Reuse and Remix of Government and Public Sector Data
IASSIST 2012

Minglu Wang (minglu@rutgers.edu)
Data Services Librarian, Rutgers-Newark

Joshua Horowitz (joshua.horowitz@rutgers.edu)
Project Coordinator, New Jersey DataBank, Rutgers-Newark

Danielle Farrie (Dfarrie@edlawcenter.org)
Research Director, Education Law Center, Newark, New Jersey

Peijia Zha (peijiaz@andromeda.rutgers.edu)
Research Associate, Rutgers-Newark Schools Research Collaborative
Rutgers-Newark: Engaged Research

• New Jersey DataBank is directed by Marc Holzer, Dean and Board of Governors Professor of the School of Public Affairs and Administration

• Education Law Center is founded by Professor Paul Tractenberg of Rutgers Law School in Newark with a start-up grant from the Ford Foundation

• Newark Schools Research Collaborative is a collaborative project of Newark Public Schools and Rutgers-Newark, and currently directed by Alan R. Sadovnik, professor of Education, Sociology and Public Affairs
Grand Mission

New Jersey / Newark, NJ

Government and Public Sector’s Data

Data Bank Database Data Warehouse

Researchers

Advocates

Community

Planners & Officials
New Jersey DataBank

• An ever-growing project
  – An attempt to aggregate “all” New Jersey policy area data into a single place
  – 14 Policy Areas (e.g. Health, Public Safety, and Business)

• Links to other databases
  – Stratified over policy areas

• Data Visualizations
  – Mainly using Google products

• Rankings
  – Mainly state
  – Soon to be county

• Comparison functions to allow cross-county comparisons of municipalities
Education Law Center

– “Extensive database of educational indicators for the state of New Jersey: school funding, student demographics, student performance, the teaching workforce, and community context”

– “The database allows ELC to provide thorough and timely analysis of the educational issues facing the state.”
Newark Schools Research Collaborative

- “To establish a large-scale data warehouse on the public schools and all charter schools in Newark, and to make available these data to researchers examining educational issues affecting Newark”

- Also serves Newark community, including the school district, the charter schools, the Mayor’s office, community based organizations, and parent organizations
Big Challenges – Data Aggregation

• Availability of data
  – NJ Databank: hard to get municipal level data, older data, some policy areas data.
  – ELC: reliant on timeliness of government release of data, consistency of reporting over years, often not as disaggregated as we would like (district, school, grade), much data not available by racial/ethnic groups or income groups
  – NSRC: hard to get clean school level data, and also follow up data after students graduate from high school
Big Challenges – Data Aggregation

• Quality of data
  – Consistency: change of reporting system, data collection methods, policy and personnel
  – Accuracy: data errors; discrepancies among different levels of data (both individual level and summary levels)
  – Missing data – some schools/districts/charters fail to report data in any given year
  – Reporting errors at district level
Big Challenges – Data Aggregation

• Sources of data are in different formats & Structures
  – Data in PDF files is difficult to transfer
  – Excel spreadsheet structure can also change over time
Big Challenges
– Management & System Architecture
The data is all stored in MySQL and the comparisons are generated dynamically from the database.

Joomla is the Content Management system to manage the web design and overall look and feel.
Education Law Center

Data Inputs
(Excel, text, or Access files)

SQL Server

Access Database

Tableau Desktop
Newark Schools Research Collaborative

New Jersey and National Public Access Data (national, state, school district, and school level data)

- New Jersey and National Public Access Data (national, state, school district, and school level data)

- Newark Public Schools (student level data)

NSRC SQL Server (Under Construction) and Desktop (Excel & SAS files)

NSRC inside users (researchers, graduate students)

- NSRC inside users (researchers, graduate students)

- Joomla(CMS)
Big Challenges - Data Presentation

• Visualization
  – Google Applications
  – Tableau

• Engage users
  – providing enough interactivity to make the charts interesting, but not overwhelming
  – Audience has varying levels of familiarity with the data – press, advocates, parents, policymakers – sometimes hard to make report appropriate for all levels
Strategies

• Self-learning
  – SQL database design
  – Google Application
  – Tableau

• Learn from other practitioners
  – Local network
  – Online user community

• Collaboration
  – Data sharing
Accomplishment

Profiles Comparison

Comparisons are an important tool provided by the NJ DataBank. The user can compare municipalities or counties on many policy indicators.

Step 1: Choose a year from the second pull down menu
Step 2: Choose a Municipality from the map that appears

Select a Timeframe:
2010

Select four or fewer municipalities from the map below that you wish to compare.

Selected Municipalities:
1. ESSEX • LIVINGSTON TWP
2. ESSEX • MONTCLAIR TWP
3. ESSEX • NEWARK CITY
4. 

New Jersey DataBank

- Comprehensive New Jersey data source collection on 14 policy areas
- Municipal level profile data comparison
- Featured rankings
Accomplishment

Students in Special Programs in NJ Schools

Select a School Year to Start

STUDENTS IN SPECIAL PROGRAMS BY DISTRICT WEALTH

- Free Lunch Eligible
- Free & Reduced Lunch Eligible
- Limited English Proficiency
- Students with IEPs

District Wealth
- Low Wealth
- Middle Wealth
- High Wealth

STUDENTS IN SPECIAL PROGRAMS IN ATLANTIC COUNTY

- Total Enrollment
- Free Lunch Eligible (%)
- Free & Reduced Lunch Eligible (%)
- Students with IEPs (%)
- Limited English Proficiency (%)

- Atlantic City
- Atlantic Co Special Serv
- Atlantic Co Vocational
- Brigantine
- Buena Regional
- Egg Harbor City
- Egg Harbor Twp
- Estell Manor City
- Folsom Boro
- Galloway Twp
- Greater Egg Harbor Reg
- Hamilton Twp
- Hammonton Twp
- Littletown
- Manahawkin
- Margate City

- District of New Jersey

• extensive historical and current education database focused on New Jersey
• Online interactive report of education data
Accomplishment

Comprehensive Newark education data for newspaper articles, journal articles, and research reports