Open Data in the Chicago Region

Leaders, geeks and grassroots
I. Explosion
II. Impact
III. How we got here
IV. Scalability
V. What's next?
I. Explosion
Chicagoland government data
2010
Chicagoland government data
2013

Cool painting by Emma Hill
Government Data Sharing in the Information Age

Spring 2007 Events
Sponsored by IDEA (Illinois Data Exchange Affiliates)

What is IDEA?
IDEA is a voluntary group of government agencies and not for profit organizations working to improve data sharing in Illinois. We invite you to join us for presentations and open discussion of government data sharing. Members include:
- Center for Neighborhood Technology (CNT)
- Chicago Metropolis 2020
- Chicago Metropolitan Agency for Planning (CMAP)
- City of Chicago
- Cook County Assessor’s Office
- Metro Chicago Information Center (MCIC)

For more information:
Visit www.chidataexchange.net or contact Greg Sanders at
What changed? Leadership!

Counter-clockwise: President Barack Obama, Governor Pat Quinn, Cook County Board President Toni Preckwinkle, Chicago Mayor Rahm Emanuel
120 days to transparency: City Hall opens

https://data.cityofchicago.org/
120 days to transparency: Cook County

https://datacatalog.cookcountyil.gov/
Federated City-County-State portal

https://www.metrochicagodata.org/
Socrata's cloud-based hosting for open data portals offered a rapid deployment strategy for governments at all levels.
II. Impact
Why open data for metro regions?

All public policy issues cross local boundaries

- Housing
- Employment
- Crime
- Education / workforce development
- Food access
- Transportation
- Natural resources
- Land use
- Tax policy
Why open data for metro regions?

Government works best when information is generally available, as a basis for decision-making and creative problem-solving.

Society works best when government processes are transparent and well-understood.

Web and mobile technologies give us unprecedented opportunities for developing rich information flows across complex urban areas.
Apps for Metro Chicago and other app contests

FROM THE FIELD

Apps for Metro Chicago Winners Announced

Published December 19, 2011

Winners of the Apps 4 Metro Chicago Grand Challenge and the Innovate Illinois Award were announced.

Read a Chicago Tribune article about the winners.
Meetups

April 11 · 6:00 PM
Thoughtful Critiques of the Open Government Movement

93 Members | ★★★★★ | 5 Photos

The international open government/open data movement has a lot of momentum, especially here in Chicago. We’ve all worked hard to get us where we are, and we’re excited... LEARN MORE

February 26 · 6:00 PM
Recent Developments in News, Data, and Gov

84 Members | ★★★★★ | 10 Photos

This month’s meeting is devoted to the work of journalists as it relates to government data. Here’s an agenda: Northwestern University Knight Lab’s Miranda Mulligan (... LEARN MORE

January 24 · 6:00 PM
Using Illinois Employment Data in Civic Innovation

65 Members | ★★★★★ | 5 Photos
Civic Hackers and Code for America

EDIFICE by Cory & Juan at Open City
Maps of Chicago’s built environment, from the Edifice database

Buildings
Demolitions
Violations
Construction

Every building in Chicago, colored by its footprint area.

All data from Chicago’s Open Data portal and OpenStreetMap. Maps designed with TileMill, site template by MapBox. Code is on GitHub.
Tech entrepreneurship and Lean Startup

1871: provides Chicago startups with affordable workspace, mentors, community

New startups focused on public data

Lean Startup Circle: based on Eric Ries’s book, the Lean Startup Circle focuses on tech & data
Intergovernmental data sharing

Cook County and the City of Chicago increase collaboration -- and information sharing

June 2011
Connecting the mega-region

MetroChicagoData.org combines data from Chicago, Cook County and State of Illinois into a single portal

Illinois Open Technology Challenge aims to "increase the number of units of government that publish data online"
Open data accelerates research

UrbanCCD created at University of Chicago (with City involvement) to leverage new data sources
Investment

Data-smart cities stimulate investment and attract a high-tech workforce.

Open public data supports better public services.
III. How we got here
Government transparency's long road

1790s: Pennsylvania Senator William Maclay argues for opening Congressional sessions to journalists

1966: Federal FOIA is enacted despite President Johnson's opposition

Phil Windley's Web Service Manifesto

2003: Utah CIO Phil Windley advocates for machine-readable data as the default information format for government agencies

"Governments control vast data resources.... data is held hostage in disconnected, legacy data resources"

"eGovernment requires that we free data from siloed systems and legacy platforms"
8 Principles of Open Govt. Data

2007: Open Government Working Group defines open government data

"The Internet is the public space of the modern world"

Data must be complete, granular, timely, accessible, machine-readable, non-discriminatory, non-proprietary, license-free and reviewable

http://opengovdata.org/home/8principles
Open data portals everywhere: 2013

39 of 50 states now have open data portals

35 U.S. cities & counties have open data portals

41 countries now have open data portals

Many open data portals around the world are based on open-source CKAN platform
Open data in the open information ecosystem, 2013

- Open data vs. open government
- Open data vs. inter-governmental data sharing
- Open data vs. open source
- Accountability (gotcha!) vs. empowerment (government as a platform)
- Privacy, confidentiality and public safety
IV. Scalability
Can open data scale across regions?

The Chicago region comprises 7 counties, 285 municipalities, hundreds of school districts, etc.
Residents are served best when government information is freely accessible to all.

Data sharing between public entities makes it easier to coordinate policies and operations.

http://www.cmap.illinois.gov/2040/improve-access-to-information
Single jurisdictions vs. metro regions

Cities have mayors, counties have executives, states have governors, countries have presidents.

But regions have... no executive authority. Open data happens community-by-community

Scaling open data across a region involves not only technology -- it requires engagement, persuasion, negotiation and perseverance.
Municipalities manage important data

- Land use and zoning
- Building permits
- Code violations
- Business licenses
- Publicly owned properties
- Infrastructure improvements and public amenities
- Employer tax subsidies
- Parks and open space
- Contracts / tax $ expenditures
Strategies for scaling up open data

Allow smaller units of government to host data on portals of larger entities (Cook County)

Create a collaborative cloud portal serving many municipalities (SmartChicago's Illinois Technology Challenge, also GIS consortia)

Create an open platform where both governmental and non-governmental organizations can post data and documents (CKAN)
Open data portals can be federated

Federation creates a "virtual pool" of data resources

Users of one portal can access data resources from linked portals

Federation can cross platform boundaries (Socrata and CKAN)
V. What Next?
Open data movement: challenges

- Discoverability
- Semantic interoperability
- ETL (extract-transform-load)
- Privacy and public safety
- Data quality
- "Gotcha" - data as political weapon
- Maturity path from civic hacker to tech firm
- "Where are the people?"
Open data movement: decisions

- Data registry vs. data file hosting
- Focus on raw data, analytics, visualizations?
- Runtime APIs vs. download APIs
- Socrata vs. CKAN vs. ______
Discoverability matrix

Open data portals support various resources:

- Uploaded data files
- Links to web-based data on other sites
- Metadata
- Visualizations

Each presents discoverability challenges
Discoverability challenges

Open data portals typically index the metadata registry, but not the content of uploaded files.

Metadata is easily indexed, but limited to fields and tags entered by portal managers.

Data files can be indexed if column-level metadata is documented.

Links to web-based resources need to be indexed by HTML crawlers.
Semantic interoperability

What do we call things? How do we count things?

Aggregating data from 285 municipalities is almost impossible without standard definitions.

Does "foreclosure" mean the same thing to all?

Closely related to metadata: what information does each data column contain?
ETL (extract-transform-load)

Enormous effort and expense required to clean, format, aggregate, geocode and link data

ETL happens BEFORE data is transferred to open data portal and AFTER it is downloaded

Great efficiency gains could be realized from open ETL routines

GitHub is your friend!
Privacy and public safety

Open data is dangerous -- really.

Data center security is crucial when sensitive data is stored.

Almost all data resources can be redacted to protect privacy and public safety, without compromising its value.

When in doubt, leave it out.
Data quality varies, and that's okay

Few data sets have perfect quality

Data quality must be balanced against the value that a data set can yield

Open data can/should have an open process for correction and improvement

Disclaimers are your friend
Gotcha!

Open data invites scrutiny --and criticism / ridicule

Chicago overcame this problem by the particular individuals and relationships involved

Media and civic groups must approach open data as partners of government, work for common goals

Acknowledge quality issues and ask for help
Where should we focus our attention?

End users of open data typically want simple tools for solving everyday problems.

Open data advocates and developers typically want raw available in bulk, in machine-readable formats.

Researchers typically want analytic capabilities.

Government as platform, not provider.
Runtime APIs: beyond downloads

Runtime API is data that can be consumed directly by a browser, mobile device or server.

Runtime APIs eliminate the need for many local copies of the same data.

Runtime APIs require a higher skill level to implement.

Apps are dependent on stability of the API host.
Socrata vs. CKAN vs ____________

Socrata's cloud-based solution is ideal for single jurisdictions with longterm funding and a need for rapid launch and continuous support.

CKAN is ideal for multi-jurisdictional collaboratives, or agencies with ample tech expertise. Also ideal for low-resource projects (free open source code base).

Alternatives such as Philly's open data platform will continue to emerge.
Where are the people?

"Dozens of developers looking at each other in conference rooms over pizza is never going to lead to making lives better in Chicago without the active involvement of real residents expressing real needs and advocating for software that makes sense to them."

--Daniel X. O'Neil

SmartChicago Collaborative's Civic User Testing group: paid to test-drive apps
Conclusions

- Open data is an idea whose time has come (Tim O'Reilly was right!)
- Chicago has succeeded because of the specific individuals in leadership positions
- Extending open data from central cities to municipalities across the metro region is a great challenge, but hugely important
- Open data will have much greater impact if end users are brought into the process at every step
Greg Sanders
gregsanders101@gmail.com
Twitter: @gregsanders

Google+ community: "Open Data for Metro Regions"

Slides are online at: http://bit.ly/16PBrMo