URBAN CENTER FOR COMPUTATION AND DATA

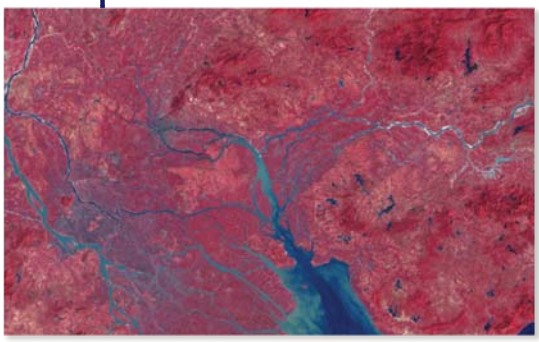
Charlie Catlett, Director
Senior Fellow, Computation Institute of the University of Chicago and Argonne National Laboratory
Senior Computer Scientist, Argonne National Laboratory
catlett@anl.gov

Presentation to the National Academy of Sciences
Committee on National Statistics
February 6, 2014

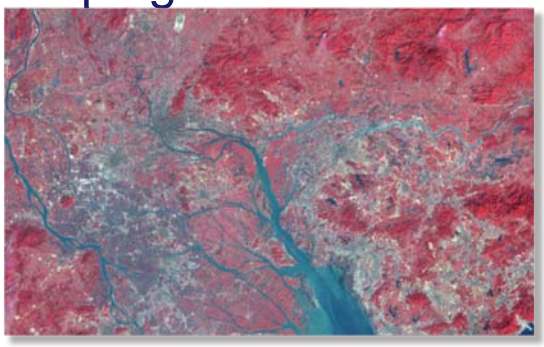
Computation Institute · THE UNIVERSITY OF CHICAGO · Argonne NATIONAL LABORATORY

NSF · MacArthur Foundation

Rapid Urbanization in Developing Economies



Landsat images of the Pearl River Delta in 1980 and 2005, illustrating the impact of urbanization on the planet.



Between now and 2020, the Guangdong province will invest $229B in 202 ongoing and 258 new transport infrastructure projects to create a single 50M person city.

In 2025:
70%
of Chinese people will live in cities with 1M or more people.

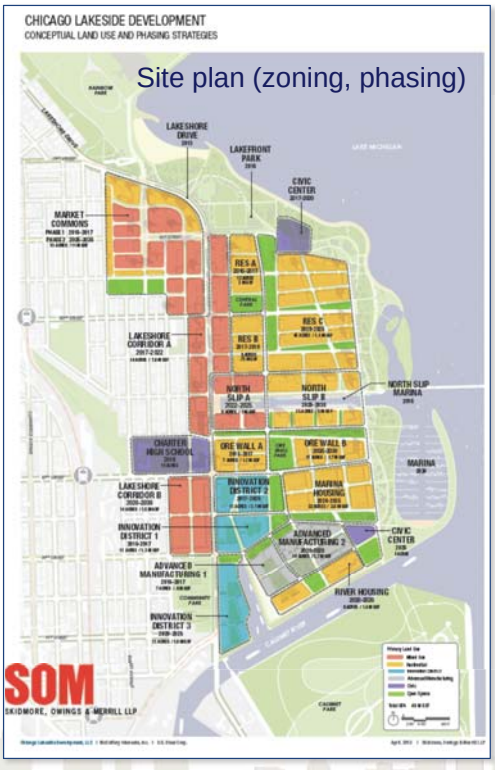
And by 2030...
221
Chinese cities will have 1M or more people.

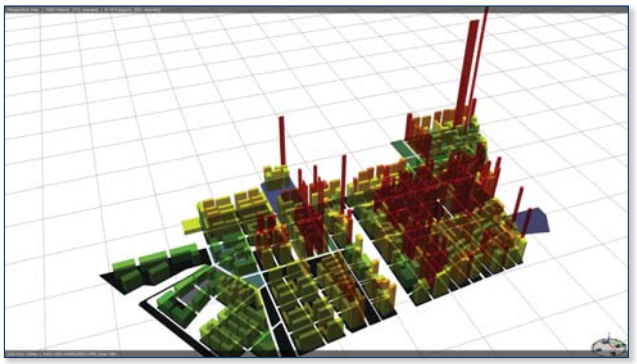
China will *add*
400
million city dwellers

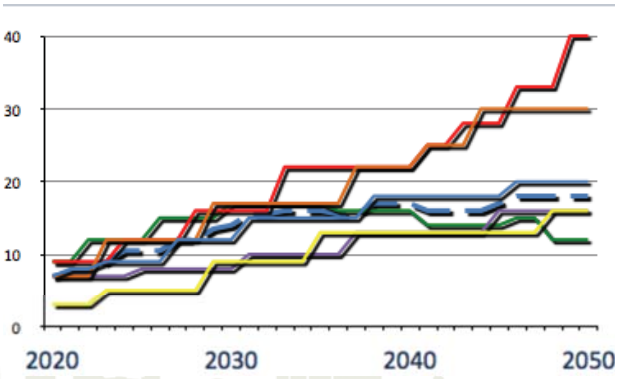
....requiring the construction of one New York City every year for several decades

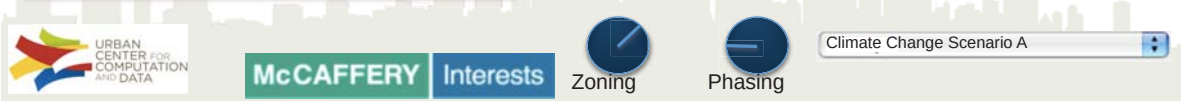
Source: Foreign Policy Magazine, Sep/Oct 2010, "Megacities," Richard Dobbs (McKinsey Global Institute)

	Environment	Infrastructure	Society
Computational Modeling	Science-based design and planning. *Timescales: years to decades.*		
Data Analytics	Evidence-based measurement and predictive analytics. *Timescales: hours to months.*		
Embedded Systems	Decision-support and new ways to interact with the built environment. *Timescales: seconds to hours.*		

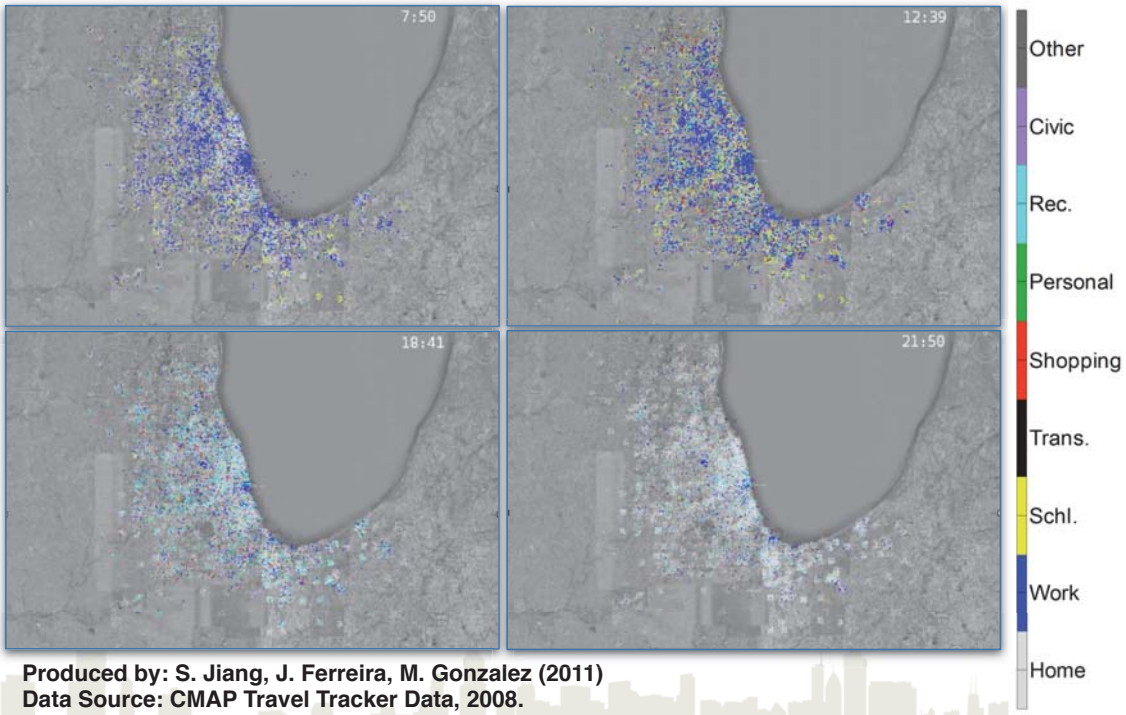




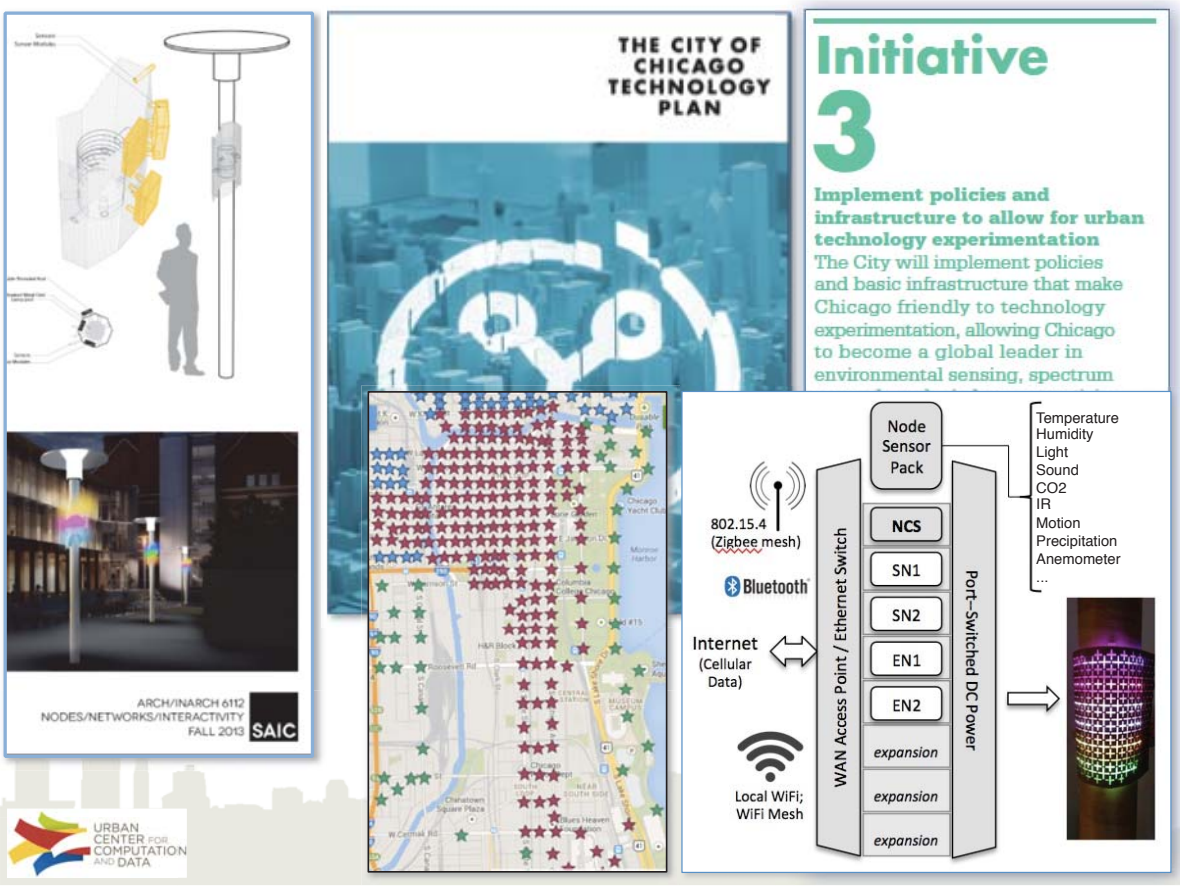




Chicago Human Activity Patterns: Weekday



Produced by: S. Jiang, J. Ferreira, M. Gonzalez (2011)
Data Source: CMAP Travel Tracker Data, 2008.
Reference: Jiang, S., J. Ferreira, and M. González. 2012.
Clustering Daily Patterns of Human Activities in the City. *Data Mining and Knowledge Discovery.* Volume 25, Number 3, Pages 478-510



Healthcare
Education
Social/Behavioral Sciences
Atmospheric Science
Environmental & Life Sciences
Infrastructure Engineering/
Optimization
Computer Science

Instrument

Citizen engagement and communication

Usage, management, access policies/processes

Instrument training materials/activities

Programming and Development Tools

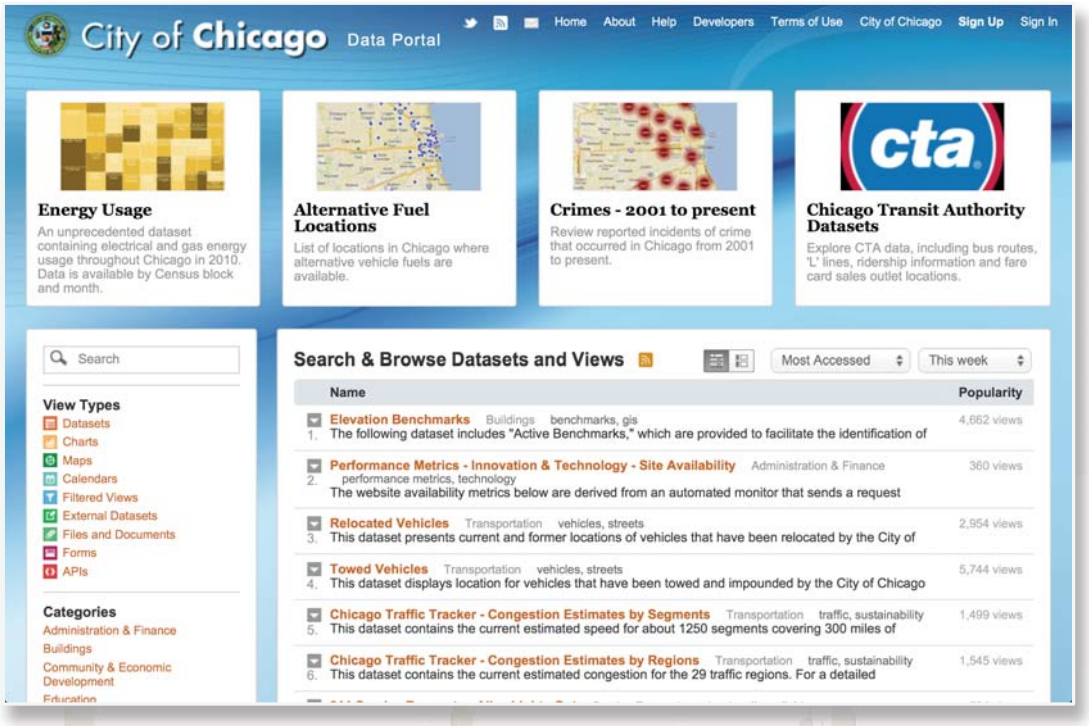
Platform

Device provisioning, diagnostics, management

Shared devices (configurations, capabilities)

Device install and management specifications

Enclosures, Placement, Density, Power, Internet...



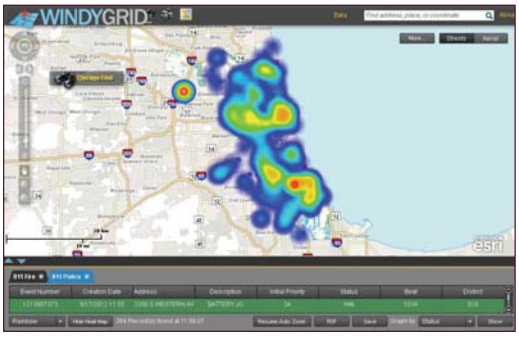
Data Access, Authorization, Privacy

External organizations, citizens

Aldermen, City Employees

Scientists and Engineers (City Partners)

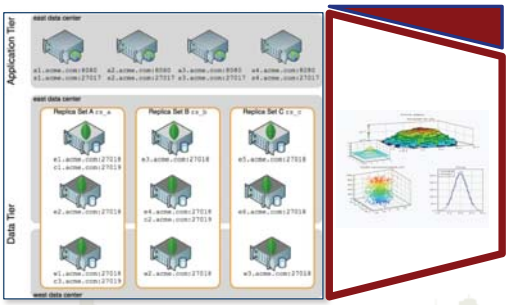
City Decision-Makers



Visual Interaction, Mapping, Analysis Tools

Scalable Data Handling

(pre-processing, cleaning, merging, de-duplication, access…)



Application Programming Interfaces

Automated Continuous Data Analytics

Data Sources

City Databases · Sensors · Video · GPS · Social Media

NSF · Bloomberg Philanthropies · City of Chicago · Urban Center for Computation and Data · Carnegie Mellon University Heinz College

Background: Crime Prediction in Chicago

Since 2009, we have been working with the Chicago Police Department (CPD) to predict and prevent emerging clusters of violent crime.

Our new crime prediction methods have been incorporated into our **CrimeScan** software, run twice a day by CPD and used operationally for deployment of patrols.

Carnegie Mellon University Heinz College

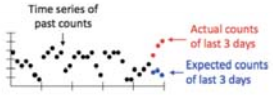
City of Chicago

CrimeScan: Cluster Detection

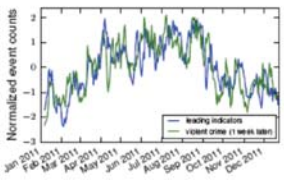
We aggregate daily counts for each leading indicator at the block level, and search for **clusters** of nearby blocks with recent counts that are significantly higher than expected.

Imagine moving a circular window around the city, allowing the center, radius, and temporal duration to vary.

Is there any spatial window and duration T such that counts have been significantly higher than expected for the last T days?



Results: Exploratory Analysis



Considering all subsets of census tracts within each of the 77 neighborhoods of Chicago, 28 different potential predictors, and a 1-week lag, we found a correlation of **r = .786** between violent crime and a subset of 12 leading indicators, for 10 census tracts in the West Englewood neighborhood.

Total run time for all 77 neighborhoods was **2.1 hours**.

Urban Center for Computation and Data

Daniel B. Neill (Carnegie Melon University - Event and Pattern Detection Laboratory),
Brett Goldstein (fmr Chicago CIO/CDO; Fellow, Harris School of Public Policy, Computation Institute, UChicago)

Garbage Cart Black	Inspection	No Building Permit &	Restaurant Complaint	Animal Abandoned
Maintenance	Vacant/Abandoned	Construction Violations	Bulk Pickup	Water Quality
Sanitation Code	Building	Missing Lid/Grate	Unwanted Animal	Sewer Odor/Bad Odor
Violation	Sewer Cave In	Block Party Request	Blue Recycling Cart	Dead Bird
Building Violation	Inspection	Recycling Pick Up	Animal Bite	Animal In Trap
Stray Animal	Garbage Pickup	Nuisance Animals	Dumpster Task Force	Animal Business
Dead Animal Pick-Up	Debris Removal	Animal - Inhumane	Inspection	Animal Fighting
Sewer Cleaning	Fly Dumping	Treatment	Demolition Inspection	

Rodent Infestation

Liquor license (decreased likelihood)
Low-risk v. high-risk gas station (both less likely than non-gas station)
Tobacco license
Neighborhood
Common names (e.g., Tom's market) in store name

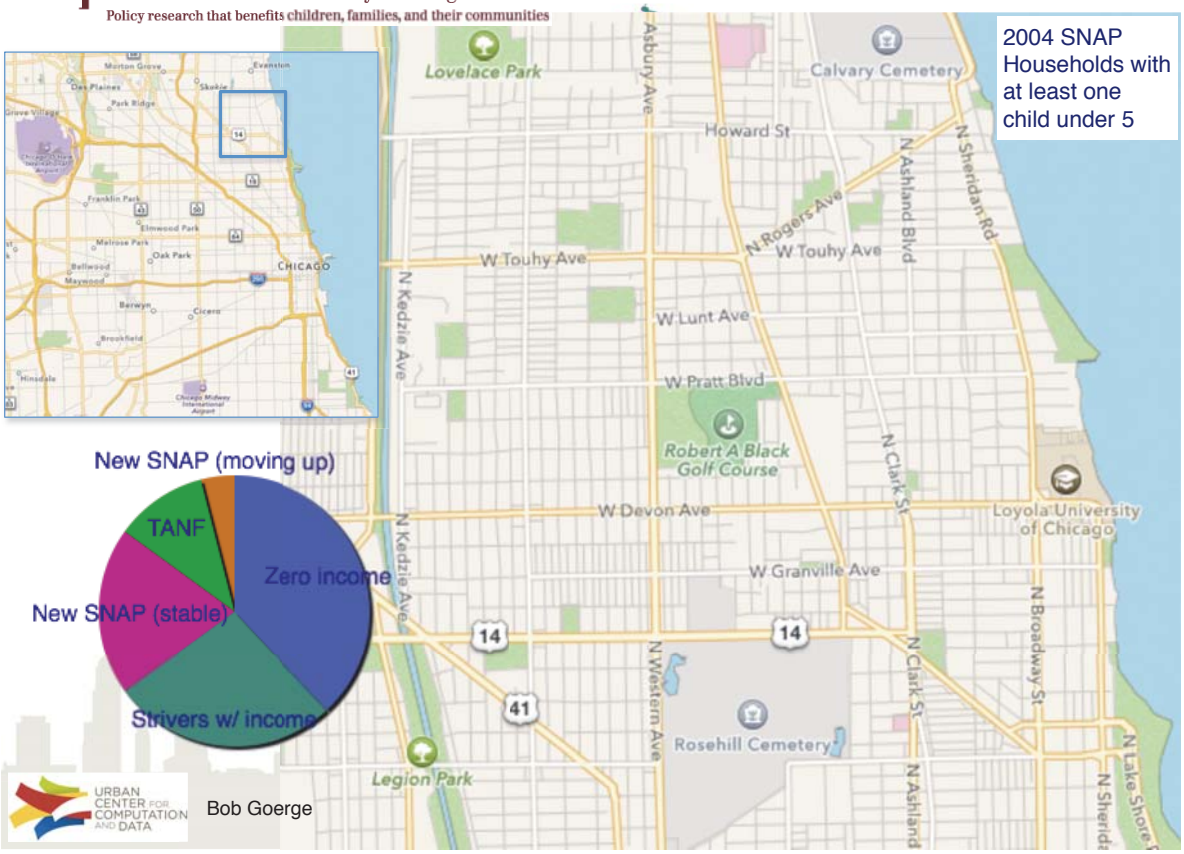
Black Market Cigarettes

Food Safety Inspections

Critical violation found in prior inspections
Sanitation code complaints through 311
Rodent sighting reported through 311
Request for garbage carts (per-ward)
Three-day moving avg of high temp before the inspection
Type of facility (e.g., restaurant versus grocery store)

Tom Schenk, City of Chicago

ChapinHall at the University of Chicago
Policy research that benefits children, families, and their communities



URBAN CENTER FOR COMPUTATION AND DATA

Bob Goerge

CHICAGO INFRASTRUCTURE TRUST

Negative Impacts

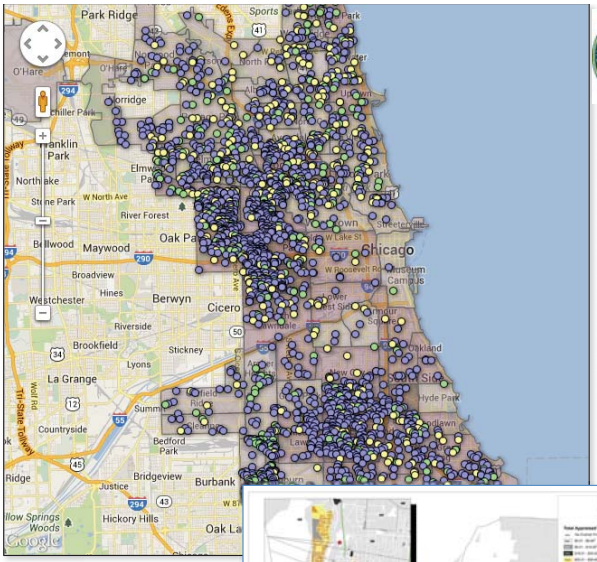
- Lowers property value
- Crime
- Lost taxes

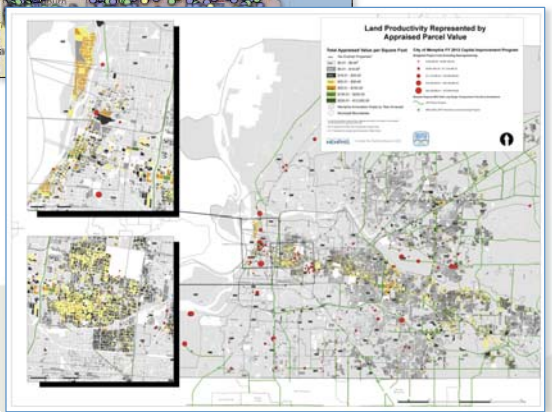
Legal Obstacles

- Unknown owner
- Unpaid bills
- Contaminated land

Guiding Indicators

- Housing stability
- Affordability
- Vacancy





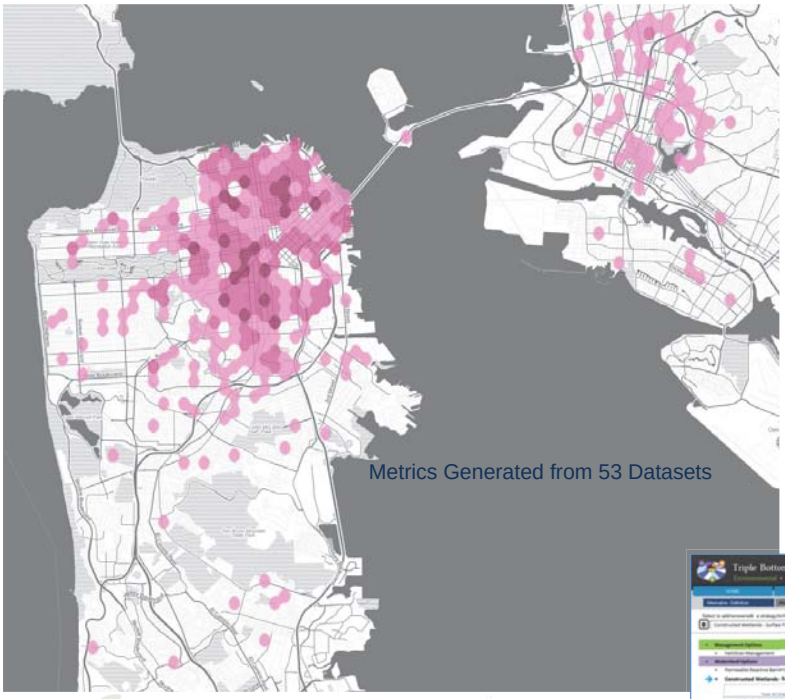
City of Memphis

URBAN CENTER FOR COMPUTATION AND DATA

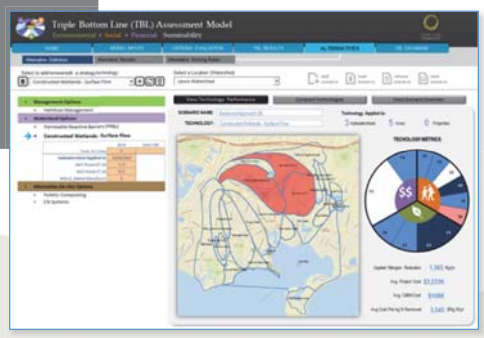
Matt Gee, UChicago

Equitable Development
Community Development
Habitat and Ecosystems Function
Water, Wastewater, and Stormwater
Energy
Materials Management and Waste
Health and Well-Being
Access and Mobility

Matt Gee, UChicago
Kate McGee, City of San Francisco

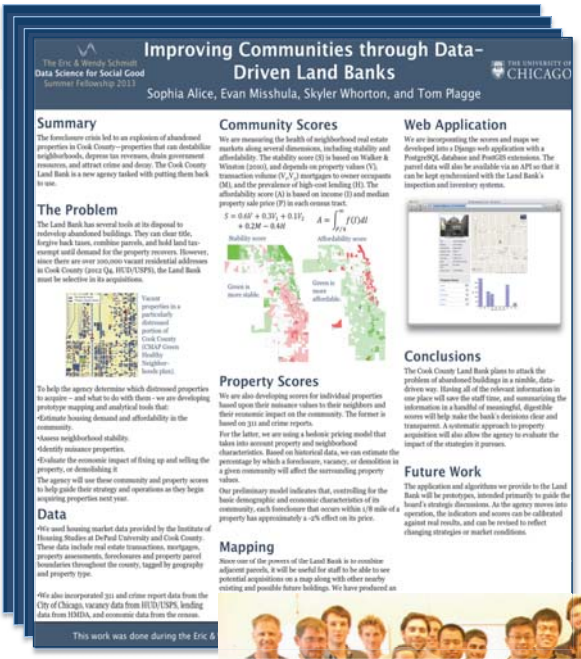


- Built Environment
- Neighborhood Assets
- Housing and Rental Prices
- Building-level Energy Use
- Solar installations
- Renewables generation
- Longitudinal Surveys
- Employment records
- Waste tonnage by block
- Transportation
- Traffic density
- Air quality
- Emissions



Matt Gee, UChicago
Kate McGee, City of San Francisco

SCHMIDT FAMILY FOUNDATION


Improving Communities through Data-Driven Land Banks

Sophia Alice, Evan Misshula, Skyler Whorton, and Tom Plagge

Summary

The Problem

Data

Community Scores

Property Scores

Mapping

Web Application

Conclusions

Future Work


The New York Times Technology | Personal Tech | Business Day

Bits

JULY 25, 2013, 7:47 AM | Comment

A Summer of Data Hacking Social Problems

By STEVE LOHR

The idea, Rayid Ghani recalled, grew out of his experience speaking to computer science students at elite schools like Carnegie Mellon, Stanford and the University of Chicago. President Obama had just won his re-election bid last fall. And Mr. Ghani, chief scientist for the campaign, was on a kind of explanatory victory tour, describing how cutting-edge data analysis and computing tools gave its side an edge.

Robert Kozloff

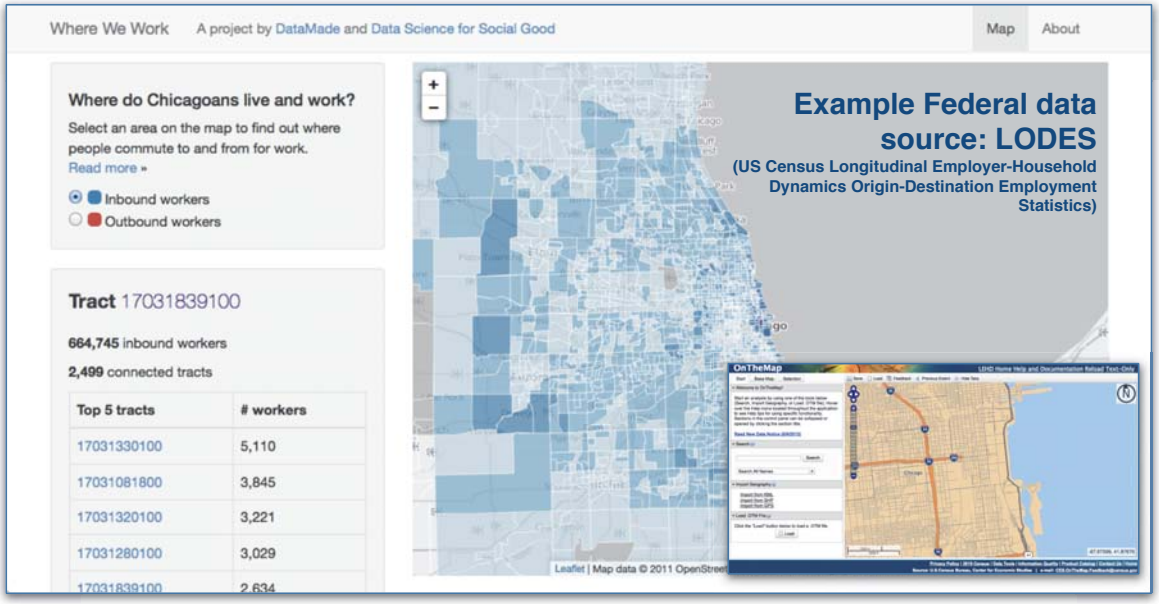
Rayid Ghani, chief scientist for President Obama's re-election campaign.

For Mr. Ghani, the Obama campaign demonstrated how those tools could be used to influence people in fields beyond the well-known commercial ones, like search, social networks and online advertising. And beyond politics, he would tell the students, were a host of social challenges in health care, education and urban development where their skills could be put to good use, working with nonprofits, civic

THE HARRIS SCHOOL OF PUBLIC POLICY | THE UNIVERSITY OF CHICAGO

Computation Institute

URBAN CENTER FOR COMPUTATION AND DATA



No API - must download individual (huge) flat files or use online tool ("OnTheMap").

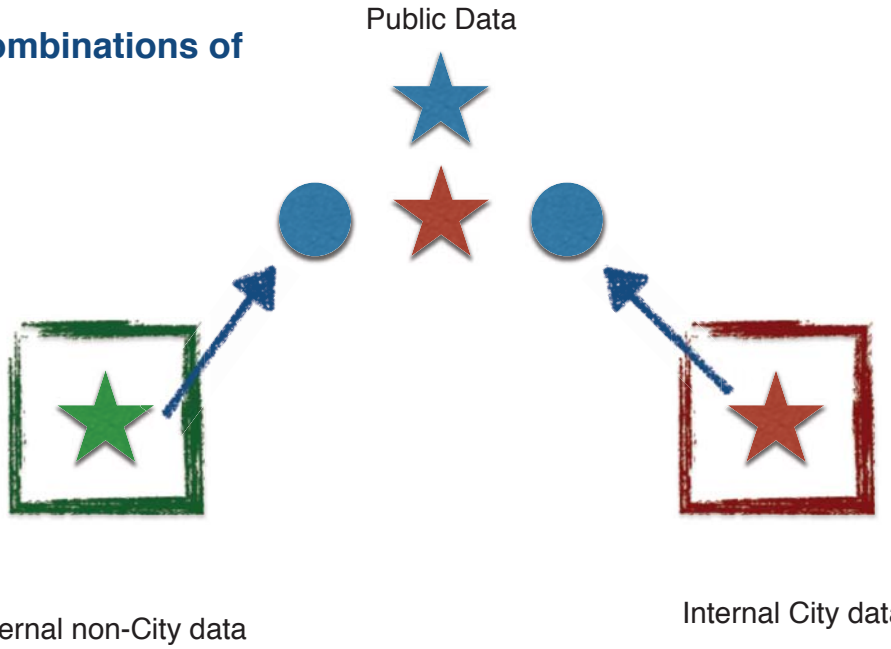
Significant effort to enable geographic queries/navigation

Data is 1-2 yrs old. Ideally could use local data to fill gaps.

URBAN CENTER FOR COMPUTATION AND DATA

Matt Gee, UChicago
Derek Eder, DataMade

New combinations of data...



Summary

Cities are creating (open) data resources

Many critical city decisions also rely on county, state, federal data

Spatial and temporal resolution are key (along with timely data)

How Can the Federal Government Help?

ACCESS	**Modern Data Infrastructure *with APIs***
PREPARATION	**Open Source Tools for Common Tasks**
INCENTIVES	**Strategically chosen standards and requirements**
ACCELERATION	**Fund city/academic exemplars to provide replicable/sharable capabilities and to assist other (esp. smaller) cities.**