New Challenges and Opportunities in Democratizing Data

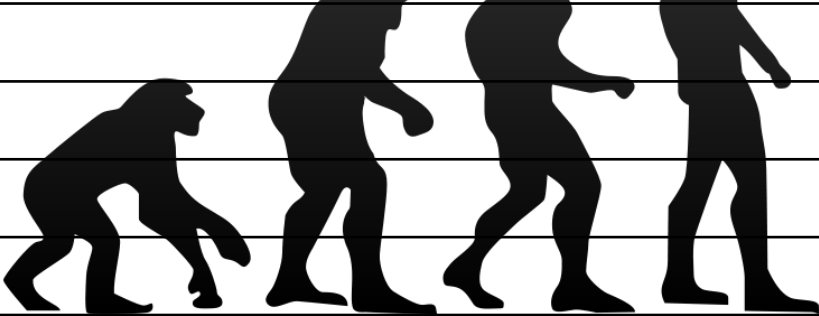## 2012 Regional Data Summit

Jordan Wirfs-Brock

Matthew Barry

"Over the next decade, cities will continue to grow larger at a rapid pace. At the same time, new technologies will unlock massive streams of data about cities and their residents. As these forces collide, they will **turn every city into a unique civic laboratory**—a place where technology is adapted in novel ways to meet local needs."

- Institute for the Future, *The Future of Cities: Information and Inclusion*
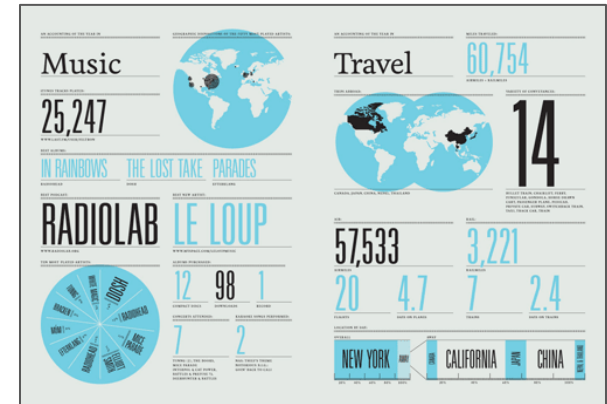
The Information Landscape Has Fundamentally Changed

	Was		**Is**
Broadband	Fixed		Mobile
Smart Devices	Voice		Multimedia
Interfaces	Terminal		Natural Interaction
Computing	Personal		Cloud

Data Creation + Access is Mainstream

True, but….

Democratization of Data (An Expanded Definition)

A Prerequisite.

Complete
Primary
Timely
Accessible
Machine processable
Non-Discriminatory
Non-Proprietary
License-free

Source: Opengovdata.org

Which is (really) more democratic?

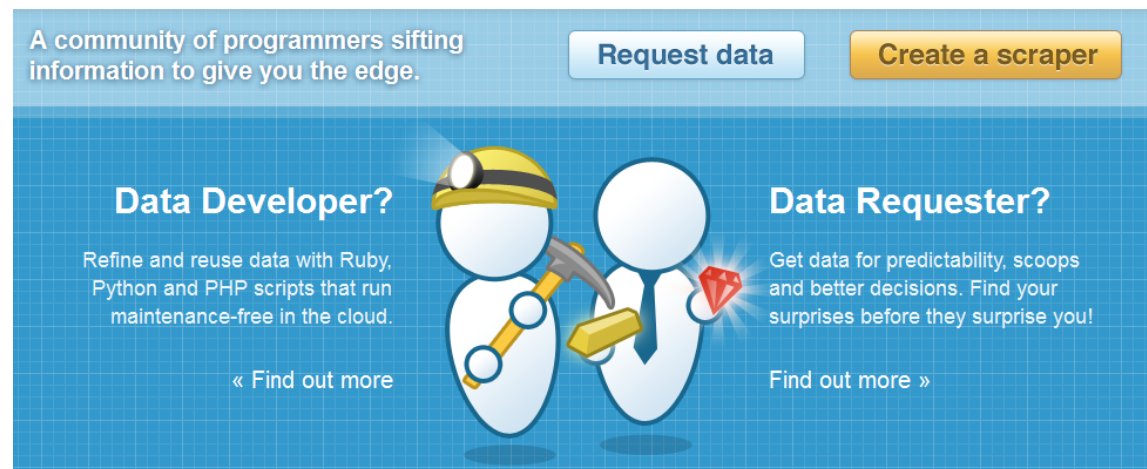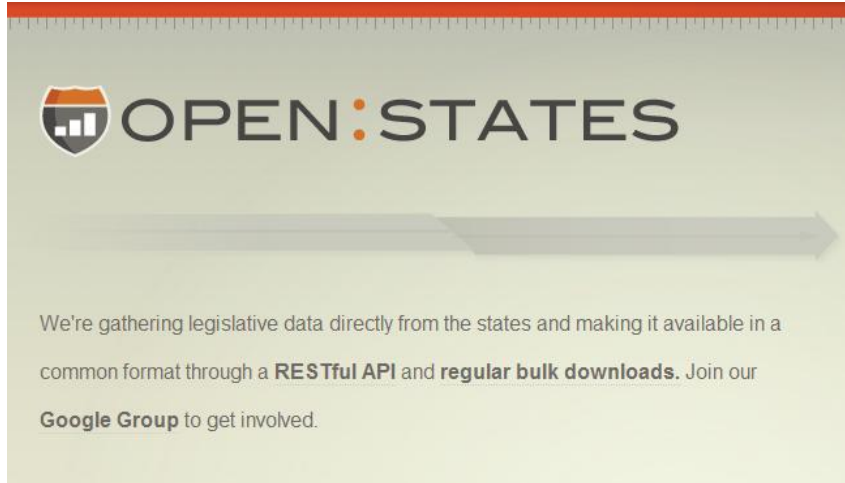| Year	Coordinates
1790	39.27500°N 76.18667°W
1800	39.26833°N 76.94167°W
1810	39.19167°N 77.62000°W
1820	39.09500°N 78.55000°W
1830	38.96500°N 79.28167°W
1840	39.03333°N 80.30000°W
1850	38.98333°N 81.31667°W
1860	39.00667°N 82.81333°W
1870	39.20000°N 83.59500°W
1880	39.06889°N 84.66111°W
1890	39.19889°N 85.54806°W
1900	39.16000°N 85.81500°W
1910	39.17000°N 86.53889°W
1920	39.17250°N 86.72083°W
1930	39.06250°N 87.13500°W
1940	38.94833°N 87.37639°W
1950	38.83917°N 88.15917°W
1950	38.80417°N 88.36889°W
1960	38.59944°N 89.20972°W
1970	38.46306°N 89.70611°W
1980	38.13694°N 90.57389°W
1990	37.87222°N 91.21528°W
2000	37.696987°N 91.809567°W
2010	37.517534°N 92.173096°W
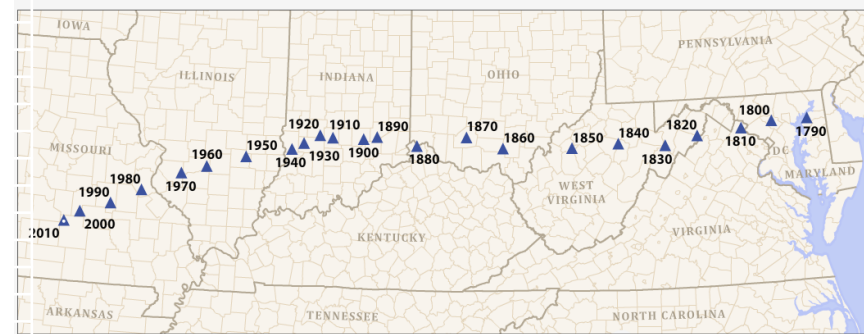
Source: U.S. Census Bureau

No one can do it alone.

"The ability of 5 billion people to instrument the world and share their experiences in a low-cost manner has forever shifted power away from the hands of the few to the network." - Vivek Kundra, *Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect*


Luckily, we don't have to.

"Network strategies" for using data to democratize innovation:

1. Leverage networks and collaborations to liberate data

We're gathering legislative data directly from the states and making it available in a common format through a **RESTful API** and **regular bulk downloads**. Join our **Google Group** to get involved.

A community of programmers sifting information to give you the edge.

Request data   **Create a scraper**

Data Developer?
Refine and reuse data with Ruby, Python and PHP scripts that run maintenance-free in the cloud.

« Find out more

Data Requester?
Get data for predictability, scoops and better decisions. Find your surprises before they surprise you!

Find out more »

"Network strategies" for using data to democratize innovation:

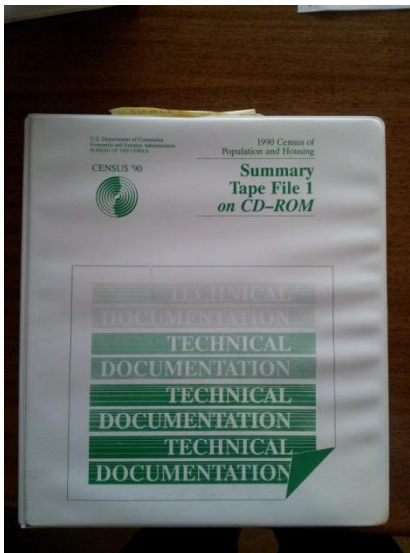## 2. Involve (unexpected) networks and collaborations in applying data
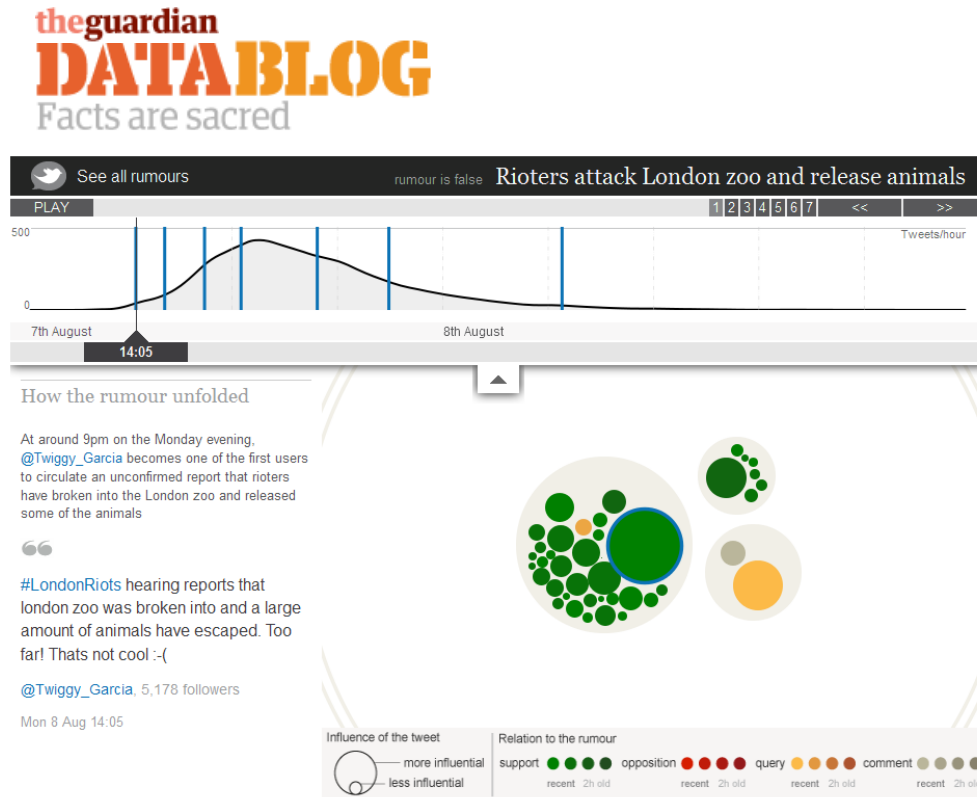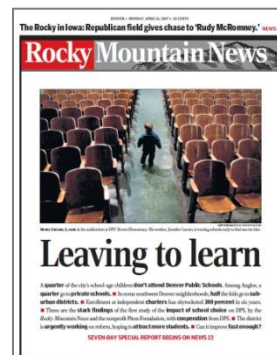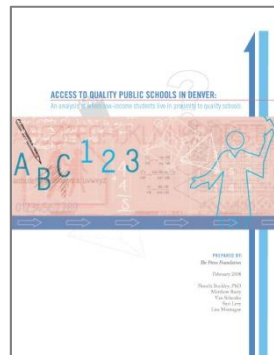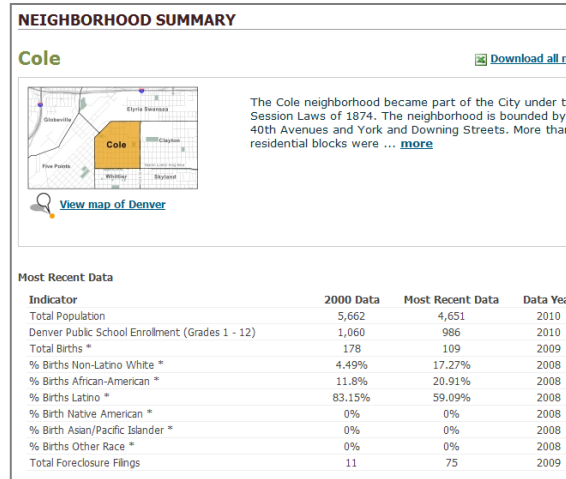
3. Complete the loop

Piton's Data Initiative (A Short History)

1. Data Access

2. Data Interpretation and (limited) Community Information Systems

3. Leveraging Data to Drive Innovation
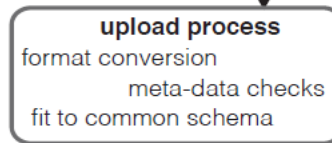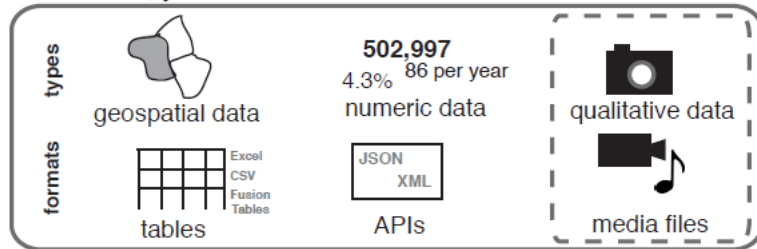
The Colorado Data Commons

and

The Citizen Atlas

The Colorado Data Commons Project

Truly democratic data should be easy to find, easy to use, and easy to share.

THE DATA COMMONS will open key socioeconomic datasets currently locked within public agency databases or hard-to-use formats, providing one-stop access for journalists, policy-makers and citizens.

data from local + federal governments, universities, journalists (vetted) + **crowd-sourced/citizen created data (unvetted)**

types

geospatial data

502,997
4.3%   86 per year
numeric data

qualitative data

formats

Excel
CSV
Fusion
Tables

tables

JSON
XML

APIs

media files

upload process
format conversion
meta-data checks
fit to common schema

relational database

open, queriable API

Colorado Data Commons Front End

A public, user-friendly website featuring:

searches across data sets

lightweight data visualization

data downloads/exports in accessible, easy to work with formats

Other open data projects

external websites
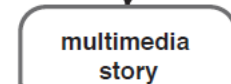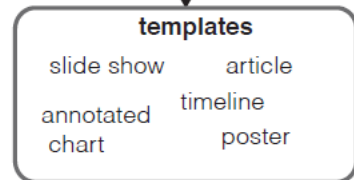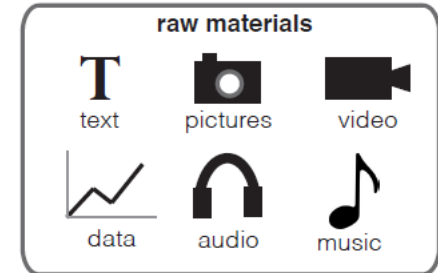
app developers

Citizen Atlas

The Citizen Atlas/StoryBase Project

Data-rich stories can be powerful drivers of change.
STORYBASE will provide user-friendly tools for community members to tell multimedia stories with data.
THE CITIZEN ATLAS will be a freely-available, public-facing website where users can publish, view, browse and comment on stories.

StoryBase

raw materials

text   pictures   video

data   audio   music

templates

slide show   article

annotated chart   timeline   poster

multimedia story

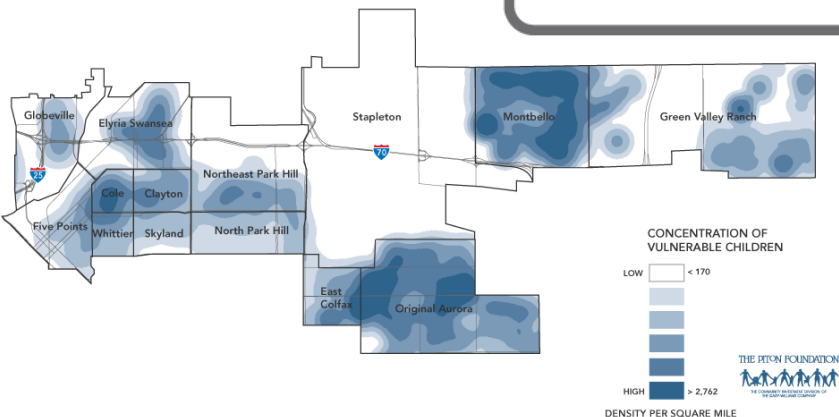website   PDF   Facebook Twitter   etc.

Citizen Atlas

A web publication showcasing:

data- and multimedia-driven stories

about important social issues

from around the metro Denver area

authored by community members and professionals.

Globeville   Elyria Swansea   Stapleton   Montbello   Green Valley Ranch

Northeast Park Hill

Cole   Clayton

Five Points   Whittier   Skyland   North Park Hill

East Colfax   Original Aurora

CONCENTRATION OF
VULNERABLE CHILDREN

LOW   < 170

HIGH   > 2,762

DENSITY PER SQUARE MILE

THE PITON FOUNDATION
THE COMMUNITY INVESTMENT DIVISION OF
THE GARY-WILLIAMS COMPANY

COLLABORATIVE

OPEN SOURCE

Knight Foundation

The Denver Foundation

I·News

opencola

Address problems in real time.

USER NEEDS

COMMUNICATIONS NEEDS

INFO. GATHERING

TECHNOLOGY

CAPACITY

launching
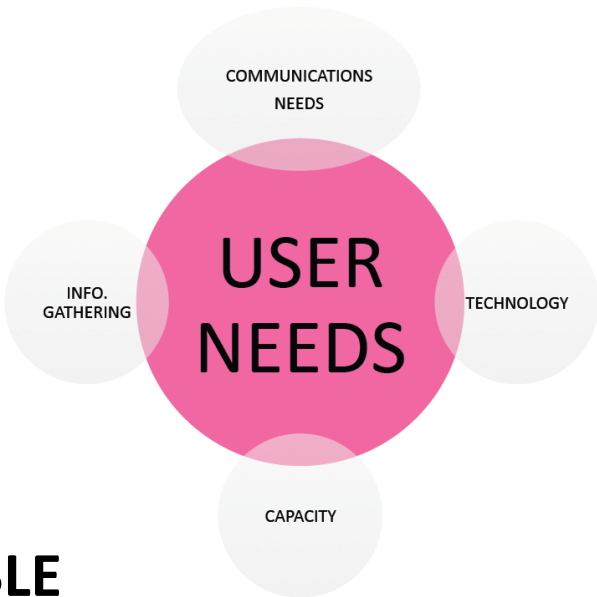
"Something" Up & Running by Year's End

analyzing and adjusting
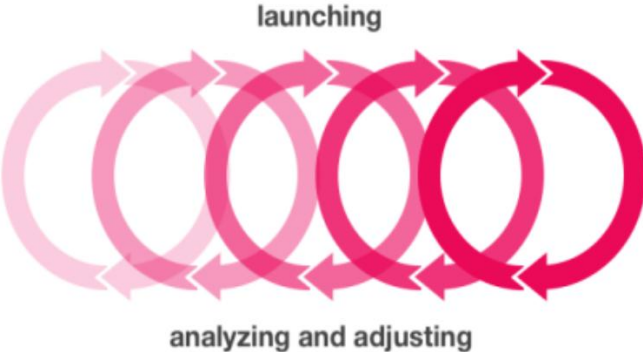
USABLE

ITERATIVE + AGILE

Challenges We Face

Data Control: safeguards v. public good

Participatory Public Services: cooperation v. offloading

More Data, More Models: visible v. actionable

Proliferating Digital Divides: access v. agency

Tyranny of Networks: freedom v. anarchy

*"The biggest threat to the open data movement is the desire for governments to operate in a closed, secretive and opaque manner and to **create a mirage of openness** by releasing low value datasets."* - Vivek Kundra

Discussion Questions

Can democratizing data go too far?

What role should public agencies play in democratizing data?

How could we build an effective network to democratize data?

How can you prove the ROI is worth the energy it takes to democratize your data?