the tutorial gives a general overview of the different applications of deep learning in NLP, including bag of words models. We will provide a discussion of NLP-oriented issues in modeling, interpretation, representational power, and optimization.

Tutorial Title: Tools for Scaling Genomics Analysis

Dr. Uri Laserson
Data Scientist
Cloudera, Inc.



Biographical Sketch: Dr. Uri Laserson is a data scientist at Cloudera. Previously, he received his PhD from MIT developing applications of high-throughput DNA sequencing to immunology. During that time, he co-founded Good Start Genetics, a next-generation diagnostics company focused on genetic carrier screening. In 2012 he was selected to Forbes's list of 30 under 30.

Abstract: The advent of next-generation DNA sequencing technologies is poised to revolutionize the way life sciences research is practiced. These new technologies are scaling significantly faster than Moore's law, and promise to catapult life sciences research and the biotech industry into the realm of big data. However, bioinformatics and data management in the life sciences has been slow to adopt the latest big data technologies pioneered by the internet industry (e.g., Google and Facebook), in part because these tools are only beginning to become necessary today. In this talk, we will review several ways in which distributed computing tools (e.g., the Hadoop ecosystem) can be used to significantly advance the state-of-the-art in life sciences research, including: scaling genome-wide association studies to find connections between your genes and your traits and large-scale data integration of the large number of public databases.

Tutorial Title: Toward Integrated Security in Cloud Based Data and Compute Infrastructures

Michael Shoffner

Research Scientist

Renaissance Computing Institute



Biographical Sketch: Michael Shoffner is a member of the Informatics group at RENCI, a University of North Carolina at Chapel Hill based research institute, and adjunct faculty in the UNC School of Information and Library Science (SILS). Michael's primary area of practice is data and metadata infrastructure and cybersecurity, with a focus on medical and life sciences data. Previously he was a technical architect, a member of an emerging technologies strategy group, and co-founder of an early Internet technology consultancy with clients in the biomedical and education spaces.

Abstract: Cloud based systems are the subject of great interest in the computing community due to their numerous advantages over traditional systems for many applications. However, for use with data subject to privacy concerns, such as medical data, adoption of this style of computing is hampered by security and privacy concerns. RENCI, in partnership with a number of other collaborators, is prototyping an integrated technology architecture to address these issues. This architecture is a fusion of Software Defined Networking (SDN), data management middleware, hypervisor based security, and endpoint security containers on cloud based virtual machines that run scientific workflows. The system is currently in the prototype stage of development. This tutorial will step through how to set up this prototype infrastructure, with background on the constituent technologies involved, and a focus on how the general approach can inform architecting analogous solutions in other contexts.

Tutorial Title: Behavioral Data Mining and Social Network Analysis in Massive Online Games

Dr. Muhammad Aurangzeb Ahmad and Jaideep Srivastava

Research Scientist and Professor

Department of Computer Science

University of Minnesota

Abstract: Massive Online Games (MOGs) refer to massive online persistent environments (World of Warcraft, EVE Online, EverQuest etc) shared by millions of people. In general these environments are characterized by a rich array of activities and social interactions with a wide array of behaviors e.g., cooperation, trade, quest, deceit, mentoring etc. Such environments allow one to study human behavior at a level of granularity where it was not possible to do so previously. Given the challenges associated with analyzing this type of data traditional techniques in data mining and social network analysis have to be extended with insights from the social sciences. The tutorial will cover predictive and generative models in the study of MOGs. Additionally we will cover some SNA techniques which are more appropriate for MOGs given the multi-dimensionality of the data (P*/ERGM Models, IR Based Network Analysis, Hypergraph based Techniques, Coextensive Social Networks etc). Based on our published work in this area we also describe the various ways in which MOGs exhibit similarities to the real world e.g., economic behaviors, clandestine behaviors, mentoring etc). Lastly we describe the scope and limitations of analysis of MOGs based on limitations in data collection, availability and ethical concerns. An overview of commercial applications is also given.

Biographical Sketch: Dr. Muhammad Aurangzeb Ahmad is a research scientist at the University of Minnesota. He has authored over 30 research papers in the area of social informatics, data mining. He completed bachelors in Computer Science from the Rochester Institute of Technology, and master and PhD in Computer Science from University of Minnesota. He is also part of the Virtual Worlds Observatory (VWO) project, the leading project on the analysis of human behavior and socialization patterns in online virtual environments. He was also a research assistant at the Minnesota Population Center where he worked on the application of machine learning to census data. In addition to his post- doc work he is currently also a data scientist at the social analytics start up Ninja Metrics. **Dr. Jaideep Srivastava** is Professor of Computer Science & Engineering at the University of Minnesota, where he directs a laboratory focusing on research in Web Mining, Social Media Analytics, and Health Analytics. He has authored over 275 papers, and his research has been supported by government agencies, including NSF, NASA, ARDA, IARPA, NIH, CDC, US Army, US Air Force, and MNDOT; and industries, including IBM, United Technologies, Eaton, Honeywell, Cargill, and Huawei Telecom. He has an active collaboration with Allina's Center for Healthcare Innovation, where he is a Distinguished Fellow. He has lead a data mining team at Amazon.com, and built a data analytics department at Yodlee. He holds distinguished professorships at Heilongjiang University and Wuhan University, China. Dr. Srivastava has BS from Indian Institute of Technology (IIT), Kanpur, India, and MS and PhD from University of California, Berkeley. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), and has been an IEEE Distinguished Visitor. He has given over 150 invited talks in over 30 countries, including more than a dozen keynote addresses at major conferences. He is a co-founder and President of Ninja Metrics, a social analytics company, whose goal is bring his research in this area into the commercial domain.

Tutorial title: Introduction to Revolution R Enterprise

Dr. Mario Inchiosa and Joseph Rickert

US Chief Scientists

Revolution Analytics, Inc.

Abstract: This tutorial will provide a detailed introduction to Revolution R Enterprise, highlighting its big data advanced analytics features built upon the concepts of “compute context” (e.g. workstations, servers, and Hadoop clusters), “data source” (local and distributed file systems and database tables), and scalable algorithms. We will then see how it all comes together by taking a tour through an end-to-end predictive analytics session, starting with data exploration, continuing through data preparation and model building, and finishing with prediction.



Biographical Sketch: Mario Inchiosa’s passion for data science and high performance computing drives his work at Revolution Analytics, where he focuses on delivering parallelized, scalable advanced analytics integrated with the R language. Previously, Mario served as Analytics Architect in IBM’s Big Data organization and US Chief Scientist in Netezza Labs, advancing Hadoop and SQL-based parallel data analytic platforms. Their success led to Netezza’s acquisition by IBM. Mario also served as US Chief Science Officer at NuTech Solutions, a computer science consultancy specializing in simulation, optimization, and data mining, and Senior Scientist at BiosGroup, a complexity science spin-off of the Santa Fe Institute. Dr. Inchiosa holds Bachelors, Masters, and PhD degrees from Harvard University. He has been awarded four patents and has published over 30 research papers, earning Publication of the Year and Open Literature Publication Excellence awards.



Biographical Sketch: Joseph Rickert is a Data Scientist and Community Manager at Revolution Analytics with a passion for analyzing data and teaching people about R. He is a regular contributor to the Revolutions blog and an organizer of the Bay Area R Users Group. Joseph is frequently invited to speak on the topic of R and big data. He has presented tutorials including the O'Reilly [Strata conference](#), [Strata/ Hadoop World](#), [DataWeek](#), [PACE](#). Joseph has worked for a number of Silicon Valley start-ups and has experience building statistical models in industries as diverse as local area networks and healthcare. Joseph holds graduate degrees in both the Humanities and Statistics. He taught statistics briefly at SJSU.

Poster Session P1

May 28, 2014

1. **Tweet This or Not? Predicting Twitter Users Actions from Sentiment** [[PDF](#)]
Jalal Mahmud (IBM Research - Almaden)
2. **The Application of Gamification Mechanics on Social Media Platforms for Creative Crowdsourcing** [[PDF](#)]
Anja Solf (Technische Universität Ilmenau), Daniel Schultheiss (Technische Universität Ilmenau), Saskia Stäudtner (Technische Universität Ilmenau)
3. **Using Specialized Support Vector Machines to Detect Helpful and Unhelpful Product Reviews** [[PDF](#)]
Scott Bolter, Teng Moh (San Jose State University)
4. **The Relationship Between Practicing Makes Perfect and Game-based Learning in Mobile Device** [[PDF](#)]
Ming-Hung Lin (Chung Yuan Christian University), Chien-Hung Lai (Chung Yuan Christian University), Wei-Ting Tseng (Chung Yuan Christian University), Bin-Shyan Jong (Chung Yuan Christian University)
5. **Space, Time, and Hurricanes: Investigating the Spatiotemporal Relationship among Social Media Use, Donations, and Disasters** [[PDF](#)]
Jared Sylvester (University of Maryland), John Healey (University of Maryland), Chen Wang (University of Maryland), William Rand (University of Maryland)
6. **Understanding Types of Users on Twitter** [[PDF](#)]
Mohammad Moeen-uddin (Lahore University of Management and Sciences), Muhammad Imran (Qatar Computing Research Institute), Hassan Sajjad (Qatar Computing Research Insti)
7. **Towards Ability Based Adaptive Learning Technology Design** [[PDF](#)]
Gahangir Hossain, Habibah Khan, Iqbal Hossain, Abdul Halim Khan (University Of Memphis)
8. **Political Polarization over Global Warming: Analyzing Twitter Data on Climate Change** [[PDF](#)]
Alireza Hajibagheri (Department of EECS, University of Central Florida), Gita Sukthankar (Department of EECS, University of Central Florida)
9. **Geographic Bias in Twitter Based Election Forecasting** [[PDF](#)]
Anoop Kumar (raytheon bbn technologies), Manish Gaurav (raytheon bbn technologies), David Stallard (Raytheon BBN Technologies), Amit Srivastava (Raytheon BBN Technologies), Scott Miller (Raytheon BBN Technologies)
10. **Transitioning from Management to Engineering** [[PDF](#)]
Ida Hashemi (CSUF), Yun Tian (CSUF), Ya-Fei Jia (Beijing University of Technology)
11. **An Energy Efficient Data Mining Scheme For Big Data Biodiversity Environment** [[PDF](#)]
Moh'D ALWADI (The University of Canberra), Girija Chetty (The University of Canberra)
12. **Towards Data Readiness Level For Structured Data** [[PDF](#)]
Yang Lu (North Carolina A&T State University), Xing Fang (North Carolina A&T State University)

13. **Scalable Real-Time Big Data IoT platform for Connected Cars** [\[PDF\]](#)
Dirk Van den Poel (Ghent University), Michiel Van Herwegen (Ghent University)
14. **A Study of Electric Vehicle Data Analytics** [\[PDF\]](#)
Vamshi Krishna Bolly (Purdue University), John Springer (Purdue University), Eric Dietz (Purdue University)
15. **Using Triangles to Improve Community Detection in Directed Networks** [\[PDF\]](#)
Christine Klymko (Lawrence Livermore National Laboratory), David Gleich (Purdue University), Tamara Kolda (Sandia National Laboratories)
16. **Query Expansion to Search Politically Relevant Tweets** [\[PDF\]](#)
Manish Gaurav (raytheon bbn technologies), Anoop Kumar (raytheon bbn technologies), Amit Srivastava (Raytheon BBN Technologies), Scott Miller (Raytheon BBN Technologies)
17. **Animation-Based Learning in Children Education: Animation-Based Learning in Children Education: An Emerging Paradigm for Children Instruction** [\[PDF\]](#)
Habibah Khan, Gahangir Hossain, AbdulHalim Khan, and Iqbal Hossain (The University of Memphis)
18. **Transitioning from Management to Engineering** [\[PDF\]](#)
Ida Hashemi (CSUF), Yun Tian (CSUF)

Poster Session P2
May 29, 2014

1. **BlogSphere - A Topical Map of the Entire Blogosphere** [\[PDF\]](#)
Philipp Berger (Hasso Plattner Institut), Patrick Hennig (Hasso Plattner Institut), Stephan Detje (Hasso Plattner Institut), Christoph Meinel (Hasso Plattner Institut), David Eickhoff (Hasso Plattner Institut), Daniel Taschik (Hasso Plattner Institut), Bjoern Wagner (Hasso Plattner Institut)
2. **Crowdsourcing Parking Lot Occupancy using a Mobile Phone Application** [\[PDF\]](#)
Erfan Davami (University of Central Florida), Gita Sukthankar (Department of EECS, University of Central Florida)
3. **Modeling Users' Topical Interests on Twitter: a Novel Framework** [\[PDF\]](#)
Mahmoud Khodier (Cairo University), Akram I. Salah (Cairo University), Abeer El-Korany (Cairo University)
4. **Can Hashtags Bear the Testimony of Personality? Predicting Personality from Hashtag Use** [\[PDF\]](#)
Hasan Al Maruf (Bangladesh University of Engineering and Technology), Jalal Mahmud (IBM Research), Mohammed Eunus Ali (Bangladesh University of Engineering and Technology)
5. **Causal Relationship between ICT Development and Portfolio Investment Network: Social Selection or Social Influence?** [\[PDF\]](#)
Shing Doong (ShuTe University)
6. **Big Data Compression Using Anamorphic Stretch Transform** [\[PDF\]](#)
Mohammad H. Asghari (UCLA), Bahram Jalali (UCLA)
7. **Linked Open Data Publication Strategies: Application in Networking Performance Measurement Data** [\[PDF\]](#)
Renan Souza (Federal University of Rio de Janeiro), Les Cottrell (SLAC National Accelerator Laboratory), Bebo White (SLAC National Accelerator Laboratory), M. Luiza Campos (Federal University of Rio de Janeiro), Marta Mattoso (Federal University of Rio de Janeiro)
8. **Big Data Analysis in a Message Oriented Framework** [\[PDF\]](#)
Arcot Rajasekar (UNC Chapel Hill), Howard Lander (UNC Chapel Hill)
9. **Big Software Deployments in a Big Enterprise Environment** [\[PDF\]](#)
Serg Mescheryakov (State Polytechnic University), Dmitry Shchemelinin (RingCentral Inc.)
10. **Obesity in Online Social Networks** [\[PDF\]](#)
Ahmad Albu-Shamah (North Carolina A&T State University)
11. **Design of a Versatile Inspection Mobile Robot with Drilling Module for Space Applications** [\[PDF\]](#)
Michal Ciszewski (AGH University of Science and Technology), Tomasz Buratowski (AGH University of Science and Technology), Mariusz Giergiel (AGH University of Science and Technology), Karol Seweryn (Polish Academy of Sciences, Space Research Centre)

12. Beyond MapReduce: The Next Generation of Big Data Analytics [\[PDF\]](#)

Stephen Turner (ET International, Inc.), Brian Heilig (ET International, Inc.), Rich Collier (ET International, Inc.), Long Zheng (University of Delaware)

13. Using ‘Big Data’ Understanding For Analysis and Performance Improvement of Public Transportation Systems in Istanbul [\[PDF\]](#)

Kevser Simsek (Bogazici University- IMM Belbim Inc), Ilgin Gokasar (Bogazici University)

14. A Novel Approach for Evaluating Semantic Similarity Measures [\[PDF\]](#)

Xing Fang (North Carolina A&T State University)

15. Towards a Model of Storage oriented NoSQL for an Ontology Database [\[PDF\]](#)

Yassine Drias (USTHB)

Poster Session P3

May 30, 2014

1. **Association Rules Mining: Application to Large-Scale Satisfiability** [\[PDF\]](#)
Habiba Drias, Youcef Djenouri (LRIA/USTHB)
2. **Prediction of Financial Time Series Using Hidden Markov Models** [\[PDF\]](#)
Lev Brailovski (Open University), Maya Herman (Open University)
3. **A Monograph on Data Mining Techniques for Email Forensics** [\[PDF\]](#)
Venkata Krishna Kota (CRL)
4. **Network Traffic Analysis of ZeroAccess Botnet** [\[PDF\]](#)
Shree Garg (Indian Institute of Technology), Anil Sarje (Indian Institute of Technology), Sateesh Peddoju (Indian Institute of Technology)
5. **Better Timing of Cyber Conflict** [\[PDF\]](#)
Elisabeth Paulson (Penn State University), Christopher Griffin (Penn State University)
6. **Jamming Attacks on VANETs** [\[PDF\]](#)
Noha Gamal El-din (Nile University, IS Dept.), Sherif Mousa (Nile University, IS Dept.), Ahmed Kosairy (Nile University, IS Dept.), Marianne Azer (National Telecommunication Institute / Nile University)
7. **An Approach to Obfuscate Password Based Authentication** [\[PDF\]](#)
Dipankar Dasgupta (The University of Memphis), Paul N. Balister (The University of Memphis), Sanjib K. Saha (The University of Memphis), Scott M. Binski (The University of Memphis)
8. **Computational Models to Capture Human Behavior in Cybersecurity Attacks** [\[PDF\]](#)
Faisal Quader (UMBC), Vandana Janeja (UMBC)
9. **Zoo CAPTCHA: Telling Computers and Humans Apart via Animal Image Classification** [\[PDF\]](#)
Darryl D'Souza (University of Louisville), Jacob Matchuny (University of Louisville), Roman Yampolskiy (University of Louisville)
10. **Malware Detection Using the Weak Estimator** [\[PDF\]](#)
Balkirat Kaur (North Carolina A&T State University)
11. **BLN-Gram-TF-ITF as a new Feature for Authorship Identification** [\[PDF\]](#)
Nawaf Ali (University of Louisville), Matthew Price (University of Louisville), Roman Yampolskiy (University of Louisville)
12. **The Target Security Breach: A Case Study** [\[PDF\]](#)
Stephen Tipton, Young B. Choi (Regent University)
13. **Securing BYOD Networks: Inherent Vulnerabilities and Emerging Feasible Technologies** [\[PDF\]](#)
Arthur B. Hernandez, Young B. Choi (Regent University)

EXHIBITOR DIRECTORY

Bay Storage Technology

4

6200 Stoneridge Mall Road #300
Pleasanton, CA 94588
Phone 855-576-9933
Email jbim@baystoragetechnology.com
Web www.baystoragetechnology.com

Bay Storage Technology offers a network-attached high performance storage solution with extremely low latency (under 10 microseconds). It is scalable, cost effective and can be delivered as Software on customer provided hardware or as an integrated Appliance.

FiU Online

#5

11200 SW 8th Street
Miami, FL 33174
Phone 305-348-4673
Email ijaubert@fiu.edu
Web www.fiuonline.edu

Florida International University – fully online Masters Program.

HAMR by ET International, Inc.

3

100 White Clay Center Drive
Newark, DE 19711
Phone 302-738-1438
Fax 302-738-1436
Email sturner@etinternational.com
Web www.hamr-eti.com

HAMR is a software **analytics engine** that offers a unique approach to solving a range of problems arising because of Big Data. It provides access to a variety of data sources. One intuitive interface enables data scientists and programmers to be more productive in streaming, real-time and batch processing modes.

Maxeler Technologies

#2

2225 E. Bayshore Road
Palo Alto, CA 94303
Phone 650-320-1614
Email info@maxeler.com
Web www.maxeler.com

We combine maximum performance computing consulting with our dataflow computing technology to drive faster deployment of next-generation algorithms and lasting competitive advantage.

VisualThreat Inc.

#1

2901 Tasman Drive, Suite 107
Santa Clara, CA 95054
Phone 408-802-2383
Email wyan@visualthreat.com
Web www.visualthreat.com

VisualThreat is a leading mobile security vendor that offers the most comprehensive threat report and threat certification score (MobileThreatCert) for mobile apps. By providing RESTful API, online analytic service and mobile agent, VisualThreat enables you to discover mobile threats in a visual way.

EXHIBIT HOURS

Oak Lounge Foyer

Wednesday, May 28, 2014

12:00 pm – 5:00 pm

Thursday, May 29, 2014

9:00 am – 3:00 pm

Friday, May 30, 2014

9:00 am – 3:00 pm

**SocialCom - The Seventh ASE International Conference on
Social Computing**

**BigDataScience - The Third ASE International Conference
on Big Data Science and Computing**

Tsinghua University, Beijing, China, August 4-7, 2014

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

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

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

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

Harvard University, Cambridge, MA, USA,

December 15-19, 2014

**The Third ASE International Conference on Big Data
Science and Computing**

**The Seventh ASE International Conference on Social
Computing**

**The Fourth ASE International Conference on Cyber
Security**

San Jose, CA, USA, June 15-20, 2015