Enabling Seamless Data Sharing in Industry and Academia

Northeast Big Data Innovation Hub Workshop
Gerri C. LeBow College of Business, Drexel University
3220 Market Street Philadelphia, PA 19104
Meeting Room 220/Rose Terrace 221

September 29, 2016, 8:00 AM-5:00 PM

8:00 - 9:00   **Check-in, hot breakfast**

9:00 - 9:30   **Welcome, Introductions and Overview of the Day**
- Welcome to Drexel and Introductions - Jane Greenberg, Alice B. Kroeger Professor, Director of the Metadata Research Center, College of Computing & Informatics, Drexel University
- Welcome to Drexel - Yi Deng, Isaac L. Auerbach Professor and Dean, College of Computing and Informatics
- Northeast Big Data Hub Introduction - René Bastón, Executive Director, NEBDIH
- Goals of the BD Sharing Spoke and Workshop goals - Sam Madden, Professor and Director of Computer Science and Artificial Intelligence Laboratory (CSAIL), MIT, and Tim Kraska, Assistant Professor, Brown University)

9:30 - 10:50   **Session 1: War Stories Part I** (Session chair: Florence Hudson)
Robert Cheetham, Founder and President, Azavea
Jason Bobe, Director of the Sharing Lab, Icahn Institute and Dept. of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai
Joe Chaya, Director – Project Management, Global Data Warehousing, IMS Health
John Lee, Senior Associate, Osage Partners

10:50 - 11:20   **Break**

11:20 - 12:00   **Session 2: IP, Policy, Legal issues** (Session chair: Tim Kraska)
Anita Eisenstadt, Research Staff Member, Science and Technology Policy Institute
Paul Ragusa, Intellectual Property Law Partner, Baker Botts

12:00 - 12:20   **Introduction of everybody** (Jane Greenberg)

12:30 - 1:45   **Brainstorming Lunch** (Session chair: Sam Madden)
Topic: What are the biggest challenges in sharing data and how can the Sharing Spoke help? Ideas include: Personal Identifiable Information (PII), Restriction of use, Access control, Training, Contribution, Result Sharing, Accountability, etc.
Goal: every group of 5 creates a ranked list of issues. The issues and potential solutions are entered in a Google Doc. Furthermore, every group has to pick its top issue and present it (next session)

1:45 - 2:30   **Result Presentation: Top Issues** (5 min per group)
Every group presents its top issue and a potential solution.

2:30 - 3:10   **Infrastructure Challenges and Opportunities** (Session chair: Jane Greenberg)
Greg Madden, Senior Advisor for Research Computing and Cyberinfrastructure, Penn State
Sam Madden, Professor and Director of CSAIL, MIT

3:10 - 4:10   **Breakout and Coffee: Requirement Gathering** (Session: Tim Kraska)
Participants meet at different topic tables derived from topics identified at lunch.
Potential Topics: Personal Identifiable Information (PII), Restriction of use, Access control, Training, Contribution, Result Sharing, Accountability, etc.
Goal: A list of potential options for the data sharing license (e.g., what should the license cover? What is most important?). Each table creates a 5 min presentation per topic. Identify people who are willing to invest in a working group afterwards.

4:10 - 4:40   **Presentation of Breakout Results** (Session: Tim Kraska)

5:00 - 7:00   **Reception**
Landmark Americana
3333 Market Street (5 minute walk from event venue)

Enabling Seamless Data Sharing in Industry and Academia

Northeast Big Data Innovation Hub Workshop
Gerri C. LeBow College of Business, Drexel University
3220 Market Street Philadelphia, PA 19104
Meeting Room 220/Rose Terrace 221

September 30, 2016, 8:00 AM-2:45 PM

8:00 - 9:00	**Check-in, hot breakfast**
9:00 - 9:20	**Introductions**

Jane Greenberg, Welcome to Drexel, introduce Aleister Saunders
Aleister Saunders, Vice Provost for Research, Drexel University
NEBDIH Introduction, René Bastón, Executive Director
Recap of Day 1, Sam Madden

9:20 - 10:20 Welcome, War Stories II (Session chair: Florence Hudson)
John Brzozowski (Fellow and Chief IPv6 Architect, Comcast)
Yana Kane-Esrig (Senior Principal Engineer, Comcast)
Jane Greenberg, Alice B. Kroeger Professor, Director of the Metadata Research
Center, College of Computing & Informatics, Drexel University

10:20 - 11:20 Industrial Requirements (Session chair: Jane Greenberg)
Ian Swanson, CEO and Founder, DataScience, Inc.
Kareem Aggour, Senior Computer Engineer, GE
Tony Orsini, Director of Decision Sciences, Experian

11:20 - 11:40 Break

11:40 - 12:00 City Perspective (Session chair: René Bastón)
Amen Ra Mashariki, Chief Analytics Officer, City of New York

12:00 - 12:40 Breakout II: Data Sharing Licenses (Session chair: René Bastón)
Topics: What problems does a "standard" data license solve?
How big is the licensing problem, in your experience?
What should we do to make a data sharing license succeed?
Goal: Every group list the key success factors

12:45 -1:45 Lunch: Networking and Next Steps

1:45 - 2:45 Open Discussion (Session chair: Sam Madden)
Next steps: Formation of working groups