Example of Big Data in Action with Controls: Big Data and City Living

Lynn A. Goldstein
Chief Data Officer, Center for Urban Science and Progress
December 4, 2014
“Applied Sciences NYC is the City’s unparalleled opportunity to build or expand world-class applied sciences and engineering campuses in New York City. We are seeking to dramatically expand our capacity in the applied sciences to maintain our global competitiveness and create jobs. These campuses would not only enrich the City’s existing research capabilities, but also lead to innovative ideas that can be commercialized, catalyzing hundreds of spinoff companies and increasing the probability that the next high growth company – a Google, Amazon, or Facebook – will emerge in New York City.”

New York City Economic Development Corporation

The NYU-led Center for Urban Science and Progress, a multi-sector research and education collaborative, was announced on April 23, 2012.
Big Cities + Big Data

- The world is urbanizing
- Cities are the loci of consumption, economic activity, and innovation

*Cities are the cause of our problems and the source of the solutions*

- Informatics capabilities are exploding
  - Storage, transmission, analysis
- Proliferation of static and mobile sensors
- Internet of things

Global network traffic, 30% CAGR
What can cities do with the data?

- **Optimize operations**
  - traffic flow, utility loads, services delivery, ...
- **Monitor infrastructure conditions**
  - bridges, potholes, leaks, ...
- **Infrastructure planning**
  - zoning, public transit, utilities
- **Improve regulatory compliance** ("nudges", efficient enforcement)
- **Public health**
  - Nutrition, epidemiology, environmental impacts
- **Abnormal conditions**
  - Hazard detection, emergency management
- **Data-driven formulation of data-driven policies and investments**
  - Low-Income Home Energy Assistance Program, road pricing and congestion charging, ...
- **Better inform the citizenry**
- **Enhance economic performance and competitiveness**
New York City as a Living Lab

The Center for Urban Science and Progress (CUSP) is a unique public-private research center that uses New York City as its laboratory and classroom to help cities around the world become more productive, livable, equitable, and resilient. CUSP observes, analyzes, and models cities to optimize outcomes, prototype new solutions, formalize new tools and processes, and develop new expertise/experts. These activities will make CUSP the world’s leading authority in the emerging field of “Urban Informatics.”
The CUSP Partnership

University Partners
- NYU (multiple schools)
- The City University of New York
- Carnegie Mellon University
- University of Toronto (Canada)
- University of Warwick (UK)
- IIT-Bombay (India)

Industrial Partners
- IBM
- Microsoft
- Xerox
- Cisco, Con Edison, Lutron, National Grid, Siemens
- AECOM, Arup, IDEO

National Laboratories
- Brookhaven
- Lawrence Livermore
- Los Alamos
- Sandia

City & State Agency Partners
- The City of New York
  - Buildings
  - City Planning
  - Citywide Administrative Services
  - Design and Construction
  - Economic Development
  - Environmental Protection
  - Finance
- Metropolitan Transportation Authority
- Port Authority of NY & NJ

Industrial Partners
- IBM
- Microsoft
- Xerox
- Cisco, Con Edison, Lutron, National Grid, Siemens
- AECOM, Arup, IDEO

City & State Agency Partners
- The City of New York
  - Buildings
  - City Planning
  - Citywide Administrative Services
  - Design and Construction
  - Economic Development
  - Environmental Protection
  - Finance
- Fire Department
- Health and Mental Hygiene
- Information Technology and Telecommunications
- Parks and Recreation
- Police Department
- Sanitation
- Transportation
- Metropolitan Transportation Authority
- Port Authority of NY & NJ
Research
Urban Science & Actionable Insight
Urban Data Sources: Acquire, Integrate, Use

Organic Data Flows
- Administrative records (census, permits, …)
- Transactions (sales, communications, …)
- Operational (traffic, transit, utilities, health system, …)
- Social media (Twitter, Facebook, blogs, …)

Sensors
- Personal (location, activity, physiological)
- Fixed *in situ* sensors
- Crowd sourcing (mobile phones, …)
- Choke points (people, vehicles)

Novel Technologies
- Visible, infrared and spectral imagery
- RADAR, LIDAR
- Gravity and magnetic
- Seismic, acoustic
- Ionizing radiation, biological, chemical
- …
Core City Services Include…

**Streets**
- Buildings
- Sanitation
- Planning
- Parks

**Safety**
- Police
- Fire
- Emergency Management

**Customer Service**
- 311
- Business Assistance
- Quality of Life

**Human Services**
- Social Services
- Health
- Education
- Youth
Research Program Structure

<table>
<thead>
<tr>
<th>Disciplines (tbd)</th>
<th>Acquisition</th>
<th>Integration</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data Management, Data Curation, Data Analytics, Visualization, Geographic Information Systems, Machine Learning, Data Mining, Modeling and Simulation, Sensing, …</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disciplines (tbd)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Domains (tbd)

- Transit
- Utilities
- Health
- …

- Project 1
- Project 2
- Project 3
- Project 4
- …

Science “Users”

Citizen Science, Quantified Community, Innovation District

- …
Partnership with NYC Agencies

Meeting with NYC Agency Commissioners, 2012 – 2013
  • Identify priorities and challenges

3-hour meetings at Agencies 2013 - ongoing
  • Establish lines of communication at staff level
  • Introduce CUSP
  • Better understand agency structure, available data, needs

Agency Collaborators (bold indicates meetings that have already occurred)

Department of Buildings
Department of City Planning
**Department of Citywide Administrative Services**
Department of Design and Construction
**Department of Environmental Protection**
Department of Finance
Department of Health and Mental Hygiene
**Office of Long-Term Planning and Sustainability**

Department of Parks and Recreation
Department of Transportation
New York City Fire Department
New York City Housing Authority
New York City Police Department
Taxi and Limousine Commission
Port Authority of New York and New Jersey
2013-2014 Research Workshops

COMPLETED
• Mobile Sensing (March)
• Ethical, Legal & Social Implications (ELSI) (March 2013)
• Persistent Synoptic Phenomenology (April 2013)
• Urban Observatory (April 2013)
• Neuroeconomics (April 2013)
• Noise (April 2013)
• MTA New York City Transit (April 2013)
• Building Energy Efficiency (May 2013)
• Simulation & Modeling (May 2013)
• Citizen Science I (May 2013)
• MTA Long Island Railroad (June 2013)
• Social Media & Peer Networks (July 2013)

• Quantified Community (July 2013)
• Citizen Science II (July 2013)
• Bio Sensing (September 2013)
• Commercialization (November 2013)
• Data Warehouse (February 2014)

UPCOMING
• MTA Metro North (May)
• Citizen Science III (May)
• Urban Sustainability

POTENTIAL
• Data Analysis & Visualization
• Urban GIS

CUSP Infrastructure Agency-Specific Major Project
Projects for the City & State

- Economic Mapping
- Greener Greater Buildings Plan
- MTA Bus Driver Optimization
- MTA Origin/Destination Study
- New York City Police Department
- Parks Attendance & Utilization
- Parks Tree Census
- Property Ownership Records Assessment
- Public Health
- School Property Use Assessment
- Taxi Visualization
- Transit Operations
- and more
Analysis of Massive Taxi GPS Data

Overview
• Data from yellow cabs 2009-2013 is almost 800 million trips; nearly impossible to manage, explore, visualize, and analyze with existing tools

Objective & Goal
• Build scalable, usable tools that can be used by experts and non-experts
• Work with relevant city agencies on development & deployment of the technology

Status
• Initial deployment of TaxiVis at NYC Taxi & Limousine Commission and Department of Transportation
Taxis as City Sensors

NYC Taxi Rides by Day in 2011

NYC Taxi Rides:
- Jan: 590k
- Feb: 29k
- Mar: 590k
- Apr: 29k
- May: 590k
- Jun: 29k
- Jul: 590k
- Aug: 29k
- Sep: 590k
- Oct: 29k
- Nov: 590k
- Dec: 29k

Event Dates:
- April 2
- April 3
- August 28
- Dec 25

Events:
- Irene

Graph shows taxi rides over time, with significant drops on April 2, April 3, August 28, and December 25.
TaxiVis: Interactive Visual Exploration of NYC Taxi Records
CUSP Research Facilities

IN EXPANSION

• CUSP Data Warehouse: data management & science
• CUSP Urban Observatory
• CUSP Quantified Community

IN DEVELOPMENT

• Modeling and Simulation
• Citizen Science: crowdsourced data and community engagement in urban science
• #Social Cities: urban social media
Data Warehouse Facility

Overview

• Omnivorous ingestion to a repository for NYC-related data for urban science and informatics

Objective and Goals

• Make data interoperable, with proper multi-layered access protocols

Data

• Data from City agencies on operations, schedules, maps, etc., including NYC Open Data
• Will include proprietary data, social media data, CUSP-generated data
• Working with the NYC Mayor’s Office of Analytics
• CUSP Chief Data Officer will oversee ethical, legal, and social issues
Urban Observatory

PERSISTENT and SYNOPTIC ANALYTICS for URBAN SCIENCE

The CUSP Urban Observatory

- unique user facility for persistent and synoptic observations of cities with detailed analysis

- seek high impact science and applications to enhance public well-being, city operations, and future urban design

- instrumentation to include both broad band and hyper-spectral from visible to infrared wavelengths

- combine correlative data including administrative records, in situ measurements, topography, etc.
City Lights Project

OBJECTIVES

- Develop a fundamentally new modality for studying the city “from a distance”
- Identify aggregate patterns of light in the time-dependent brightnesses of city lights
- Leverage these patterns into foundational contributions to urban science and urban functioning

IMPACT

Urban Science

- Determine the underlying drivers of the “pulse of the city”
- Understand the effects of perturbations

City Life

- Monitor energy consumption by proxy using light patterns as a measure of building occupancy
- Evaluate the effects of disturbances (e.g., light/noise pollution) on public health
Privacy Protections

- Institutional Review Board approval of all projects involving non-open data
- Close oversight by CUSP Chief Data Officer, Lynn Goldstein
- Limited # of pixels per window (but atmosphere/instrument effects typically dominate)
- Aggregate and de-identified analysis only
Quantified Community
PERSISTENT and SYNOPTIC ANALYTICS for URBAN SCIENCE
Quantified Community Overview

• A fully instrumented urban neighborhood that uses an integrated, expandable sensor network and citizen engagement to support the measurement, integration, and analysis of neighborhood conditions.

• Through an informatics overlay, data on physical and environmental conditions and use patterns will be processed in real-time to maximize operational efficiencies, improve quality of life for residents and visitors, and drive evidence-based planning.
What will the CUSP Quantified Community measure?

**Infrastructure**
Solid waste, storm-water management, power generation/distribution

**Buildings**
Resource consumption; indoor air quality; productivity, health measures

**Safety and Security**
Network Security, Situational Awareness, Emergency Management Integration, Event Forecasting

**Environment**
carbon emissions; air pollution and particulates; noise; climate

**People**
Behavior; mobility; health; activity; social networks, metagenomics
CUSP-Initiated Projects

- Building Informatics
- Sound
- City Lights
- Transportation Research
- Etc.
Objectives & Goals

• Understand and model patterns and flows of energy use and carbon emissions within and across cities, across spatial and temporal scales
• Study the relationship between the urban built environment, urban development, and resource efficiency, particularly focusing on issues of socioeconomic and demographic disparities
• Provide policymakers, community organizations, and the real estate community with actionable insight to support strategies to reduce resource consumption and carbon emission in cities

Data

Currently hold one of the largest databases of heterogeneous, non-self-selected building energy and water and building attribute data available, sources include:

– NYC Local Law 84 – approx 20,000 buildings over three years
– Urban Land Institute – approx. 6,000 buildings across 46 cities over five years
Building Energy Efficiency (NYC Local Law 84)

Source Energy Use Intensity, Multi-Family Buildings, New York City
Source: Local Law 84 Energy Disclosure Data, Kontokosta 2013

N = 7,505
Mean = 137.9
s.d. = 46.8

Source Energy Use Intensity, Office Buildings, New York City
Source: Local Law 84 Disclosure Data, Kontokosta 2013

N = 1,150
Mean = 219.5
s.d. = 101.7

Source: Kontokosta 2013
Sound Project

Overview
- Noise the biggest 311 complaint
- Significant public health issue
- Noise code enforcement is difficult

Objective and Goals
- Characterize the “soundscape” of the City
- Inform policy makers
- Build an archive of sound around the city that can be used for research

Project Plan: initial phases underway
- Audio measurements throughout the City via phones or fixed microphones
- Panel on ethical, legal, and social implications will be convened and led by CUSP Chief Data Officer
The book will identify ways in which vast new sets of data on human beings can be collected, integrated, and analyzed to improve urban systems and quality of life while protecting confidentiality. Sponsored by CUSP, the American Statistical Association, its Privacy and Confidentiality subcommittee, and the Research Data Centre of the German Federal Employment Agency.

Editors: Julia Lane, American Institutes for Research; Victoria Stodden, Columbia; Stefan Bender, The German Federal Employment Agency; Helen Nissenbaum, NYU

Chapter Authors
Steve Koonin, CUSP; Frauke Kreuter, U-MD and Richard Peng, Johns Hopkins; Alessandro Acquisti, Carnegie Mellon University; Robert Goerge, UChicago; Helen Nissenbaum, NYU; Kathy Strandberg, NYU; Paul Ohm, Colorado; Victoria Stodden, Columbia; Alan Karr, National Institute of Statistical Sciences and Jerry Reiter, Duke University; John Wilbanks, Sage Bionetworks/Kauffman Foundation; Cynthia Dwork, Microsoft; Alexander Pentland, et al., MIT; Carl Landwehr, George Washington University; Peter Elias, University of Warwick
Education
Urban Science
& Informatics
CUSP Educational Programs

Train professionals who will understand how cities function and the potential of urban informatics.

CURRENT
• Master of Science in Applied Urban Science & Informatics
• Advanced Certificate in Applied Urban Science & Informatics
• Executive education, “Boot Camps” for City and other professionals

FUTURE
• PhD in Urban Informatics
• Global Executive M.S. in Urban Analytics and Innovation
• Dual/joint degrees (NYU & Academic Partners)
• Distance and online learning modules
M.S. in Applied Urban Science and Informatics

DEGREE
Master of Science

LENGTH
One Year, 3-semester (Full-time)

CLASS SIZE
23 / 1

TERM
August - July
Admissions Summary, Class of 2014
Cycle Dates: December 18, 2012 through June 30, 2013 (~6months)

<table>
<thead>
<tr>
<th>Inaugural Class (including 1 Adv. Cert.)</th>
<th>Selectivity</th>
<th>Years Average Age</th>
<th>Female</th>
<th>Average Undergraduate GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>21%</td>
<td>27</td>
<td>36%</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Undergraduate Disciplines: 20
International: 48%
Countries Represented: 9
Years Average Work Experience: 4
With Graduate Degree: 28%
In 5-10 years, CUSP will be a major center for research and education in Urban Informatics

- Defined practice of Urban Science and Informatics
- Projects that impact the City and its Citizens
- New tools for the social sciences
- Global Reach
- 50 senior researchers, 30 Postdocs, ~500 students (M.S., PhD, Certificate)
- Several hundred new experts trained
- Commercialization and new services development
THANK YOU

lynn.goldstein@nyu.edu
cusp.nyu.edu

Privacy, Big Data, and the Public Good: Frameworks For Engagement

Join us on June 16, 2014 in New York City for a preview event

http://www.dataprivacybook.org/
http://cusp.nyu.edu/event-registration/?ee=44