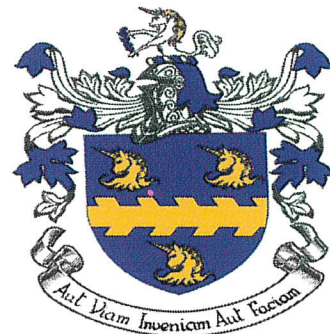*The Leir Retreat Center*
Ridgefield, Connecticut

Symposium on Data Science for Healthcare (DaSH)

"Data Sharing: the Art, Science, and Ethics"

The Leir Retreat Center
Wednesday, October 5, 2016
Thursday, October 6, 2016

NJIT
New Jersey's Science &
Technology University

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Table of Contents

Wednesday, October 5, 2016
Thursday, October 6, 2016

Dr. H.C. Henry J. Leir (1900 – 1998)

Henry J. Leir was known for charitable endeavors as the founder and chairman of three foundations- The Henry J. and Erna D. Leir Foundation, Inc. of Luxembourg, The Ridgefield Foundations, Inc. and The Leir Foundation, Inc.

Mr. Leir was born in Germany in the year 1900. In 1933 he and his wife, Erna, took refuge in Luxembourg. There he wrote a book entitled, *La Grande Compagnie de Colonisation.*

In the book he wrote about his visions for a new world where tyranny would be defeated; he further anticipated the Marshall Plan, the United Nations, renewal projects in underdeveloped countries and the unification of world economies.

Foreseeing the coming war, Henry Leir also predicted the invasion of Luxembourg, and so it was that he and his wife came to the United States in 1938. Mr. Leir then built a veritable empire in the minerals and metals industry. Over the years, he earned a worldwide reputation as an international industrialist, visionary, philanthropist, and patron of the arts.

The Leirs never forgot the brave, humanitarian country of Luxembourg. They therefore promoted its business and culture throughout their lives. Today, *a Leir Foundation, in Luxembourg, a Leir Luxembourg Program at Clark University and annual cultural events in Luxembourg* sponsored by the Leir Charitable Foundations all attest to their gratitude.

The wealth of the Leirs has been devoted primarily to those charitable purposes having the widest benefit to humankind. Following their precedent*, ten Leir Chairs* have been endowed at universities and hospitals in medical research. In addition Leir Chairs were created for *humanitarian studies* on the relief of poverty, famine, conflict resolution and international trade and development, along with a *Chair in Foreign Languages and Cultures.*

There are also *Leir fellowships at hospitals and museums, and Leir Scholars Abroad* in programs and universities. *Major medical research* programs have been established in their names, and *Leir halls and galleries* grace several museums.

At the heart of the Leirs' charity is the wish to promote understanding. Thus, the Leir Charitable Foundations supports an outstanding series of conferences on religious understanding sponsored by the *Center for Christian-Jewish Understanding at Sacred Heart University.* These conferences reflect the moral concepts which motivated the Leirs throughout their lives.

In the last years of his life, Henry Leir and his wife planned charitable uses of their 37-acre property in Ridgefield, Connecticut. At age 97, he planned and oversaw the construction of Leir House, focusing on programs for disadvantaged children of all races and religions, and on scientific and educational conferences. The first session at Leir House commenced two weeks after his death, and the series and other activities still continues to flourish to this day as he had hoped.
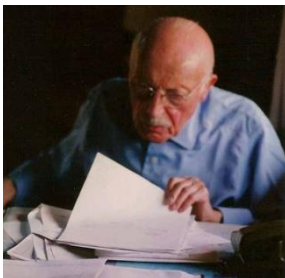
The Leir Charitable Foundations are committed to maintaining Mr. Leir's vision and philanthropy.

Founded in 1881, NJIT now offers 125 undergraduate and graduate degree programs in six specialized schools. These academic programs are fully accredited by the appropriate accrediting boards, commissions and associations such as Middle States, ABET, and NAAB. With an enrollment of more than 10,000 graduate and undergraduate students, NJIT offers small-campus intimacy with the resources of a major public research university.

The university is constantly developing new educational opportunities designed to meet the changing needs of the marketplace. These initiatives may take the form of new degrees programs, online learning programs, programs for pre-college youngsters, elementary and high school teachers, or professionals in the workforce. Opportunities abound for interdisciplinary learning at the intersections with the humanities, social sciences and economics, as we help our students to consider where technology is going to take us as a society locally, regionally, and globally.

NJIT is New Jersey's public technological research university. Research funding has grown dramatically over the last decades. In 2014, NJIT's research expenditures totaled more than $110 million. As of September 1, 2014, NJIT had 185 issued U.S. Patents and 133 pending U.S. patent applications. In FY2014, NJIT was awarded 15 new U.S. patents. NSF has ranked NJIT in the top 50 among colleges without a medical school for research expenditures and in the top 10 among universities whose main research is in engineering.

NJIT is located in the vibrant University Heights' district of downtown Newark, just 20 minutes from Manhattan by train.

NJIT Strategic Research Plan:
A seamless integration of collaborative synergy at all levels with faculty, staff and students will be developed. With the goal of significantly enhancing the external academic research funding, the NJIT Strategic Research Plan has been developed through the collective input from faculty, chairs and deans. The Strategic Research Plan identifies NJIT research enterprise into13 research focus groups that are organized in four research clusters. These research clusters with a target of 20 multidisciplinary research centers and 5-6 interdisciplinary research institutes are:

Life Sciences and Healthcare:
Life Sciences and Healthcare research cluster includes both basic and applied research in the areas of neuroscience, neural engineering, regenerative medicine, and point of care technologies. Research at NJIT includes understanding functions of brain and spinal cord under normal, injured, and diseased

states at molecular, cellular and functional levels through experimental, theoretical and computational methods. Regenerative medicine research deals with the process of replacing regenerating cells, tissues or organs to restore normal functions. The cluster also represents healthcare information system and managements involving primary care, hospitals and emergency care resources and protocols. The cluster includes research areas that are well aligned with NAE and NAS Grand challenges, Reverse Engineer the Brain, Engineer Better Medicine, and Tools for Scientific Discovery.

Data Science and Information Technology:

This research cluster includes the study and practice of extracting information and knowledge from data that can then be used for medical, finance, science and engineering applications. The focus groups in this cluster are engaged in bioinformatics, medical informatics, image processing and data mining, solar terrestrial physics, transportation, finance management, and life sciences and healthcare. Cyber security group studies how to design cyber systems so they are secure, and how to improve the cyber Information and Communications Technology (ICT) is shaping many aspects of the society and economy, and is evolving rapidly to provide access to unprecedented amounts of information anytime and anywhere from any device. Currently, there are 7.4 billion mobile devices connected throughout the globe with more than Global mobile data traffic reached 2.5 Exabytes per month.   It is expected that by the year 2050 more than 100 billion devices will be connected across the globe. The scope of the proposed cluster addresses NEA and NAS Grand Challenges including Secure Cyberspace, Virtual Reality, and Tools for Scientific Discovery.

Sustainable Systems:

This cluster represents research areas in urban ecology and sustainability, advanced materials and nanotechnologies, and smart manufacturing systems. Urban ecology and sustainability area emphasizes on sustainable infrastructure, ecological communities, and urban modeling and simulation. The area also focuses on water-energy nexus to investigate impact of ocean levels on the environments, and developing technologies for clean water and energy resources such as biofuel cells. The scope of Nanotechnology is to understand all science and engineering phenomenon at the minutest and fundamental level to develop technologies for environmental and pharmaceutical applications. The interdisciplinary group on engineered material and particulates focuses on technology development for preparation, processing and use of engineered particulate materials and their composites for a spectrum of applications. The research in manufacturing systems group is focused on developing new methods and technologies for design innovation and process automation. One of the specific emphases is to develop new processes, and tools for pharmaceutical manufacturing. The NAE and NAS Grand challenges covered by this research cluster incudes Solar Energy, Energy from Fusion, Clean Water and Urban Infrastructure.

Transdisciplinary Areas:

This cluster includes research focus groups and centers on mathematical sciences, transportation systems and science and technology impact in society. All of these three areas have broader multidisciplinary and interdisciplinary scope over large number of applications including life sciences, smart cities, and societal behavior and developing enabling technologies for market needs. The scope of the proposed cluster addresses NEA and NAS Grand Challenges including Urban Infrastructure, Smart Transportation, tools for Scientific Discovery, Advanced Personalized Learning.

Symposium on Data Science for Healthcare (DaSH)
Data Sharing: the Art, Science and Ethics

The Leir Retreat Center
Ridgefield, CT

2016 AGENDA

WEDNESDAY, OCTOBER 5, 2016

12:00 PM -01:00 PM	**LUNCH AT LEIR RETREAT CENTER**

01:00 PM -01:30 PM   **OPENING REMARKS AND INTRODUCTIONS**
- Jack Fong, MD, *Director, The Leir Retreat Center*
- Joel Bloom, PhD, *President, NJIT*
- Yi Chen, PhD, NJIT, *Event Organizer*

01:30 PM -02:45 PM   **Panel 1: "Big Data and Cancer Moonshot"**
- Reynold Panettieri, MD
- Ari D. Brooks, MD
- Bradford Hesse, PhD
- Jianying Hu, PhD
- Shridar Ganesan, MD, PhD

02:45 PM - 03:10 PM   **COFFEE BREAK**

03:10 PM - 04:55 PM   **Session Chair:** *Soon Ae Chun, PhD*

03:10 PM - 03:45 PM   *"Preventative Personalized and Precision Healthcare with Point-of-Care Technologies"*
- Atam P. Dhawan, PhD

03:45 PM - 04:20 PM   "*Observational Health Data Sciences and Informatics" (OHDSI)*
- George M. Hripcsak, MD

04:20 PM - 04:55 PM   *"Privacy in Healthcare Data Sharing: Challenges & Strategies"*
- Nan Zhang, PhD

Symposium on Data Science for Healthcare (DaSH)
Data Sharing: the Art, Science and Ethics

The Leir Retreat Center
Ridgefield, CT

2016 AGENDA

WEDNESDAY, OCTOBER 5, 2016

05:00 PM - 05:30 PM   **RECEPTION**

05:30 PM - 06:30 PM   **DINNER**

06:20 PM - 06:30 PM   **WELCOME REMARKS**
- Fadi Deek, PhD, *Provost and Senior V.P., NJIT*

06:30 PM - 08:00 PM   ***Breakout Discussions on Data Sharing:***
"The Art, Science and Ethics"

Leaders: *James Geller, PhD; Julia Stoyanovich, PhD; Dantong Yu, PhD.*

Group A: How to Encourage Data Sharing
Dantong Yu (Leader)
- *Cesar Bandera*
- *Réne Bastón*
- *Ari D. Brooks*
- *Ashok Chennuru*
- *Shridar Ganesan*
- *Shanthi Gopalakrishnan*
- *Ram Ramesh*
- *Jinhe Shi*
- *Heng Xu*

Group B: How to Facilitate Data Sharing
James Geller (Leader)
- *Judith Argon*
- *Reggie Caudill*
- *Soon Ae Chun*
- *Jack Fong*
- *James Gomes*
- *Zachary G. Ives*
- *Jianying Hu*
- *Vincent Oria*
- *Nan Zhang*

Group C: How to Ensure Responsible Data Sharing
Julia Stoyanovich (leader)
- *David G. Belanger*
- *Kathy Grise*
- *William Mcdermott*
- *Reynold Panettieri*
- *Cheickna Sylla*
- *Dimitri Theodoratos*
- *Fusheng Wang*
- *Hua Zhong*

08:00 PM   *ADJOURN AND DEPART FOR ETHAN ALLEN HOTEL*

Wednesday, October 5, 2016
Thursday, October 6, 2016

Symposium on Data Science for Healthcare (DaSH)
Data Sharing: the Art, Science and Ethics

The Leir Retreat Center
Ridgefield, CT

2016 AGENDA

THURSDAY, OCTOBER 6, 2016

08:00 AM - 08:30 AM	**ARRIVE TO LRC - BREAKFAST**

08:30 AM -08:45 AM **INTRODUCTIONS**
- Reggie J. Caudill, PhD, *Prof. & Dean, Tuchman School of Management, NJIT*

08:45 AM - 10:35 AM **Session Chair:** *Vincent Oria, PhD*

08:45 AM -09:20 AM **"Linking data, Linking science"**
- Zachary G. Ives, PhD

09:20 AM -10:35 AM **Panel 2: "Big Data and U.S. Health Reform"**
- Kathy L. Grise
- Ashok Chennuru
- Jack Fong, MD
- James Gomes, JD
- Ram Ramesh, PhD

10:35 AM -11:00 AM **COFFEE BREAK**

11:00 AM -11:15 AM **"Introduction to NSF NE Big Data Hub"**
- Réne Bastón

11:15 AM - 12:30 PM **"Summary of Breakout Discussions"**
- Yi Chen, PhD
- James Geller, PhD
- Dantong Yu, PhD
- Julia Stoyanovich, PhD

12:30 PM - 01:30 PM **LUNCH, CLOSING REMARKS AND ADJOURN**

Wednesday, October 5, 2016
Thursday, October 6, 2016

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Conference Organization and Attendees

Program Chair:

- Yi Chen, PhD – Henry J. Leir Chair in Healthcare, Co-director of Center for Big Data Analytics, Associate Professor, Martin Tuchman School of Management, NJIT

Advisory Board:

- Susan Davidson, PhD – Founding Co-Director, Weiss Professor, Computer & Info. Science, UPenn

- Kathy L. Grise – Senior Program Director of Big Data Initiatives, Future Directions, IEEE

- Bradford Hesse, PhD – Chief, Health Communication & Informatics Research, National Cancer Institute

- Cheickna Sylla, PhD – Professor, Martin Tuchman School of Management, NJIT

DaSH 2016 Attendees:

- Judith Argon – Chief Administrative Officer, Research, Geisinger Health System

- Cesar Bandera, PhD – Assistant Professor of Entrepreneurship, Martin Tuchman School of Management, NJIT

- René Bastón – Executive Director, Northeast Big Data Innovation Hub, Columbia University

- David G. Belanger, PhD , Senior Research Fellow, Stevens Institute of Technology

- Joel Bloom, EdD – President, NJIT

- Ari D. Brooks, MD - Dir. of Integrated Breast Center, Pennsylvania Hospital, Dir. of Endocrine Surgery, Prof. of Clinical Surgery, UPenn

- Reggie Caudill, PHD – Professor and Dean, Tuchman School of Management, NJIT

- Yi Chen, PhD – Henry J. Leir Chair in Healthcare, Co-director of Center for Big Data Analytics, Associate Professor, Martin Tuchman School of Management, NJIT

- Ashok Chennuru – Staff VP, Provider & Clinical Systems, Anthem, Inc.

- Soon Ae Chun, PhD – Professor, School of Business, The City Univ. of New York

- Fadi Deek, PhD – Provost and Senior Executive Vice President, NJIT

- Atam P. Dhawan, PhD – Vice Provost for Research and Development, Distinguished Professor, Electrical and Computer Engineering, NJIT

Wednesday, October 5, 2016
Thursday, October 6, 2016

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

- Jack Fong, MD – CM, MSc, FAAP, FRCP (C), Member of the Board of Directors, Leir Retreat Center

- Shridar Ganesan, MD, PhD – Medical Oncologist, Associate Director for Translational Science, Rutgers University

- James Geller, PhD – Professor of Computer Science, Associate Dean for Research of the Ying Wu College of Computing Sciences, NJIT

- James Gomes, JD – Director, Mosakowski Institute for Public Enterprise, Clark University

- Shanthi Gopalakrishnan, PhD – Professor, Martin Tuchman School of Management, NJIT

- Kathy L. Grise – Senior Program Director of Big Data Initiatives, Future Directions, IEEE

- Bradford Hesse, PhD – Chief, Health Communication & Informatics Research, National Cancer Institute

- George Hripcsak, MD – Vivian Beaumont Allen Professor & Chair, Dept of Biomedical Informatics, Columbia Univ.

- Jianying Hu, PhD - IEEE Fellow, IBM Distinguished Research Staff Member, Program Director for Comp Health, IBM

- Zachary G. Ives, PhD – Professor, Computer & Information Science, UPenn

- William Mcdermott – Director of Development, Univ. Advancement, NJIT

- Vincent Oria, PhD – Professor, Ying Wu College of Computing Sciences, NJIT

- Reynold Panettieri, MD – Professor, Robert Wood Johnson Medical School, Vice Chancellor of Clinical & Translational Science, Rutgers Univ.

- Ram Ramesh, PhD – Professor, School of Management, SUNY at Buffalo

- Jinhe Shi – PhD Candidate, Ying Wu College of Computing Sciences, NJIT

- Julia Stoyanovich, PhD – Assistant Professor of Computer Science, College of Computing & Informatics, Drexel University

- Cheickna Sylla, PhD – Professor, Martin Tuchman School of Management, NJIT

- Christie Teigland, PhD – Vice President, Advanced Analytics at Avalere Health

- Dimitri Theodoratos, PhD – Associate Professor, Ying Wu College of Computing Sciences, NJIT

- Fusheng Wang, PhD – Assistant Professor, Department of Biomedical Informatics, Department of Computer Science, Stony Brook University

Wednesday, October 5, 2016
Thursday, October 6, 2016

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

- Heng Xu, PhD – Associate Professor of Information Sciences and Technology, Pennsylvania State University

- Dantong Yu, PhD - Associate Professor, Martin Tuchman School of Management, NJIT

- Nan Zhang, PhD – Program Director, National Science Foundation

Wednesday, October 5, 2016
Thursday, October 6, 2016

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Executive Summary

Edited by Yi Chen

The aims of the DaSH symposium are: i) to understand the grand challenges in developing data science technologies that use big data to improve outcomes and to reduce costs in healthcare; ii) to discuss research directions and priorities; iii) to foster collaboration and community building between data science and healthcare across multi-sectors.

The theme of the DaSH 2016 symposium is "Data Sharing: the Art, Science and Ethics". Data sharing is a complex and challenging issue with both technical and social implications.

The DaSH participants discussed the importance of linking and mining data from diverse sources, and *the call for open platform, open data, and open minds* (Dr. Bradford Hesse). Data sharing is an essential step to obtain comprehensive information for better understanding situations in order to improve quality of care, and to reduce the cost.

Many challenges in data sharing were discussed, from social, technical and ethical aspects.

- Social challenges of data sharing due to low incentives. Data owner may have low trust to others, have concerns about the loss of commercial value of the data, and/or the loss of advantage of publishing, as well as potential flaws in the data. In current practices, there is significant under-appreciation in data sharing, and lack of business model to provide incentives to encourage data sharing.

- Technical challenges of data sharing due to the lack of interoperability of different systems. Data are stored in different systems (various database systems, files, etc.), with different schemas, whose interpretation should be context-dependent as modules may interrelate to each other. After the schema are matched, we also need to perform record linkage at data instance level, with challenges including data conversion and data uncertainty.

- Ethical challenges: There are security and privacy concerns, as well as legal issues in data sharing and the use of shared data. Challenges include policy making and policy enforcement.

Several domains of data sharing were discussed. One domain is sharing electronic health records of patients. There is no master patient database in United States. Several approaches were discussed to address this: demand to build such a database at policy level, or encourage patients to voluntarily share data with certain regulatory processes. Some states, such as New York, has made initiatives on electronic health record exchange among different institutions. Another domain is data sharing and data linking between environmental, social, and behavior data to healthcare to better understand health disparity and factors that impact health status.

Wednesday, October 5, 2016
Thursday, October 6, 2016

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Then several initiatives of data sharing in practice were discussed, including OHDSI: Observational Health Data Sciences and Informatics initiatives, IEEG.org for epilepsy data sharing, and the health information exchange initiatives at New York State.

Participants had breakout sessions to discuss several aspects in data sharing. 1) How to encourage data sharing? What are the types of data relevant to healthcare research? What are current barriers to sharing? How can we provide incentives for data sharing? 2) How to facilitate data sharing? What infrastructures and standards should be developed to facilitate the sharing? What are the technological challenges in sharing? 3) How to ensure integrity in data sharing? The report of the three breakout groups are included:

- Breakout Session I: Encouraging Data Sharing,

- Breakout Session II: Facilitating data sharing,

- Breakout Session III: Responsible data sharing.

DaSH also had focused discussions about how big data can contribute to the Cancer Moonshot mission. Several unique challenges in cancer treatment are discussed. Clinic challenges include diverse factors of different patients, but yet no much diversity in phenotype among patients, and high cost associated with saving each individual with unique characteristics. Challenges in data processing include: missing data, data noises, data sparsity, and data volume. It's important to develop techniques for defining meaningful patient similarity, providing personalized model for precision medicine, risk early warning systems and interpretable insights. The evaluation of big data techniques should be patient-centric, with the goal of improving quality of life.

Several technical advances of data sciences on healthcare were discussed. In particular, Point-of-care-technologies (POCT) promises the benefits of increasing healthcare accessibility, cost-effectiveness, as well as promptness for care. The challenges lie in both technological (data quality, inter-operability, analytics) as well as to gain acceptance.

For more details, please refer to the DaSH website for presentation slides of the talks: https://web.njit.edu/~ychen/DaSH/

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Breakout Session 1: Encouraging Data Sharing

Edited by Dantong Yu

List of Participants:

Dantong Yu (NJIT) (Discussion Leader)
Cesar Bandera (NJIT)
René Bastón (Executive Director, NSF NE Big Data Hub)
Ari Brooks (UPenn)
Chennuru, Ashok (CUNY)
Shridar Ganesan (Rutgers)
Shanthi Gopalakrishnan (NJIT)
Ram Ramesh (UB)
Heng Xu (PSU)

Modern integrated healthcare practice and research increasingly rely on collaborations of healthcare providers, insurance organizations, and data scientists who share data in an unprecedented scale. The requirements for data sharing are increasingly important for improving healthcare quality and driving down the healthcare costs. We need to build a data ecosystem to encourage data sharing. However, the technical, social, organizational, economic, ethical, and regulatory barriers that commonly exist in today's healthcare industry must be addressed to ensure cost-effective data sharing among all stakeholders. Improper data handle and sharing without sufficiently addressing these barriers lead to high cost and liability to all involved parties. The number of complaints filed with the U.S. Department of Health and Human Services has steadily risen. For example, the number of complaints filed in 2013 was 12,915 and nearly the double of the number in 2004. Many complains are the result of insufficient privacy measurement and control that placed patients in risk. The government regulation enforces data sharing, however, existing government monetary incentives are not very effective and will not be extended indefinitely. To ensure sustainable data sharing and its supporting exchange systems, we environ that a business model for data sharing and information exchange is necessary.

To encourage data sharing, we must first identify all barriers, challenges and difficulties in data sharing, introduce appropriate incentives to mitigate and eliminate barriers, and make recommendations to stakeholders based on the findings and principals outlined as follows.

Introduction to Data Sharing

Data sharing is the practice of making (a fraction of) data that are collected from daily (healthy care) practice, subject monitoring and tracking, and scholarly research activities available to others including investigators, peers, communities, and even competitors. Data sharing does not exclusively focus on the raw format of data. Actually, it involves higher level meta data for

understanding and reproducing data. Data sharing either involves data transfer over physical media from owners to sharers, or does not require data exchanges at all. For example, data owners allow others to submit data analysis algorithms and then collect statistics and high level knowledge without exposing and transferring raw data.

In general, there are two types of data sharing in healthcare industry: the first data sharing is among care providers to facilitate decision making by primary care physicians, specialist, and major hospital where medical records are the targets to be shared, and the second type is sharing with researchers for knowledge discovery where high-level data records, mostly de-identified, are involved. These two types of sharing ultimately serve the objectives of reducing the cost of health care, discovering human knowledge on disease and treatment, and improving the quality of patient care.

Case Study of Data Sharing

HEALTHeLINK is an online health link exchange platform that involves major hospital systems, labs, radiologists, insurance organizations in the Western New York. All medical records will be available through the electronic medical record (EMR) system that enables all radiology and lab test results to be shared by primary care providers (PCP), insurers, and specialists. There are six levels of access and control that are attached to different type of data records - from the level that allows anyone to access to the strictest level that bans access to many parties. The EMR supports "informed consent" where patients grant or deny public access to their data. Within HEALTHeLINK, some patients grant public access to their data. Currently, HEALTHeLINK is a multi-sided platform, each side has to pay for the services that they use. Insurance company are the ultimate user of this health link exchange.   The health link exchange already has a good market penetration: 95% of 9000 physicians in the Western New York are members to share lab results via HEALTHeLINK. Since the introduction of HEALTHeLINK, ER takes advantages of HEALTHeLINK, resulting in an effective ER care, i.e., the number of repeated visits to ER and hospital admissions is significantly reduced.

Identify The Barrier, Challenges and Difficulties for Data Sharing

Technical Barriers: Data appear in different formats, hand-writing notes, electronic records, structured (patient records) and unstructured notes (lab results, doctor prescriptions, and radiology graphics), and belong to different data owners, and are usually scattered among various components of information systems.   Consequently, only a fraction of data is availability for access beyond their original owners. The whole data is not availability simultaneous, which presents a significant barrier to effective sharing. The variation in data quality is another technical barrier to be considered in data sharing. Many factors contribute to data quality: the proficiency of data operators, the accuracy of data equipment, sufficient sample space, missing and corrupted data entries, and well-curated metadata.   Data quality varies even among different components in the same dataset. Data must pass a certain quality threshold to ensure a meaningful data sharing.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Recent years, healthcare industry generates big data and overwhelms current management information system. Seven million patients are newly enrolled in Obamacare, and the volume of information is going to explode. The existing data sharing system is not designed to handle the big data at such a scale.   One potential approach is to put data into public clouds, such as Amazon S3, or private clouds.   The vulnerability of data cloud and risk of data breach must be sufficiently addressed. Regardless where sensitive data are stored, they must be protected to meet the basic requirements and privacy regulations, such as the health insurance portability and accountability act of 1996 (HIPPA). Data owners must stipulate the requirements that commercial or private clouds store and process data according to the privacy rules of HIPPA, and separate their sensitive data from other clients' data.

Social and Organization Barriers: Lack of trust on data source is the primary reason why one user rejects data from others. For example, ER personnel does not trust data from paramedics, in particular diagnosis based on vital signs. Another type of social barrier is risk aversion in which one avoids taking any risk to (un)share his/her data.   Due to the potential excessive litigation costs, practitioners and hospitals hire attorneys to govern and oversee data sharing among individual institutes. Instead of using appropriate risk management, they often resort to risk aversion, which greatly impede the processing of data sharing.

Economic Barriers: Data owners invest significantly to generate and collect proprietary data that might contain commercial values for related products and intellectual properties for current and future publications. For example, pharmaceutical companies own high quality clinical trial data. Only can government possibly obligate them to share their clinical trial data for drug approval.   However, such a regulation does not exist. It is usually difficult to provide incentive to these data providers to share their valuable data. In addition, data sharing incurs a steep cost because heterogeneous data and data systems are not fully interoperable, and to exchange data with heterogeneity always incurs significant operation costs. Data sharing involves a steep learning curve for anyone to share and often distracts care providers from their primary activities of dealing with patients. In addition, membership, hiring and training medical personnel to participate and use sharing systems also incur financial cost to care providers.

Stakeholders

Many person (stakeholders) are involved in the data sharing activities: data owners and data users. Each type of stakeholder has their objective in data sharing activities. For example, physicians who pay high cost (time) aim to improve the quality of care, insurance companies often have the objective of eliminating repeated procedures, trimming the costs and maximizing profits, patients who share their data want to improve his/her healthcare experience and quality, the billing department benefits from data sharing via a smooth processing across multiple entities, and researchers who request data sharing design and evaluate innovative care procedures and methods and making new science discoveries.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Incentives to Data Sharing

The objective of introducing the following incentives is to improve the quality of care at a low cost.   We outline the incentives for data sharing to insurance company, care providers, and patients.

When data sharing occurs with insurance companies, the direct benefits to insurance companies include sharing risk managements, reducing duplicated tests, and minimizing the cost.   By participating direct exchanges, insurance companies can verify whether a patient claim matches with real patient care history, evaluate the validity of the claim, minimize duplicated medical tests and wastefulness, and identify the potential abuse and even frauds in medical claims.

In the healthcare informatics, data sharing is an essential part of the process that engages practitioners and researchers to obtain and verify diagnosis and find new results. In particular, various participants often contribute their different expertise and develop multiple interpretations of data from different perspectives. This leads to a high data quality control as many data users will exam data and identify data outliers, and further improve data as appropriate.   As many physicians share data, a collection of wisdom will improve data analysis and data-driven diagnosis, shorten the time on diagnosis, and reduce the risk by involving a second opinion.

Physicians often perform comparison studies to evaluate the effectiveness of a new treatment. They typically select control and treatment groups manually with enough members in each group. Some patients receive the standard treatment or no treatments at all, and form a control group, while others get alternative treatments and form a treatment group. The confidence of this type of comparison study is sensitive to the size of two groups.   Data sharing among physicians will pool together patients with the same symptom and therefore increase the size of two groups for a high-fidelity study.

Data sharing between patient and physicians also involves a great amount of incentives to both patients and physicians. First of cost, a patient does not redo medical exams and treatments by reusing the lab reports and doctor notes prescribed by different physicians, avoids the latency to attain duplication exams and wait for duplicated treatments, and greatly reduces his/her time and monetary costs of medical care.

Recommendations:

Data sharing is a complex activity involving multiple parties. We craft our recommendations that consider multiple aspects, including individual information right and obligations, property right, system and data quality to conform to regulatory requirements on sharing data, and accountability and control. Given the aforementioned analysis, we summarize several guiding principles to ensure effective data sharing practices: 1) the quality of data must meet a predefined standard for sharing, 2) we must set a clear demarcation between data for research and data for care providers and apply different access controls to data users, 3) if barriers can not be eliminated with a reasonable cost, we need to reduce and mitigate barriers to reach a trade-off between

sharing and protection, and 4) data literacy training is necessary to objectively measure and analyze risks in data sharing.   We enumerate our recommendations as follows:

- Demonstrate the value chain to get insurance company to sponsor the exchange point. For example, show the number of claims to decrease, and demonstrate that all constituents (payers and providers, members) will benefit from the value chain.

- Introduce data brokering and streamline the approval pipeline for data sharing, to ease the processing of exporting patient records, eliminate a labor-intensive process (scan and fax) to get medical records, share data digitally (raw, interpreted, meta data) and minimize the burden of data owners

- Establish data quality control to ensure seamless data integration and formally recognize and value the practices of uploading high-quality data and adding values to exchange point in return by reducing their fee for participation.

- Provide training of the best practice for sharing data to attorneys, practitioners and medical personnel, so data sharing becomes a part of standard practice. Furthermore, training content includes risk analysis, risk benefit balanced decision making.

- Adopt open data format and standard and eliminate the proprietary forms.

- Empower patients to access their data and and involve patience into the loop.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Breakout Session 2: Facilitating Data Sharing

Edited by James Geller

List of Participants:

James Geller (NJIT) (Discussion Leader)
Judith Argon (GHS)
Soon Chun (CUNY)
Jack Fong (Leir Foundation)
Zachary Ives (UPenn)
Jianying Hu (IBM)
Vincent Oria (NJIT)
Nan Zhang (NSF)

The group was charged with identifying technologies to facilitate Healthcare data sharing, including clinical data and research data.

Background:

A number of technologies have been developed or are in the process of development that were specifically designed for medical data sharing. These technologies include:

- i2b2 (Informatics for Integrating Biology and Bedside), SHRINE, the i2b2 interface

- PCORnet, CDM (Common Data Model)

- HL7, FHIR (Fast Healthcare Interoperability Resources), CDA (Clinical Document Architecture)

- Blue Button, CommonWell Health Alliance, Microsoft Health Vault, ONC's S&I Framework

- caBIG (the Cancer Biomedical Informatics Grid) which was terminated and replaced by the National Cancer Informatics Program (NCIP)

- OHDSI, OMOP (Observational Medical Outcomes Partnership)

- CARPEM, EFPIA and other European projects

A detailed description of all these systems goes beyond the scope of this brief summary.

Data Sharing of clinical data and medical research data has various challenges:

- Data owners or data collectors often do not see the immediate and long-term value/benefits of sharing the data. How can we motivate the data stakeholders to encourage data sharing?

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

- Structural and Semantic heterogeneity continue posing challenges to data sharing. Value creation from data sharing often requires aggregation and linking data from different sources. Any technology that allows the use of data (e.g. in table format) from two or more independently developed sources but in the same subject domain will face structural and semantic problems.

- Lack of a technology platform that is easy to adapt to transient research networks. Data sharing often involves a network of data holders, depending on the research projects. A technology platform should allow forming a team and a data sharing network dynamically and effortlessly. The design requirements of such a platform need to be spelled out in detail.

Semantic and Structural Issues in Data Sharing

Structural problems include, for example, the use of different data types by different data repositories. Thus one repository might represent patient IDs as integer numbers, while another implementation would use character strings. Both these choices are defensible. Using character strings makes it easy in a future upgrade to change patient IDs from pure numbers to mixed letter number codes. By and large, structural problems are considered solved or fairly easy to solve with current technology.

Semantic problems are much harder to solve. Examples of semantic problems include the following:

Synonym problem: One system uses a term (e.g. Kohler's disease) while the other system uses a different term of the same meaning, in this case, Navicular Osteochondrosis. The knowledge that these two terms are synonyms is not present in either of the two medical data repositories. An additional knowledge resource is needed to express this synonymy.

Homonym problem: One term is used in both data repositories (e.g., cold). However, in one repository "cold" stands for "the common cold," a disease, while in the other data repository it stands for the qualitative body temperature "cold."

Abbreviations: Abbreviations create additional complications with homonyms. Thus the abbreviation PCP could stand for Primary Care Physician, or for phenylcyclohexylpiperidine (aka Angel Dust) or for Pneumocystis Pneumonia. In fact, an online dictionary lists 107 different meanings of PCP, several of them from the medical domain.

Wrong granularity (i.e., wrong level of generality): Two data repositories might express the same data at different levels of abstraction. Even more commonly, a data query against the combined data of two repositories might be expressed at a very general level, but specific answers are expected. (Refer to Figure 1 for the following text.) Thus a medical database or an EHR might contain diagnosis data at the level of angiosarcoma and hepatoblastoma. A query might be expressed for patients with cancer. Data on all patients with angiosarcomas and hepatoblastomas should be returned. However, the fact that an angiosarcoma is not expressed in the database.

Similarly, a query for patients with liver diseases should return angiosarcoma, as it is a liver cancer. However, the location of angiosarcoma is probably not explicitly stored in the database.

Ontologies

Ontologies have been developed to deal with the above problems. In the simples case, an ontology is a set of concepts. Pairs of these concepts are connected by IS-A links that point from the more specific concepts to the more general concept. Thus an ontology would contain angiosarcoma and an IS-A link to the concept "liver cancer." Furthermore there would be a link from "liver cancer" to "cancer." Any query to a medical database for patients with cancer would be internally translated into a query that includes all the kinds of cancer linked in the way described here. Thus a query for cancer might become a query for [angiosarcoma or hepatoblastoma].

Synonym information is normally stored directly in the ontology. Thus there would be one single concept for Kohler's disease that would be linked (would "store") all the existing names of Kohler's disease, including Navicular Osteochondrosis. To make concepts truly unique they are usually assigned a unique ID number.

Homonym information is also stored in an ontology. Thus the term "cold" would be linked to the concept cold-temperature and the concept cold-disease. Acronyms are included with the concepts that stand for the expanded terms.

From the above it would appear that ontologies are the silver bullet for all semantic data integration problems. However, this is unfortunately not the case, because building ontologies is very difficult. As a result, ontologies often miss different elements. If an ontology is missing a synonym of a given concept, then the synonym problem is not solved for this concept.

Many concepts allow several "parents" (i.e., more general terms) for one given term. Thus viral pneumonia might be marked as "IS-A lung disease," but also as "IS-A infection." When concepts need two or three such IS-A links, it frequently happens that one of them is accidentally omitted. Thus, a query application program against a medical database that peruses a "defective" ontology of this kind will return incomplete results.

The SABOC research group of Dr. Perl and Dr. Geller at NJIT has spent the last 23 years on developing methods for making it easier to find errors and omissions in ontologies. Some of these methods are fully automatic, but most of them consist of a combination of algorithms that suggest "suspicious" concepts to a human expert who then has to verify the suspicion or exonerate the concepts. An important idea that was developed by Dr. Perl and Dr. Geller is the Abstraction Network. As many medical ontologies are very large and complex, different kinds of Abstraction Networks may be derived from them and then visualized with specially-developed software, making it easier to find suspicious concepts.

Symposium on Data Science for Healthcare (DaSH)

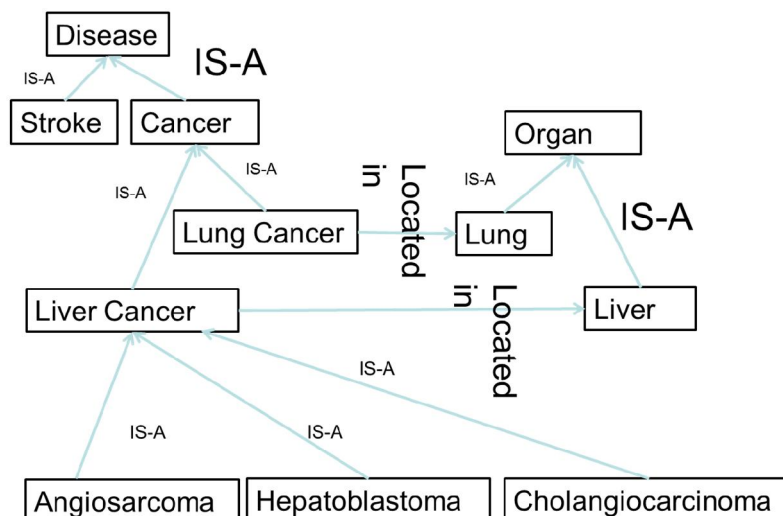## Data Sharing: The Art, Science, and Ethics

Figure 1: A "toy ontology"

Even an ontology that was initially perfectly correct, which is very unlikely, might become incorrect due to semantic drift over a longer period of use and update. Thus, the discovery of a new kind of diabetes (neither Type 1 nor Type 2) in essence changed the concept "Diabetes" itself. If the medical database is changed to accommodate the new type of diabetes, but the ontology lags behind, then the query results derived by using the ontology will be wrong.

Yet, with all the well-known problems of ontologies, there is no known alternative for using an ontology or ontology-like structure (controlled vocabulary, etc.) to deal with the semantic problems listed above. Thus, investments should be made in continued work on ontology quality assurance. With apologies to Winston Churchill, "Ontology is the worst mechanism for data sharing except for all the others."

Motivational and Economic Obstacles to Data Sharing

According to a REPORT TO the US CONGRESS on Health Information Blocking ("Report"), dated April 2015, and prepared by the Office of the National Coordinator for Health Information Technology (ONC), a new phenomenon has been reported: Information blocking. According to the report, information blocking is defined as "when persons or entities knowingly and unreasonably interfere with the exchange or use of electronic health information." Quoting from the Report, "Market conditions create business incentives for some persons and entities to exercise control over electronic health information in ways that unreasonably limit its availability and use."

As an example, the Report describes the scenario where a patient receives medical services from two large competing hospitals. When the patient moves from Hospital A to Hospital B, she requests that her medical record be forwarded to Hospital B. Hospital A follows the order by sending a record of several 100 pages that exists in Hospital A in an electronic format by *fax* to

Hospital B. This unreasonably interferes with the effectiveness of operation at Hospital B and, if done knowingly (with this intention), would be a case of information blocking.

The Report lists three categories of steps to be taken against Information Blocking: Targeted Actions, Broader Strategies, and how to address Gaps in Current Knowledge, Programs, and Authorities. In each category there are multiple proposals.

We discuss Information blocking to make it clear that there are massive financial incentives to counteract any efforts directed at facilitating data sharing. Thus, the existing problems cannot be solved on a technical level alone. Rather, solutions have to be found as part of a regulatory framework.

Suggestions for Solutions

Different kinds of data will need different kinds and extents of preparation for sharing. There is a big difference between personally identifiable patient data and anonymized data from a big cohort, derived by a single query, extracting a small number of features of the patients.

Homebase for Patient Data: We propose to create the regulatory environment, followed by the technical infrastructure to allow, and possibly require, every patient to choose a "health data home base." All healthcare providers will then be required to ascertain where the health data home base of a patient is, and will have to upload any data that they have created or collected about the patient to this health data home base. Furthermore, cost will have to be kept at a nominal level, comparable to the cost of an electronic funds transfer between two bank accounts, which is typically free for smaller cash amounts.

Research Study Data: With respect to research study data we notice that there are two problems that impede data sharing. One problem is that owners of data often do not want to share it. For example, a research group might have written two academic papers about a data set that they have collected at great effort and great expense. However, there are enough discoveries in the data to justify writing two more papers. Naturally, the data owners would not want to share the data until they have written those papers and exhausted the data set, because competing research groups might preempt the efforts of the data creators by publishing papers about the data before them. Clearly there are many more scenarios, but we will not delve deeper into this side of the issue.

On the other hand, we have more strong anecdotal evidence that data that is made available to others is often *not picked up.* Thus the problem is not just on the side of the party who owns the data but also often on the side of potential users who either don't know that relevant data exists or who are too deeply involved and busy with their own work and data to consider retrieving and using data from another source.

In some cases there is an additional problem. If data is shared because this is required by a funding agency or a journal policy, but the data owner has reasons not to share the data, then the owner can follow the letter of the law and share the data; however, explanations or context

information of the data creation (i.e. metadata) will not be provided, or will be provided with a great delay. In many cases this might make the data unusable.

As a solution we draw on the analogy of the US National Library. By law, every book that appears in the United States in paper form must be submitted in one exemplar to the US National Library for archiving. A similar method should be used for all medical research study data. The institution for this purpose exists already, the National Library of Medicine (NLM). However, the NLM apparently has been underfunded compared to many other National Institutes of Health. The charge to become data steward and archive for all medical study data in the United States should be added to the NLM, with an according increase in appropriation of budget to create the physical infrastructure (server farm and staff) that is necessary.

To encourage individual researchers to not just share data, but do so willingly and support users, we propose the creation of a mechanism similar to the Citation Index. This will require the creation of a Share Index in which is recorded which data sets are used by researchers from other institutions that were not involved in creating the data set. In parallel to that, a cultural change will be necessary. Promotion and Tenure Committees will need to start to get Share Index reports in the same routine way as they are now getting Citation Index reports.

Design of Sharing Network Platforms

The scientific discovery for new treatments or new drugs for patients requires a network of partners, sharing clinical data, clinical trials data, genomic data, scientific experimental data, drug data, health monitoring data, healthcare management data, etc. To establish such a **sharing network**, a data sharing platform requires a stack of technology components. These components include:

Metadata: Describe the data with metadata, including contextual data, to easily discover the right data at the right time.

Provenance: Data lineage is important in the scientific discovery process. The provenance of data records and information on the history of how a piece of data has been processed and handled are essential.

Workflow: Clinical research often requires a set of data composed from many different sources. The platform should provide the capability to specify steps and processes on the data to be shared and the data granularity of every source.

Data discovery: The data sets and their APIs in the network should be registered and catalogued for discovery and access.

Privacy: To provide the basis for ethical or responsible data sharing, privacy-preserving technologies should be part of the platform. Deidentification of data and anonymization techniques must be considered. Access control for the network partners and their roles have to be required for preventing unauthorized exposure of clinical data.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Usability: The system should provide ease of use for end users, and the network should be easy to configure and adopt for different projects.

In order for the technology solutions to be widely adopted, we may have to consider business questions: Which institutions may be suitable trusted parties to install, maintain and sustain the sharing network? What is the business model for the trusted intermediary provider, as the maintenance and management will need infrastructure and human resources? A value-driven business model may emphasize the values created by sharing, in terms of the data products, so the partners would pay for the services and products. One can also consider a payment model where higher quality data and/or completed experimental data are charged for at a higher rate.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Breakout Session 3: Responsible Data Sharing

Edited by Julia Stoyanovich

List of Participants:

Julia Stoyanovich (Drexel U) (Discussion Leader)
David Belanger (SIT)
Kathy Grise (IEEE)
William McDermott (NJIT)
Reynold Panettieri (Rutgers)
Cheickna Sylla (NJIT)
Dimitri Theodoratos (NJIT)
Fushen Wang (SBU)
Jinhe Shi (NJIT)

Introduction

The break-out session on responsible data sharing discussed the ethical issues that arise in the context of data collection, sharing and use in the health domain. The session was moderated by Prof. Stoyanovich (Drexel, Computer Science). Participants included academic researchers and practitioners who offered their perspectives on various aspects of the problem, and engaged in a lively discussion.

To set the stage, the group discussed the 1979 study of alcohol consumption in the Native American population in Barrow, Alaska, which is often used as a case study in ethical research practices (see University of Alaska Fairbanks / IRB at http://www.uaf.edu/irb/readings/ for references). The study was fraught with methodological issues, its results were published in a report "The Inupiat, Economics and Alcohol on the Alaskan North Slope", and were made public in a press release simultaneously with informing study participants.   The results were subsequently sensationalized by the media, bringing tangible harms (reputational and economic) to the Native American population of Barrow, Alaska and beyond, and eroding trust for scientific research in these communities.

These events are unlikely to be repeated today, due to the safeguards put in place by the IRBs and to a higher degree of awareness about ethical issues in the scientific community. However, the Barrow Alcohol Study serves to highlight that *ethical principles* can be violated even when the *ethical and legal rules* of the day are followed. Rules and regulations necessarily lag behind the rapid development and adoption of research methodologies, which in this day and age are increasingly data-centric. In the context of data sharing, this means that all stakeholders of the data sharing ecosystem bear a *responsibility* for their actions (or inaction) and must act ethically.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

In what follows, we will describe the stakeholders of the data sharing and analysis ecosystem, and will analyze the harms to which the stakeholders may be subjected, their incentives for ethical data sharing practices, and their responsibilities.

Stakeholders: Harms, Responsibilities and Incentives

The group briefly discussed the stages of the data sharing pipeline, which include (1) data collection, (2) integration, (3) cleaning and pre-processing, (4) analysis and meta-analysis, (5) publishing of datasets, analysis methodologies and results; and (6) interpretation of results and interrogation of methodologies. We noted that *curation*, which may be manual, semi-automatic or automatic, permeates all stages of a responsible data sharing pipeline.

The *stakeholders* in the data sharing pipeline include (1) the data provider, e.g., a patient or a participant in a clinical study; (2) the data collector, e.g., a healthcare provider, a pharmacy, an insurance company, or the government; (3) the data publisher, such as a health informatics exchange (HIE); (4) the data scientist who performs data analysis; (5) the medical and scientific community; and (6) the public.

The group considered a range of possible harms to the stakeholders.

- Privacy violations present a classic risk of inclusion that may impact data providers.

- In contrast to privacy, lack of representation of patient cohorts is a risk of exclusion, which works contrary to the goal of ensuring uniform quality of service and availability of treatment for different subpopulations, such as those based on ethnic group membership, gender, age, or disability status. Lack of representation is an aspect of fairness.

- Another harm is due to misinterpretation of results, as illustrated by the Barrow Alcohol Study discussed in the introduction, which poses a risk to data providers but also to members of subpopulations who are indirectly associated with the data providers, e.g., share demographic characteristics with them.

- Lack of reproducibility in data analysis methodologies, and lack of robustness or data analysis results, harms the medical and scientific community (http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970) and the public. This issue is often caused by methodological issues in data collection and sharing, including (a) low data quality, due, e.g., to ambiguity and noise; (b) bias in the data being analyzed -- including non-uniform coverage, lack of diversity, over-representation -- in data collection, integration, cleaning, and analysis; (c) insufficient samples size; (d) effects due to multiple hypothesis testing, known as p-hacking; (e) blurring the line between exploratory and confirmatory research methodologies.

Having considered the harms, the group discussed the responsibilities of the stakeholders to mitigate these harms. While responsibility for ethical conduct is shared by all stakeholders, specific harms should be mitigated by the least cost avoider.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Data collectors and data publishers should exercise due diligence in data cleaning and annotation, ensuring the veracity and interpretability of the data. Data publishers are also responsible for ensuring the privacy of data providers, and for providing information about bias with respect to coverage, diversity and representativeness of the data. Further, we observed that the act of sharing data compels the publisher to consider the potential harms that the data may bring to individual participants and to population groups. Importantly, while data publisher typically has no legal responsibility for subsequent use of data, but there may still be ethical concerns.

Data scientists, taken broadly to represent individuals and organization who are performing data analysis, are responsible for making explicit that research hypothesis is appropriate for data, precisely stating assumptions and qualifying applicability of results. Data scientists are responsible for making results interpretable in context, and for ensuring transparency of the data analysis process, enabling "analysis of the analysis". Finally and importantly, data scientists have the responsibility to be skeptical, especially when attempting to analyze and make conclusions based on data of unknown provenance, i.e., for which the origin, the assumptions under which it was collected, and the history of cleaning and transformations, is unknown.

Next, the group discussed the ways to *incentivize ethical data sharing*. We observed that ethical and principled data sharing should be rewarded by existing *academic structures* on par with scientific publications, and emphasized as part of peer review. Currently, academic structures instead introduce a lag in data sharing, as there is a perceived need to publish first, withholding data until a publication. Beyond this, data generation and curation are insufficiently valued, e.g., citing data is cumbersome (see a description of ongoing work on data citation in CACM: http://cacm.acm.org/magazines/2016/9/206243-why-data-citation-is-a-computational-problem).

The group also suggested that research *funding* should be available to support research that emphasizes ethical data sharing, and that ethical data sharing *training* should be part of standard research training.

Research collaborations can serve as a natural incentive, since the data owner is more likely to share data with someone who will recognize the contribution, usefulness, potential, and beauty of the data.

The group also discussed the importance of legal and policy frameworks (e.g., IRB) in regulating data sharing and use. However, these mechanisms are reactive in nature, and cannot be the sole gatekeepers of ethical data sharing.

Towards A Code of Ethics In Data Sharing

The group recognized that there is a need to develop recommendations and guidelines that support effective and ethical sharing of data, methodologies and results in the health domain. These guidelines will include aspects of fairness, transparency, repeatability, and interpretability that are in common with other areas of data-intensive science and engineering, and are the

subject of ongoing discourse about data-driven algorithmic decision making. The main question that a code of ethics should help answer is *"Are you, the stakeholder, acting professionally?"*

A code of ethics can be developed by consulting a wealth of available resources, including:

- EFPIA and PhRMA: Joint Principles for Responsible Clinical Trial Data Sharing to Benefit Patients: http://transparency.efpia.eu/responsible-data-sharing
- Data Science Association: Code of conduct: http://www.datascienceassn.org/code-of-conduct.html
- American Statistical Association: Ethical guidelines for statistical practice(http://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx)
- Certified Analytics Professional: Code of ethics / conduct: https://www.certifiedanalytics.org/ethics.php
- ACM code of ethics, under revision: https://www.acm.org/about-acm/code-of-ethics

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

Biographies of Attendees

Wednesday, October 5, 2016
Thursday, October 6, 2016

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

JOEL S. BLOOM, EdD

Joel S. Bloom, President of New Jersey Institute of Technology started his career in industry working as an economist. He subsequently became an educator and administrator for the New York City public schools. Prior to coming to New Jersey, he worked as a research director and instructor at Teacher's College, Columbia University. In New Jersey, he managed state and federally funded curriculum development and training centers for the state. From 1983 through 1990, Dr. Bloom had served as assistant commissioner in the NJ Department of Education. He was responsible for managing many of the education department's initiatives: competency testing, curriculum content standards, pre-school programs, establishment of model effective schools, alternative schools, and science and technology academies.

From July 1990 through September of 2011, Dr. Bloom had served as NJIT's Vice President for Academic and Student Services. He was responsible for nine divisions of the university, including pre-college programs, enrollment management, continuing professional education, career development services, dean of student services, freshman studies, educational opportunity programs, library services, physical education and athletics.

In 1998, Dr. Bloom was appointed as the first Dean of the Albert Dorman Honors College, in addition to his vice president responsibilities. The Honors College currently enrolls over 700 students who come from 15 states and several foreign countries. The combined average SAT score (math and reading) is over 1340 and the honors scholars are in the top 10 percent of high school graduates.

In September 2011, Dr. Bloom was appointed interim President of NJIT, and appointed President January 2012. Under his leadership, the "Warren St. Village," comprised of a residential Honors College and ten Greek houses was opened September 2013 adding 600 residential students and amenities including a fitness center, restaurant and convenience store. Dr. Bloom's leadership has also resulted in a fall 2015 enrollment of more than 11,000 students, the hiring of 68 new faculty and 30 lectures over the past four years, annual research expenditures of over $100 million, a renewal and new facilities plan of nearly $300 million, a new strategic plan "2020 Vision," and the development of a NJIT corporation, the New Jersey Innovation Institute (NJII). NJII was

created as an industry-facing organization to partner with companies, government and other universities for the purpose of economic development. Since its April 2014 inception it has generated grants, contracts and corporate memberships of $15 million. Dr. Bloom is presently Board Chair of the University Heights Science and Technology Park, Vice Chair of the NJ President's Council, Vice Chair of NJ's Edgenet board; and, serves as a member of the Governor's Council on Innovation, and board member of the Philadelphia Alliance for Minority Participation, La Casa don Pedro, and the APLU Presidents' Council. Dr. Bloom has most recently been named "Educator of the Year" for 2015 by the New Jersey R&D Council, and been the recipient of other national and state education awards, as well as federal, foundation and corporate grant awards. He has published journal articles and presented papers on organizational development, curriculum development and student assessment.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

FADI P. DEEK, PHD

Fadi P. Deek is Provost and Senior Executive Vice President at New Jersey Institute of Technology (NJIT) where he began his academic career as student in the early 1980s. He received his B.S. Computer Science, 1985; M.S. Computer Science, 1986; and Ph.D. Computer and Information Science, 1997, all from NJIT.

In addition, Dr. Deek is Distinguished Professor with appointments is in two departments: Information Systems (in the Ying Wu College of Computing Sciences) and Mathematical Sciences (in the College of Science and Liberal Arts). He also holds a joint appointment in the Information Technology Program and serves as a member of the Graduate Faculty - Rutgers University Business School.

Dr. Deek maintains an active funded-research program. His research interests include software engineering and open source software development, with applications to learning/collaborative/decision-support technologies, and computer science education. Dr. Deek has published over 150 articles in journals and conference proceedings, ten book chapters, and three books: O*pen Source – Technology and Policy*, (Cambridge University Press, 2007, 369 pages co-authored with J. McHugh), *Strategic Software Engineering – An Interdisciplinary Approach* (Taylor & Francis Group - Auerbach Publications, 2005, 350 pages co-authored with J. McHugh and O. Eljabiri), *Computer-Supported Collaboration with Applications to Software Development* (Kluwer Academic Publishers, 2003, 265 pages co-authored with J. McHugh). Dr. Deek has also given over 40 invited and professional presentations and has been a Principal Investigator on a number of large projects, most recently the *Computation and Communication: Promoting Research Integration in Science and Mathematics (C^{2}PRISM)*, $3 million from the National Science Foundation (2007-2011; R. Friedman, CP; B. Bukiet, CP). In addition, Dr. Deek has served as advisor or co-advisor to 13 PhD students.

Dr. Deek has received numerous teaching, research and service awards: The NJIT Student Senate Faculty of the Year Award, given to him in 1992 and 1993; the NJIT Honors Program Outstanding Teacher Award in 1992; the NJIT Excellence in Teaching Award in 1990 and 1999; the NJIT Master Teacher Designation in 2001 and the NJIT Robert W. Van Houten Award for Teaching Excellence in 2002. He has also been awarded the NJIT Overseers Public and Institute Service Award in 1997, the IBM Faculty Award in 2002, the NJIT Student Senate Administrator of the Year Award in 2007, and the NJIT Alumni Association Alumni Achievement Award in 2009.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics


REGGIE J. CAUDILL, PHD

Dr. Reggie J. Caudill is currently the Dean of the Martin Tuchman School of Management at New Jersey Institute of Technology (NJIT) and the Panasonic Chair for Sustainability.   He received his PhD degree in Mechanical Engineering from the University of Minnesota and earned a BS in Mechanical Engineering and MS in Engineering Mechanics from the University of Alabama.   Dr. Caudill has over 30 years of experience in research related to industrial ecology, sustainability and industrial systems, with over $10 million in research funding since 1990 including a recently funded million-dollar grant from PSE&G bringing advanced energy efficiency and sustainability techniques to New Jersey industry.   A significant aspect of his teaching and research has focused on dynamic systems analysis, simulation and data-driven decision making in industrial and management systems, including design for environment, production operations and corporate sustainability.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics
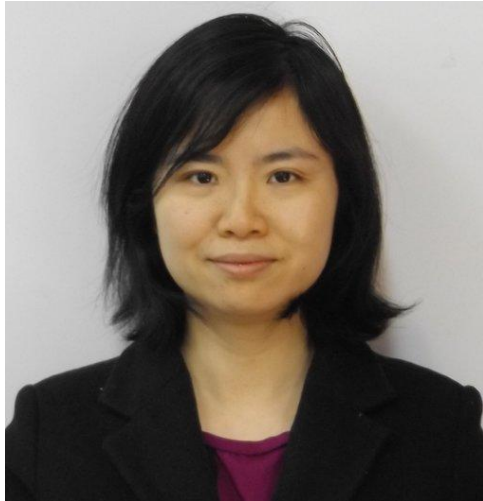
YI CHEN, PHD

Yi Chen is the first Henry J. Leir Chair in Healthcare, co-Director of Big Data Analytics Center, and an Associate Professor in the Martin Tuchman School of Management, with joint appointments in the Ying Wu College of Computing Sciences, at New Jersey Institute of Technology. Prior to this, she was an Associate Professor at Arizona State University (ASU). Yi Chen received her Ph.D. and M.S. degrees from the University of Pennsylvania in 2005 and 2000, respectively, and received her B.S. degree from Central South University (China) in 1999. Yi Chen's research provides cutting-edge technologies in big data analytics, with applications in healthcare. She has developed technologies for harnessing social media to provide clinicians and researchers the knowledge mined from patient-shared data, and to assist patients with relevant information for informed decision making. She also developed technologies for extracting bio- medical networks from literature, and for providing effective retrieval mechanisms to biological workflows and networks for knowledge discovery. Yi Chen has published over 70 journal and peer-reviewed conference articles. She also gave tutorials at premier conferences in big data, including the ACM International SIGMOD (Special Interest Group on Management of Data) Conference and the IEEE International Conference on Data Engineering (ICDE). Yi Chen's research has been funded by the Leir Charitable Foundations, National Science Foundation (NSF), Department of Defense (DOD), IBM, Google, and Science Foundation Arizona. Yi Chen is a recipient of an NSF CAREER award, a Google Research Award, an IBM Faculty Award, Excellence in Research at NJIT, an Outstanding Researcher Award in Computer Science and Engineering in ASU, and a Peter Chen Big Data Young Researcher Award. Yi Chen served as an associate editor for the Journal of Distributed and Parallel Databases, the Proceedings of Very Large Databases (PVLDB), and Electronic Commerce Research and Applications. She also served as an editor for special issues at IEEE Transactions on Knowledge and Data Engineering (TKDE) and PVLDB. She was a General co-Chair for ACM SIGMOD 2012 conference, and a PC Chair for international workshops: DBRank 2012 and KEYS 2009. She served on program committees for over 50 conferences.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

JUDITH ARGON

Judith Argon is an independent consultant with expertise in research operations, management, compliance, and administration, as well as grant writing and research ethics. Previously, she held the position of Chief Administrative Officer of Research at Geisinger Health System from 2006 -2015. While at Geisinger, she developed the administrative and operational infrastructure to support a growing research program. Among her achievements were a Board-approved ten-year plan for research, building of new research space, and the development of a biobank linked to a research-quality searchable version of Geisinger's EHR. Ms. Argon came to Geisinger following her role as the Vice President of Research Administration at the Children's Hospital of Philadelphia. In addition to expanding the research space, developing new core facilities and re-designing the administrative structures for research, during this time, she also guided the formulation of a research strategic plan, a plan for developing research IT and the implementation of HIPAA requirements. She was also the Director of Research Services in the Biological Sciences Division and the Pritzker School of Medicine at The University of Chicago, and prior to that, Director of Research Support at Duke University, where she established Duke's first pre-award and research policy office. Ms. Argon led Geisinger's participation in the Federal Demonstration Partnership and CHOP's membership in COGR, and was active nationally in research administration and policy matters. She is a member of a number of professional societies, such as the National Council of University Research Administrators and the Society of Research Administrators, as well as organizations devoted to ethics and compliance, including PRIMR and the Health Care Compliance Association.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics
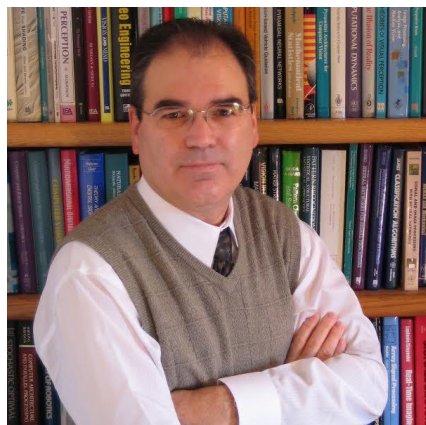
CESAR BANDERA, PHD

Cesar Bandera is assistant professor of entrepreneurship at the Martin Tuchman School of Management of the New Jersey Institute of Technology. His research interests include cross-cultural experiential entrepreneurship education and business incubation with a focus on innovation-driven ventures. Bandera is also founding partner of the mobile health company Cell Podium, which conducts wireless public health campaigns for CDC, NIH, EPA, and foreign ministries of health. Bandera received his Ph.D. in Electrical and Computer Engineering from the University at Buffalo, NY. His work has yielded the NASA Space Act award, Small Business of the Year Nominations from the US Department of Defense, four patents, and numerous publications. Bandera is Director of the Academy of the International Council for Small Business, Associate Editor of the Journal of Translational Engineering in Health and Medicine, and Senior Member of the IEEE.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

RENÉ BASTÓN

René Bastón has over 20 years of experience in innovation strategy, business development, economic development, technology commercialization, and venture startup. He is an instructor of the lean startup methodology for National Science Foundation entrepreneurs via its Innovation Corps program, and is a member of the External Board of Advisors for the Columbia University Center for Advanced Information Management.

René has applied lean principles directly as the co-founder of 3 startups – REAL, a B2B cloud-based marketing analytics company transforming marketing and e-commerce at the intersection of recommendation technologies and neuropsychology; HVBio, employing biotechnology and self-assembly methods to enable single-walled carbon nanotube devices, and Medihub, an early Health IT company.

He has also served as the Director of Industry Interactions and Entrepreneurship at the Columbia University Institute for Data Science and Engineering (IDSE), where his role was to cultivate relationships with industry partners and develop the Institute's entrepreneurial ecosystem while integrating it with Columbia's, the City's, and beyond. Prior to the IDSE, René was the Chief Business Officer at the New York Academy of Sciences, where he ran NYAS' domestic and international business and economic development initiatives; and an Associate Director at Columbia's Science; Technology Ventures, one of the world's leading academic technology commercialization organizations. René also spent several years performing neurobiology research in the laboratory of Nobel Laureate, Eric Kandel, at the Columbia University Center for Neurobiology and Behavior. He earned both his Master's in Biomedical Informatics and his B.A. from Columbia University.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

DAVID G. BELANGER, PHD

David Belanger is currently a Senior Research Fellow at Stevens Institute of Technology. He continues his work in Big Data Technology, Applications, and Governance, and is a leader in the Business Intelligence & Analysis Masters Degree program, He also leads the IEEE Big Data Initiative (bigdata.ieee.org).

Dr. Belanger retired as Chief Scientist of AT&T Labs, and Vice President of Information, Software, & Systems Research. He created the AT&T InfoLab, a very early (1995) participant in "Big Data" research and practice. Prior to that, he led the Software Engineering Research Department at Bell Labs. He holds a PhD in Mathematics from Case Western Reserve University.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

ARI BROOKS, MD

Dr. Brooks is a Professor in the Department of Surgery at the Perelman School of Medicine at the University of Pennsylvania. He is the Chief of Endocrine and Oncologic Surgery and Director of the Integrated Breast Center at Pennsylvania Hospital. He is passionate about access to healthcare for all, and has started and run two programs to provide access to breast screening and treatment for the uninsured in Philadelphia. He is very involved with surgical education, and just completed a Master's Degree in Medical Education in the Graduate School of Education at the University of Pennsylvania. His research interests lie at the intersection of engineering, surgery and oncology. He is a member of multiple interdisciplinary research teams and has assisted with multiple small company startups. Dr. Brooks is an author of multiple patents for surgical devices and has run pre-clinical and clinical trials with many new devices.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

ASHOK CHENNURU

Ashok Chennuru is a staff vice president within the office of the chief technology officer, where he leads Enterprise Information Architecture and Health Information Technology areas in WellPoint. He has been with WellPoint for six years. Mr. Chennuru is also the IT lead for the Clinical Strategy working with Comprehensive Health Solutions group. He has over 20 years experience in IT systems development and Implementation with focus on enterprise architecture, information management, health information technologies (HIT), data architecture, business intelligence, healthcare analytics, program management and systems engineering disciplines. Prior to WellPoint, Mr. Chennuru managed the Enterprise Information Management group at Liberty Mutual Insurance. He also worked as software product development engineer, consultant, and manager at the Oracle and SAP corporations. He has deep domain expertise in the healthcare (payer and provider), property and casualty, and financial services industries. Mr. Chennuru has a master's degree in computer science from the University of Missouri.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

SOON AE CHUN, PHD

Soon Ae Chun is a professor and coordinator of the Information Systems and Informatics program at the College of Staten Island and a doctoral faculty member of Computer Science at the Graduate Center, the City University of New York. She is the director of the Information Security Research and Education Lab (iSecure Lab) sponsored by NSF. She is the recipient of the 2014 CSI President's Dolphin Award for Outstanding Scholarly Achievement by a Member of the Full-time Faculty. She is currently the President-Elect for the Digital Government Society. She was a visiting scholar at the Columbia University Network Security Lab and a visiting research professor at CIMIC at Rutgers University. Dr. Chun's research interests include Security and Privacy, Semantic Web, Data Integration, Social Data Analytics and Workflow. Dr. Chun's current research projects include building a social media data integration and analytics platform for public healthcare, and developing Security Learning tools using a cyber-security ontology. Her research has been sponsored by NSF, NOAA and New Jersey State Agency and CUNY PSC. She is a senior member of IEEE & Computer Society, a member of the Association for Computing Machinery (ACM), the Digital Government Society, and the Beta Gamma Sigma National Business Honor Society.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

ATAM P. DHAWAN, PHD

Atam P Dhawan is Distinguished Professor of Electrical & Computer Engineering and Vice Provost for Research and Development at NJIT. Dr. Dhawan is an elected Fellow of the National Academy of Inventors (NAI), Fellow of the Institute of Electrical and Electronics Engineering (IEEE), and Fellow of the American Institute of Medical and Biological Engineering (AIMBE) for his contributions in medical imaging and image analysis, and healthcare innovations. Recently, he has been elected to be a Fellow of the International Academy of Medical and Biological Engineering (IAMBE). He has published over 215 research papers and book chapters. He has also authored and co-authored several books in medical imaging, and image analysis. He is a recipient of numerous awards including Martin Epstein Award (1984), NIH FIRST Award (1988), Sigma-Xi Young Investigator Award (1992), IEEE EMBS Early Career Achievement Award (1995), Doermann Distinguished Lecture Award (1999) and EMBS Distinguished Lecturer award (2012-2013). He has served as the Conference Chair of the IEEE 28th International Conference of Engineering in Medicine and Biology Society, New York (2006). He has also served as the Senior Editor of the IEEE Transactions on Biomedical Engineering, Editorial Board Member for International Journal of Pattern Recognition, and steering committee member for IEEE Transactions on Medical Imaging. Currently, Dr. Dhawan serves as the Editor-In-Chief of the IEEE Journal of Translational Engineering in Health and Medicine. He is the founding chair of the IEEE EMBS technical committee on Translational Engineering and Healthcare Innovations. He has organized and chaired the IEEE-EMBS International Conferences on Point-of-Care Technologies and Healthcare Innovation in Bangalore, India (2013), and in Seattle (2014), and co-chaired the NIH-IEEE Strategic Conference on Point-of-Care Technologies for Precision Medicine held at the NIH NIAID Conference Center, Bethesda in 2015. He is the conference chair of NIH-IEEE Conference on Healthcare Innovation and Point-of-Care Technologies to be held in Cancun on November 9 through November 11, 2016.

Dr. Dhawan has chaired numerous NIH special emphasis and review panels including the NIH Chartered Study Section on Biomedical Computing and Health Informatics (2008-11). His research interests lie in medical imaging, medical image analysis, point-of-care technologies, pattern recognition and computer-aided-diagnosis.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics
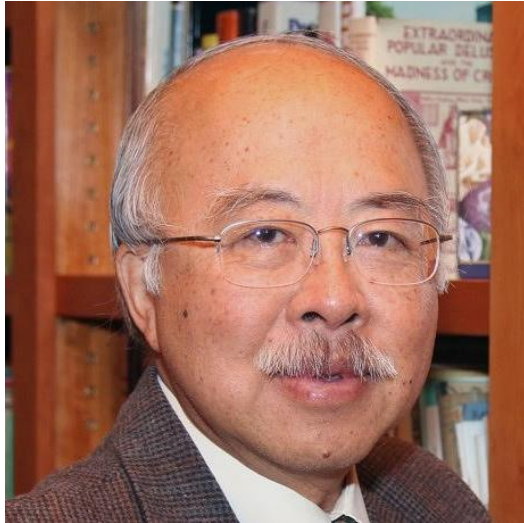
JACK FONG, MD

Fong is the School Medical Advisor of Danbury Public Schools and New Fairfield Public Schools. He is a Member of the Board of Directors of the Leir Retreat Center, Ridgefield and Advisory Board of Directors of the Families Network of Western Connecticut, Danbury and Immaculate High School, Danbury. Dr. Fong is a Fellow of the Royal College of Physicians and Surgeons of Canada, a Fellow of the American Academy of Pediatrics and a Fellow of the New York Academy of Medicine. He is also a member of the American Association of Immunologists and Clinical Immunology Society. He was formerly the Chairman (1985-2008), Department of Pediatrics at Danbury Hospital, Danbury.

For college education, Dr. Fong entered McGill University, Montréal, Québec, Canada and was awarded the B.Sc. degree in Honors Biochemistry in 1964 and MD, CM and MSc degrees in 1968. Following a year of Rotating Internship at St. Joseph Hospital of the University of Western Ontario, London, Ontario, Canada, he went to Minneapolis where he completed a Pediatric Residency and a Fellowship in Clinical Immunology at the University of Minnesota. In 1974, he returned to Montréal and was appointed as an Assistant Professor in Paediatrics at McGill University. He was promoted to the position of Associate Professor of Paediatrics in 1977, a position he held until 1985 when he became the Chairman of Pediatrics at Danbury Hospital. From 1996 to 2012, he was an Associate Clinical Professor Pediatrics at Yale University School of Medicine and from 1998 to 2012, he was a Clinical Associate Professor of Pediatrics at New York Medical College.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

SHRIDAR GANESAN, MD, PHD

Shridar Ganesan came to The Cancer Institute of New Jersey in 2005 from the Dana-Farber Cancer Institute/Harvard Medical School where he was an instructor in medicine and staff physician. Shridar started his medical training at Yale University, where he graduated from with both an MD and a PhD in cell biology. He then completed a medical residency at the Brigham and Women's Hospital in Boston and served as a Chief Medical Resident. This was followed by a fellowship in medical oncology at the Dana-Farber Cancer Institute.

With a research interest in breast cancer biology and DNA repair, Shridar is currently exploring how DNA repair defects in cancers can be exploited to develop novel effective treatments. He is also active in applying next-generation sequencing technology to identify specific genomic changes in cancers that can be therapeutically targeted. As a physician/scientist he both runs a basic research laboratory focused on breast cancer biology and sees patients in the Stacy Goldstein Breast Cancer Center. In the clinic, he works collaboratively with experts across multi-disciplines and has the opportunity to put theory into practice as they aim to develop the next generation of targeted treatments for breast cancer. Working with a team of radiation oncologists, surgical oncologists, nurses, social workers, genetic specialists and others, he helps patients understand their specific disease and their treatment options so that they can make informed decisions. He is also an associate professor of medicine and pharmacology at Robert Wood Johnson Medical School.
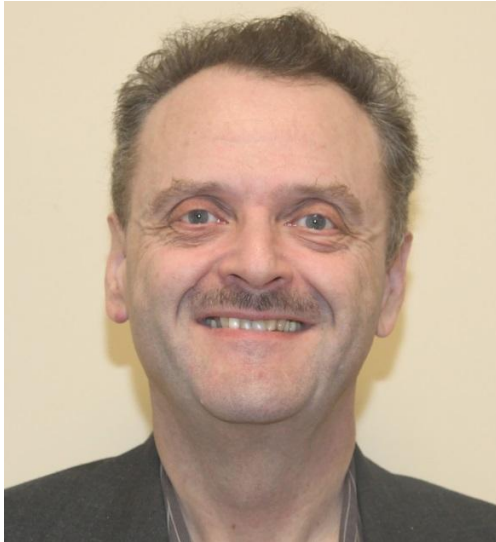
He is the author or co-author of more than 40 publications and serves on the editorial board of the journal *Frontiers in Molecular and Cellular Oncology.*

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

JAMES GELLER, PHD

Dr. James Geller is Professor, Associate Dean for Research of the Ying Wu College of Computing Sciences and Chair Emeritus of the Computer Science Department at the New Jersey Institute of Technology. Dr. Geller received his PhD from the State University of New York at Buffalo in 1988. He cofounded the Medical Informatics Laboratory ("SABOC") and founded the Semantic Web Laboratory at the Computer Science Department. He has published over 150 journal and conference papers and 12 book chapters in Medical Informatics, Semantic Web Technology, Object-Oriented Database modeling and Knowledge Representation.

Between 2006 and 2012 Dr. Geller was Co-Principal Investigator on several federal grants, totaling over $2,500,000, on auditing methods, abstraction algorithms and software tools for important medical terminology systems such as the Unified Medical Language System (UMLS) and the Systematized Nomenclature of Medicine (SNOMED). Currently he is co-PI on a $1.75 million NIH grant extending this previous research to families of ontologies. In the fall of 2012, Dr. Geller was inducted as a Fellow of the American College of Medical Informatics (ACMI). He has received the NJIT Master Teacher Designation (2005) and three other teaching awards, including the award for Professional Development in Teaching (2011) for introducing Medical Informatics as Computer Science teaching subject at the Ying Wu College of Computing Sciences. Dr. Geller also received an NJIT College Research Award (2010). In 2014, Dr. Geller received a three year BRAID grant for the promotion of Women in Computing, from the Anita Borg Institute.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

JAMES R. GOMES, JD

Jim Gomes is the Director of the Mosakowski Institute for Public Enterprise at Clark University. He began his career as a lawyer at the Boston firm of Hale and Dorr and as a Massachusetts Assistant Attorney General. He left law practice to become Chief of Staff to the newly-elected Massachusetts Lieutenant Governor John Kerry. Jim then served as the Policy Director of Mr. Kerry's successful first campaign for United States Senate and as Executive Assistant to Senator Kerry.

Jim was Massachusetts' Undersecretary of Environmental Affairs from 1989-91, where he directed work on all major policy issues. He was the chief representative of the administration of Governor Michael Dukakis in negotiations with the legislature, industry, and citizens' groups that led to the enactment of Massachusetts' landmark Toxics Use Reduction Act. In 1993, Jim became President of the Environmental League of Massachusetts (ELM), a leading policy and advocacy organization. During his tenure at ELM, Harvard Law School named him a Wasserstein Fellow for his "outstanding contribution to public interest law." In 2007, Governor Deval Patrick appointed him Chairman of the Massachusetts Environmental Trust, the commonwealth's public environmental philanthropy. Jim became the first Director of the Mosakowski Institute in 2007. The Institute seeks to mobilize university research to address important societal issues and to enrich the education of students by providing uncommonly valuable and transformative experiences. At Mosakowski Jim oversees a portfolio of research and community engagement projects in a wide array of subjects including education, health and well-being, urban economic development, and the environment. An expert on the politics of policy implementation, Jim has consulted with state governments on improving their programs and has taught or lectured at several colleges and universities including Brandeis, Clark, Harvard, Trinity, Tufts, Wellesley, and Williams. He holds a B.A. in Political Science from Trinity College, a Master's Degree in Public Policy from Harvard's Kennedy School of Government, and a J.D. from Harvard Law School.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

SHANTHI GOPALAKRISHNAN, PHD

Shanthi Gopalakrishnan is a Professor of Management at the Martin Tuchman School of Management, at NJIT, New Jersey. he received her Ph.D. in Organization Management from Rutgers University. Her research interests are in the area of innovation management, strategic alliances and knowledge management. he has published over 50 articles on these topics. Dr. Gopalakrishnan is a member of the Academy of Management and was the President of Eastern Academy of Management and currently a Fellow of the EAM.

he has received several Best Research Paper Awards, from the Eastern Academy of Management in 1995, Academy of Management in 1997 and 2009, and from IEEE Transactions in Engineering Management in 2001. he currently serves on the Editorial Boards of IEEE Transactions in Engineering Management, Group and Organization Management and on the advisory board of Organization Management Journal.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

KATHY GRISE

Kathy Grise, IEEE Future Directions Program Director, works directly with IEEE volunteers, IEEE staff, and consultants in support of new initiatives, and is the IEEE staff program director for the Big Data Initiative and the IEEE Technology Navigator, and manages the digital presence team for Future Directions. Prior to joining the IEEE staff, Ms. Grise held numerous positions at IBM, and most recently was a Senior Engineering Manager for Process Design Kit Enablement in the IBM Semiconductor Research and Development Center. Ms. Grise lead the overall IT infrastructure implementation, and software development in support of semiconductor device modeling verification, packaging, and delivery; device measurement and characterization data collection and management, and automation for device modeling engineers using cloud services. Ms. Grise is a graduate of Washington and Jefferson College, and an IEEE Senior member.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

BRADFORD W. HESSE, PHD

Bradford (Brad) Hesse is Chief of the National Cancer Institute's Health Communication and Informatics Research Branch. Dr. Hesse received his degree in social psychology from the University of Utah in 1988 with an accompanying internship in the nascent field of medical informatics. After completing his degree, he served as a postdoctoral fellow within the Department of Social and Decision Sciences at Carnegie Mellon University. For more almost three decades since that time, he has been conducting research in the interdisciplinary fields of social cognition, health communication, health informatics, data science, and user-centered design. Dr. Hesse was recruited to the National Cancer Institute in 2003 and has since been focusing his energies on bringing the power of evidence-based health communication to bear on the problem of eliminating death and suffering from cancer. He continues to direct the Health Information National Trends Survey, a biennial general population survey aimed at monitoring the public's use of health information during a period of enhanced capacity at the crest of the information revolution. He has authored or co-authored over 170 publications, including peer-reviewed journal articles, technical reports, books, and book chapters. In 2009, his coauthored book titled "Making Data Talk: Communicating Public Health Data to the Public, Policy Makers, and the Press" was named Book of the Year by the American Journal of Nursing. In newest edited volume, "Oncology Informatics: Using Health Information Technology to Improve Processes and Outcomes in Cancer" to be published by Elsevier in 2016, has brought together some of the best and brightest thinkers to create a blueprint for accelerating progress against cancer through the health information technologies being woven into the modern care system.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

GEORGE HRIPCSAK, MD, MS

George Hripcsak, MD, MS, is Vivian Beaumont Allen Professor and Chair of Columbia University's Department of Biomedical Informatics and Director of Medical Informatics Services for NewYork-Presbyterian Hospital/Columbia Campus. He is a board-certified internist with degrees in chemistry, medicine, and biostatistics. Dr. Hripcsak's current research focus is on the clinical information stored in electronic health records and on the development of next-generation health record systems. Using nonlinear time series analysis, machine learning, knowledge engineering, and natural language processing, he is developing the methods necessary to support clinical research and patient safety initiatives. He leads the Observational Health Data Sciences and Informatics (OHDSI) coordinating center; OHDSI is an international network with 160 researchers and 600 million patient records. He co-chaired the Meaningful Use Workgroup of U.S. Department of Health and Human Services's Office of the National Coordinator of Health Information Technology; it defines the criteria by which health care providers collect incentives for using electronic health records. He led the effort to create the Arden Syntax, a language for representing health knowledge that has become a national standard. Dr. Hripcsak chaired the U.S. National Library of Medicine's Biomedical Library and Informatics Review Committee, and he is a fellow of the National Academy of Medicine, the American College of Medical Informatics, and the New York Academy of Medicine. He has published over 250 papers.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

JIANYING HU, PHD

Jianying Hu is a Distinguished Research Staff Member and Program Director of Center for Computational Health at IBM T. J. Watson Research Center, NY. Prior to joining IBM in 2003 she was with Bell Labs at Murray Hill, New Jersey. Dr. Hu received the Ph.D. degree in Computer Science from SUNY Stony Brook in 1993. Her main research interests include machine learning, data mining, statistical pattern recognition, and signal processing, with applications to healthcare analytics and medical informatics, business analytics, document analysis, and multimedia content analysis. In the past 7 years her focus has been exclusively on healthcare, leading an increasingly large research team to develop advanced machine learning, data mining and visual analytics methodologies for deriving data-driven insights from real world healthcare data to facilitate "learning health systems". Dr. Hu has published over 100 technical articles and holds 27 patents. She has served as associate editor for IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, Pattern Recognition, and International Journal for Document Analysis and Recognition, and is currently on the advisory board of the Journal of Healthcare Informatics Research. Dr. Hu is a fellow of IEEE (class of 2015), a fellow of the International Association of Pattern Recognition (class of 2010), and a recipient of the Asian American Engineer of the Year Award (class of 2013).

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

ZACHARY IVES, PHD

Zachary Ives is a Professor of Computer and Information Science at the University of Pennsylvania, as well as the Associate Dean for Masters and Professional Programs at Penn's School of Engineering and Applied Science. He is a co-founder of Blackfynn, Inc., a company focused on enabling life sciences research and discovery through data integration. His research interests include data integration and sharing, managing "big data," sensor networks, and data provenance and authoritativeness. He is a recipient of the NSF CAREER award, and an alumnus of the DARPA Computer Science Study Panel and Information Science and Technology advisory panel. He has also been awarded the Christian R. and Mary F. Lindback Foundation Award for Distinguished Teaching. He serves as the Director for Penn's Singh Program in Networked and Social Systems Engineering, and he is a Penn Engineering Fellow. He is a co-author of the textbook *Principles of Data Integration*, and received an ICDE 2013 ten-year Most Influential Paper award. He has been an Associate Editor for Proceedings of the VLDB Endowment (2014) and a Program Co-Chair for SIGMOD (2015).

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

WILLIAM MCDERMOTT

Billy McDermott is the Director of Development at New Jersey Institute of Technology (NJIT). He is responsible for development of the Martin Tuchman School of Management (SOM) and Ying Wu College of Computing Sciences (CCS) (largest in region). Billy's strength is building mutually beneficial partnerships with Sr Executives from large corporations. Billy is responsible for both the Boards at SOM and CCS. A cross section of companies include: Juniper Networks, Novartis Oncology, J&J, MDC Partners, Intel, GENBAND, Alcatel Lucent, TATA, Enterprise Rent a Car, Huawei, EMC, S&P Capital IQ, VMware, TD Bank, Capital One, Merck, Lockheed Martin, Frontier Communications, GE Billy acts as principal building innovative programs including a new Corporate Affiliates Program (see attached below) (McGraw Hill Financial is a founding member); STEP (see PR below), a Saturday High School program, which identifies top students early interested in Science, Technology, Engineering, and Math (STEM), and is currently 2x over-subscribed (The Berger Group is the sponsor); many tailored programs suited to corporate needs (see attached S&P CapIQ Big Data Visualization program); Student labs with equipment from sponsoring companies; a Summer Pre-College Financial Literacy program (see PR below); the Innovation Acceleration Center (see PR below) and many Hack-a-thons (see PR below) and Capstone projects -- too numerous to mention. Billy was responsible for bringing Healthcare leaders to campus, resulting in the award of a $23 million White House grant and the creation of NJ HiTEC (see PR below) and the NJ Healthcare Innovation Center. NJIT, NJ's science and technology University, enrolls approximately 10,000 students pursuing bachelor's, master's and doctoral degrees in 120 programs. The university consists of six colleges: Newark College of Engineering, College of Architecture and Design, College of Science and Liberal Arts, Martin Tuchman School of Management, Ying Wu College of Computing Sciences and Albert Dorman Honors College.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

VINCENT ORIA, PHD

Vincent Oria received a diplôme d'ingenieur from the Institut National Polytechnique Houphouet-Boigny (formerly INSET) in Yamoussoukro, Côte d'Ivoire(Ivory Coast), in 1989 and a Ph.D. in computer science from the Ecole Nationale Supérieure des Télécommunications (Telecom-Paritech), Paris, France, in 1994. His research interests include multimedia databases, spatio-temporal databases and Recommender Systems.

From 1994 to 1996, Vincent worked as a researcher at ENST, Paris; from 1996 to 1999, he was a post-doctoral fellow at the University of Alberta, Edmonton, Canada. From January 2000 to May 2006, he was an assistant professor of computer science at the New Jersey Institute of Technology and since June 2006 I am an Associate Professor in the same department. Vincent has held visiting professor positions at various institutions including National Institute of Informatics (Tokyo, Japan), ENST (Paris, France), Université de Paris-IX Dauphine (Paris, France), INRIA (Roquencourt, France), CNAM (Paris, France), Chinese University of Hong Kong (Hong Kong China) and the Université de Bourgogne (Dijon, France). He is an associate editor the journals Multimedia Tools and Application (MTAP), the International Journal of Multimedia Data Engineering and Management (IJMDEM) and International Journal of Multimedia Information Retrieval. Vincent has served on a number of multimedia and database conference program committees including ACM Multimedia (MM), ACM World Wide Web (WWW) and IEEE International Conference on Data Engineering (ICDE).

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics
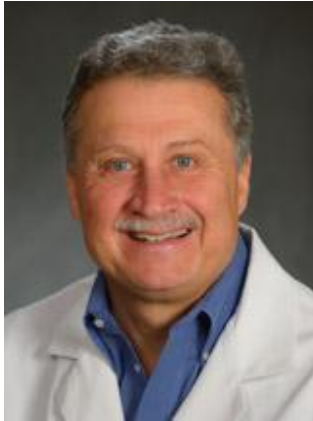
REYNOLD A. PANETTIERI, JR, MD

Reynold Panettieri is the Director of the Institute for Translational Medicine and Science and Vice Chancellor for Translational Medicine and Science at Rutgers University, and the former Director of the Airways Biology Initiative at the University of Pennsylvania. His interests are in the cellular and molecular mechanisms that regulate airway smooth muscle cell growth and the immunobiology of airway smooth muscle. Consequences of increases in airway smooth muscle growth promote the development of irreversible airflow obstruction and airway remodeling seen in patients with chronic severe asthma. Dr. Panettieri's lab also focuses on cytosolic signaling pathways that mediate gene expression and alter myocyte function.

Dr. Panettieri also served as the Deputy Director of the Center of Excellence in Environmental Toxicology at the University of Pennsylvania. He directed the human exposure chamber that defines the molecular mechanisms regulating ozone- and particulate matter-induced airway hyperresponsiveness. In parallel with his basic science interests, Dr. Panettieri managed the comprehensive clinical program for the care of patients with asthma and is actively involved in clinical investigations focused on the management of asthma and COPD.

In addition to his research and clinical interests, Dr. Panettieri served as chairperson of the NIH Lung Cellular, Molecular, and Immunobiology Study Section, is a member of the NIH Distinguished Editorial Panel, and is a member of the American Society for Clinical Investigation and Association of American Physicians.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

RAM RAMESH, PHD

Ram Ramesh's research spans economics of IT, conceptual modeling and database systems, operations and decision analysis. In particular, his contemporary works deal with predictive analytics of availability-aware cloud computing and high performance computing systems, design of service-level contract mechanisms in cloud computing markets, and predictive analytics of health information exchanges. Methodologically, his research is established in predictive modeling, mathematical programing and stochastic optimization. His research has been funded by the National Science Foundation, Google, Samsung, Raytheon and Westinghouse, besides various U.S. military research programs including U.S. Army Research Institute, U.S. Air Force Office of Scientific Research, U.S. Air Force Research Laboratory and the U.S. Naval Training Systems Center.

Ramesh serves as an area editor for the machine learning area of INFORMS Journal on Computing and is a founding co-editor-in-chief of Information Systems Frontiers (published by Springer). He also serves on the editorial advisory board, Information Technology and Systems Network of Social Science Research Network (SSRN). Ramesh served as an editor of two volumes of the Annals of Operations Research on the theme "Interface between Operations Research and Information Systems." He also edited a special issue of the Journal of AIS on "Ontologies in the context of Information Systems" and an issue of CACM on "Distributed Mission Training Technologies: Emerging Frontiers in Training Systems." He has edited two research books, "Ontologies - A Handbook of Principles, Concepts and Applications in Information Systems" (published by Springer) and "National Security Research: Technological, Organizational and Policy Perspectives" (published by Elsevier).

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

JULIA STOYANOVICH, PHD

Dr. Stoyanovich's research is in the area of data and knowledge management. She focuses on developing novel information discovery approaches, with the goal of helping the user identify relevant information, and ultimately transform that information into knowledge. She is particularly excited about the challenges that arise in life sciences applications and in social information processing - domains that are experiencing unprecedented growth in the amount and richness of available data. Having graduated from college, Dr. Stoyanovich spent 5 years in the start-up industry, as a software developer, data architect and database administrator. This experience has motivated her to work with real datasets whenever possible, and to deliver results of her research to the communities of target users, as part of open-source systems or as stand-alone prototypes. Dr. Stoyanovich teaches courses in the data management curriculum at CCI.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

CHEICKNA SYLLA, PHD

Cheickna Sylla earned his PhD in Operations Research from the Department of Industrial and Systems Engineering of the State University of New York at Buffalo. His research interests include decision support systems and decision sciences techniques (which comprise applied operations research and statistics, business intelligence and data mining), analysis and evaluation of management information systems, human machine systems, supply chain management and distribution logistics. Dr. Sylla is the author of over 85 peer reviewed journal and conference articles. His publications appear in European Journal of Operations Research, IEEE Transactions on SCM, IEEE Transactions on Engineering Management, Computers and Industrial Engineering, International Journal of Technology Management, International Journal of Environmental Agreements, International Journal of Networking and Virtual Organizations, International Journal of Information Technology and Management, International Academy of Business and Economics, Human Factors, Manufacturing Review, Control Engineering Practice, Computers in Industry, Cost Management, and others. He has won 3 best paper awards at international conferences. Additionally, he was an invited keynote and panel speaker at several international conferences throughout the years. He was also a Guest Editor of the International Journal of Management and Decision Making and the International Journal of Information Technology and Management in 2010. With respect to Data Mining (DM) and Healthcare, Dr. Sylla's interests include (1) investigations of work- related injuries toward the development of Ergonomics Improvement Programs in Workplaces (producing a Ph.D. thesis in 2001 and several related publications), and (2) development of a comprehensive neural network based data analysis framework to analyze patient's historical annual-checkup records using a hybrid model, which combines Cumulative Sum Control Charts (based on Statistical Process Control theory) and the 100% Rule (from Optimization theory) to investigate simultaneous variations in signals and noises in health parameters. As Healthcare providers and insurers are increasingly turning to big data and analytics, to help them understand their patients and the contexts of their illnesses costs and practice, the goal of this hybrid model is to devise DM based analysis procedures to track the slow growing cancer and heart diseases which are likely to remain undetected by any single annual check-up evaluation. Dr. Sylla was involved in advising and/or co-advising 8 Ph.D. dissertations and 15 Master of Science theses. His teaching areas are closely related to his research interests. He is currently a Professor of Decision Sciences and MIS and the Associate Dean for Academic Affairs of NJIT Martin Tuchman School of Management.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

DIMITRI THEODORATOS, PHD

Dimitri Theodoratos is currently an associate professor in the Department of Computer Science at the New Jersey Institute of Technology (NJIT). He received the diploma in Electrical and Computer Engineering from the National Technical University of Athens, Greece, the MSc degree from the Ecole Nationale Supérieure de Telecommunications of Paris, France, and the PhD degree from the University of Paris at Orsay, France, both in computer science. His research interests include database management, big data, data integration, data warehousing, query processing and optimization, keyword search, semi-structured data, pattern mining, medical informatics and the Semantic Web. He has published over 75 papers in conference proceedings and journals in those areas. He has advised or co-advised 8 PhDs and is currently supervising two PhD students. His work has been funded by the NSF, the DoD, Research Army Lab, the European Union, the IBM and the IBM Center of Advanced Studies (CAS). He has taught at the University of Ioannina and the National Technical University of Athens in Greece, at the University of Paris XI, in France, and at NJIT. He holds the Master Teacher Designation—the highest distinction given by NJIT in recognition of excellence in teaching—and he is also the recipient of an excellence in teaching award in graduate instruction from NJIT and an IBM faculty award.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics
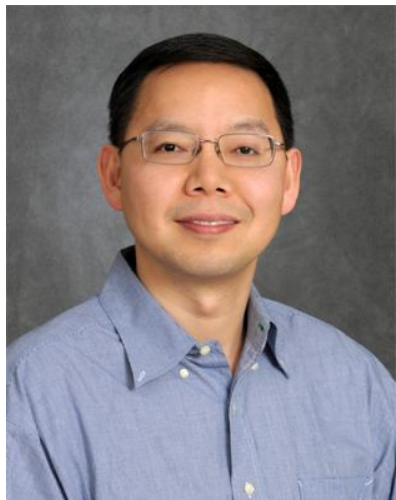

FUSHENG WANG, PHD

Fusheng Wang is an assistant professor at Department of Biomedical Informatics and Department of Computer Science at Stony Brook University. He received his Ph.D. in Computer Science from University of California, Los Angeles in 2004, and his M.S. and B.S. in Engineering Physics from Tsinghua University, China, in 1997 and 1994 respectively. Prior to joining Stony Brook University, he was an assistant professor at Emory University. He was a research scientist at Siemens Corporate Research from 2004-2009.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics
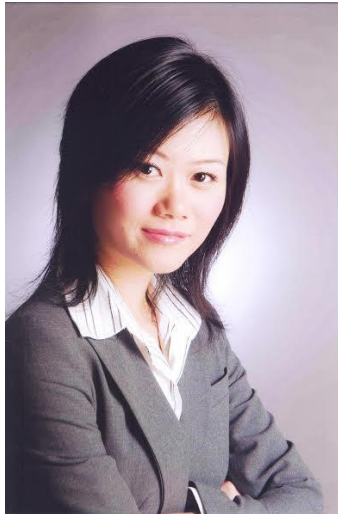
HENG XU, PHD

Dr. Heng Xu is an associate professor of Information Sciences and Technology at the Pennsylvania State University. She is the founder and director of Privacy Assurance Lab, an interdisciplinary research group working on a diverse set of projects with the overall goal of understanding and assuring privacy in different contexts, including location-based services, online social networks, medical practices, and children and adolescent online safety. During 2013-2016, Dr. Xu served as a program director in the Social and Economic Sciences Division of the Social, Behavioral & Economic Sciences (SBE) directorate of the National Science Foundation (NSF) during 2010-2013. Her main responsibility at NSF focuses on facilitating novel interdisciplinary research between Computer and Information Science and Engineering (CISE) and SBE, to address major challenges in Big Data and Cybersecurity.

Dr. Xu has authored and co-authored over 100 research papers on information privacy, security management, human-computer interaction, and technology innovation adoption. Her work has been published in premier outlets across various fields such as Information Systems, Law, Computer Science, and Human-Computer Interaction, including MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Decision Support Systems, University of Pennsylvania Journal of Constitutional Law, Proceedings of the International World Wide Web Conference (WWW), Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI) and many others. Dr. Xu was a recipient of an NSF Career award (2010) and the endowed PNC Technologies Career Development Professorship (2010-2013). She has received a number of best paper nominations and awards at many conferences.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

DANTONG YU, PHD

Dantong Yu received the BS degree in Computer Science from Beijing University, China in 1995 and the Ph.D. degree in Computer Science from State University of New York at Buffalo, USA, in 2001. He joined Brookhaven National Lab in 2001, and coordinated the Grid computing group at BNL. He currently leads the computer science group at the Computational Science Center. His research interests include high-speed network performance, network qualify of service, cluster/grid Computing, information retrieval, data mining, database, and data warehouse. He designed and implemented a novel high-dimensional indexing algorithm (termed ClusterTree) using the semantics of datasets. He has published papers in leading technical journal and conferences. He served in the review Panels of NSF CDI, DOE Early Career Principle Investigator for networking research and DOE Small Business Innovative Research (SBIR) and co-chair of several DOE Advanced Networking Workshops for Distributed Petascale Science.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

NAN ZHANG, PHD

Dr. Nan Zhang is an Associate Professor of Computer Science at the George Washington University, Washington, DC. He is currently on leave at the National Science Foundation, serving as a Program Director in the Information and Intelligent Systems Division. He received the B.S. degree from Peking University in 2001 and the Ph.D. degree from Texas A&M University in 2006, both in computer science. His current research interests include databases and information security/privacy. He received the NSF CAREER award in 2008.

Symposium on Data Science for Healthcare (DaSH)

Data Sharing: The Art, Science, and Ethics

JUDY ZHONG, PHD

Dr. Zhong is an Assistant Professor of Biostatistics at New York University School of Medicine. She has developed statistical approaches to correct the bias in odds ratios from genome-wide association (GWAS) studies and approaches that leverage information from genetics of gene expression studies to identify biological pathways enriched for expression-associated genetic loci associated with disease in GWAS results. Dr. Zhong has a demonstrated record of publications in developing novel statistical methods and has also collaborated on various scientific research projects.

Her research interest is to develop powerful statistical methods for data generated from next generation sequencing technology and to use causal inference framework to integrate a diversity of molecular and clinical trait data to uncover models that predict complex system behavior. She received her PhD. from the Department of Biostatistics at the University of Washington.