MAKING DATA OPEN & USEFUL: PROVIDING ENHANCED ACCESS TO PUBLIC DATA THROUGH WEB AND APP TOOLS TO FACILITATE BETTER LOCAL DECISIONMAKING

J.W. Sipple[1], J. Vink[1], J. Francis[1], J. Williams[1,] J. Morgan[2,] S. Brady[3], A. Tait[4]
[1] *Cornell University (UNITED STATES)*
[2] *Spencer-Van Etten Central School District (UNITED STATES)*
[3] Prism Decisions Systems (UNITED STATES)
[4] *Decision Mechanics Limited (UK)*
jws28@cornell.edu, jkv3@cornell.edu, jdf2@cornell.edu, jmorgan@svecsd.org, *andrew.tait@decisionmechanics.com*

Abstract

Problem: School and community leaders are inundated with information and political pressures as they make short- and long-term decisions. Faced with ongoing demographic and enrollment changes and the current fiscal challenge (fiscal crisis in many areas), improved understanding of demographic, performance, and financial trends is central to improved and enhanced long-term planning. School district budget planning is often narrowly focused on the next budget year without adequate attention paid to the longer term (2-10 year) implications of the current decisions. The very notion of longer term budget projections is sometimes foreign, but often just not calculated. Hence, local budgetary decisions may have unknown (or harmful?) effects on the long-term stability of the school district, which will have a substantial impact on the community served by the school.

Current state of data sharing: State and national governments systematically collect and (often with a delay) disseminate demographic, performance, and financial data to the public. While the dissemination occurs, the format of the data typically takes one of two forms. The data may appear in a spreadsheet or database with each year of data found in a new table or form. The burden of linking together data tables to allow for the analysis of trends falls to the end user - most without the technical skills or inclination to functionally use the available data. Second, the school or school district data is often released in single .pdf files, as is the case with U.S. school districts responding to the reporting requirements of the No Child Left Behind Law. This allows for an end user to see one year of data for one school district at a time with no ability to compare across districts or across time. Most municipal data systems have similar limitations.

Innovation: For three years, we have been immersed in a project to use publicly available data, use our collective expertise and connections, and create a set of data tools to allow enhanced access to data such that it informs local decision makers. This has been a four stage process thus far. First, we identified relevant data sets made available by New York State and the U.S. Census Bureau that include demographic, financial, and performance data. Second, we assessed need and interest across the state using our connections and access to school district and community leaders across the state. Third, we built a suite of web-applications that made use of the available data, but in a way that takes advantage of GIS, visual data analysis, long-term trend analyses (18 years of data), and data forecasting to allow end users intuitive and user-friendly formats, visualization, and manipulation. Finally, we have built an iOS application (for the iPad) for budget forecasting and scenario building. The proverbial “best case” and ”worst case” scenarios can be compared and manipulated in real time using intuitive sliders. The budget data can also be viewed in one of three budget details (simple, moderate, and detailed), in dollar or annual percent change formats.

At INTED we will present our work, as part of a larger research project, including the data tools and evidence from local decision makers about their use of the data tools and implications for local decision making and future tool development. Central to this paper is the analysis of the value added by the data tools in local decision making.

Keywords: Data, school districts, demographics, finance, community, web application, iApps

1 CONTEXT

The public educational system in the United States consists of over 15,000 separate school districts located inside the 50 states. Districts range in size from just a few students to 1.1 million students in New York City. New York State (U.S.A.) has just under 700 school districts. Of these, the five largest are fiscally dependent on their city municipality for their local revenues (but also receive considerable state and federal revenues) and include New York City, Buffalo, Rochester, Syracuse, and Yonkers. The remaining school districts have the legal authority to levy their own school tax (typically property tax) and self govern with lay school boards (typically elected) who hire a superintendent to oversee the professional operation of the district.

2 PROBLEM

School and community leaders are inundated with information and political pressures as they make short- and long-term decisions. Faced with ongoing demographic (e.g., growth in minority populations) and enrollment changes (large losses in rural areas) and the current fiscal challenge (fiscal crisis in many areas since 2008), improved understanding of demographic, performance, and financial trends is central to improved and enhanced long-term planning. In the U.S, and in New York in particular (with fiscally independent school districts), school district budget planning is often narrowly focussed on the next budget year without adequate attention paid to the longer term (2-10 year) implications of the current decisions and the decisions made by municipalities. The very notion of longer term budget projections is sometimes foreign, but often just ignored in favor of the immediate pressure on the current negotiations. Hence, local budgetary decisions may have unknown (or harmful?) effects on the long-term stability of the school district, which will have a substantial impact on the community served by the school.

Using a grant from the New York state government, the NYS Center for Rural Schools (CRS) set out with the goal to be "useful" to local school and community leaders. Speaking at 20 different gatherings of superintendents and school board members in 2008 and 2009, we began to see the void that existed in the capacity of local school and community leaders. Two early presentations are illustrative:

The very first presentation of the beta data tools was to a group of superintendents and school board members. At the conclusion of the demonstration of a financial tool, a superintendent raised his hand and asked, "What if the media get a hold of this?" This signalled something very clear to the participants in the room and to the presenter, who responded, "With all due respect to you and your colleagues, I hope they do." This common practice of keeping data inaccessible is so widespread that people seem to forget that it is indeed public data. Every bit of data referenced by our data tools is made available by the state to the public posting of files on various web sites on the state education department domain.

The second telling experience was during the very first superintendent-only presentation Sipple gave. Upon showing a tool that visually displays a school district's financial or demographic (e.g., local revenues, enrollment) data over 17 years, one superintendent yelled to another, "Hey Joe, look what this can do. I've been trying to do make this chart for two years."

These examples are more typical than exceptional, and signal the traditional view of data in and around school districts (and local governments). First, that it is scarce and must be carefully controlled, and second, the capacity of school district leaders to make use of the annual release of data by the state is inadequate. "This represents no criticism of superintendents, as they operate in a highly localized and political arena that makes it safer to rely on no data rather than on data that would be open to critisicm or debate."

In short, our goal is to make data available in a form that is not only accessible to greater numbers of people, but that it is easy, intuitive, and even fun to use. We are not aiming to conduct analyses that would direct decisions by local decision makers. We, instead, are working under the assumption that

local decision makers, when armed with useful data in a useful format, will use the data to inform themselves and their community and in the process make better decisions.

3 CURRENT STATE OF OPEN DATA SHARING

State and national governments systematically collect and (often with a delay) disseminate demographic, performance, and financial data to the public. While the dissemination occurs, the format of the data typically takes one of two forms. First, the data may appear in a spreadsheet or database with each year of data found in a new table or form. The burden of linking together data tables to allow for the analysis of trends falls to the end user - most without the technical skills or inclination to functionally use the available data. Second, the school or school district data is often released in single .pdf files, as is the case with U.S. school districts responding to the reporting requirements of the No Child Left Behind Law. This allows for an end user to see one year of data for one school district at a time with no ability to compare across districts or across time. Most municipal data systems (e.g., county, town) have similar limitations.

Not all governments are releasing data to the public and researchers in .pdf and Excel formats. Some, for example the UK and even the United States federal government are now driving an open data movement.[1] But, while there are some trendsetters, most government agencies and service providers are still woefully behind the curve. This paper is not about the technical aspects of the cutting edge of the open data movement, but rather our purpose is to illustrate the utility making data readily available to local decision makers and how this process could be made quicker and easier through increased availability of open-sourced data.

4 INNOVATION

For three years, we have been immersed in a project to use our collective expertise and connections, use publicly available data, and create a set of data tools to allow enhanced access to data such that it assists local decision makers. This has been a four-stage process thus far. First, we identified relevant data sets made available by New York State and the U.S. Census Bureau that include demographic, financial, and performance data (see http://nyruralschools.org/pages/capacity.php). Second, we assessed need and interest across the state using our connections and access to school district and community leaders across the state. Third, we built a suite of web-apps that made use of agile development principals, available data, but in a way that takes advantage of GIS, visual data analysis, long-term trend analyses (18 years of data), and data forecasting to allow end users intuitive and user-friendly formats, visualization, and manipulation. Finally, we have built an iOS applicaition (iApp; for the iPad) for budget forecasting and scenario building. The proverbial “best case” and ”worst case” scenarios can be compared and manipulated in real time using intuitive sliders. The budget data can also be viewed in one of three budget details (simple, moderate, and detailed), in dollar or annual percentage changed formats.

5 DATA

We draw on data published by the New York State Education Department and the U.S Census Bureau. This includes public data on school and district enrollment, school district finances, and school district performance.

Enrollment: The NYS Basic Educational Data System (BEDS) is updated annually in October with enrollment at each grade level at each school in each district. This is released by the state of NY in December and establishes the official enrollment for each district and for the state.

Finances: Each fall, every school district must submit its financial report to the state for auditing and aggregation. This consists of nearly 550 variables (expenditures and revenues including debt service, bonds, and reimbursable expenses such as transportation and BOCES expenditures. These 550 variables are aggregated into about 35 variables that represent standards categories of expenses and revenues for all districts. This data is released annually as a single Excel spreadsheet and we currently have 18 years of harmonized data.

---

[1] While many definitions of “open data” exist, we rely on a broad definition advanced by http://linkedgov.org/, http://data.gov.uk/data/tag/education, and http://www.data.gov).

Performance: Driven by both state and federal reporting requirements, the School Report Card Database is published annually in a Microsoft Access relational database. We currently have 6 years of harmonized performance data. The data consist of test scores(grades 3-8, and high school end-of-course examinations) - aggregated to the district level - but also broken out by 18 different subgroups required by the Federal No Child Left Behind Law (poor vs. Not-poor, special education vs. Regular education, male vs. Female, and Racial categories).

6 TECHNICAL DETAILS

We have built a database on a Cornell University server that holds the public state and federal data. The base call that lists the resources is: http://pad.human.cornell.edu/api/BudgetPlayground. The individual district data can be extracted through: http://pad.human.cornell.edu/api/BudgetPlayground/Districts/{id}/FiscalPlan. We implement calls to the service using a REST-framework. Once the base schema was set, it is not hard to expand the capabilities of the API. Starting with this database, and making it available to anyone who wished to link to it, we built a series of web apps and then an iApp.

7 TOOLS

Our web tools consist of single use apps that draw on the same data base, though often just small sets of variables to allow a tight focus to the app. We have divided the apps into County level apps and School District level apps. For each respective app, the user can select data from individual county or school district in the state. We encourage users to not only examine their own county or school district data but rather to examine their broader region. Users often hear about other counties or districts that are more or less efficient than theirs, or that very different trends are occurring than in the home location. These tools allow for fact checking and gaining a broader regional perspective on demographic, performance, and performance trends.

New York Counties Data

Maps

An interactive map with selected County statistics

Projections

Population size and characteristics, projected until 2035

Latest county profile

The latest information from the American Community Survey (2006-2010 5-year estimates)

Recent trends

Data from select demographic and economic annual time series

Historic trends

Data from the latest decenial Census, compared with data from past censuses

New York School districts Data

Maps

An interactive map with selected school district statistics

Projections

Enrollment projections

Latest ACS school district profile

The latest information from the American Community Survey (2005-2009 5-year estimates)

Recent trends

Data from select enrollment and financial annual time series

Enrollment details

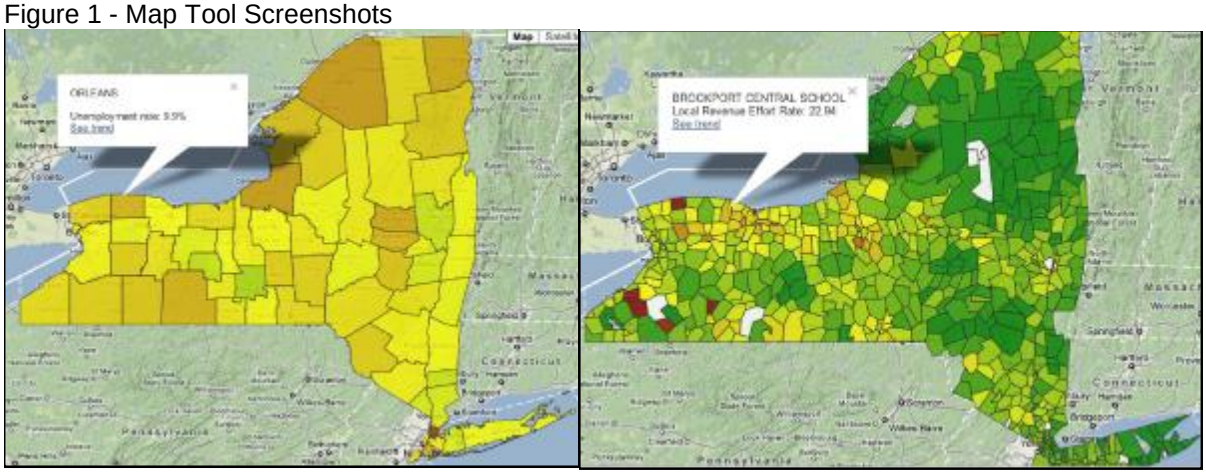
Detailed demographic enrollment data

Budget playground

Study the impact of future budget assumptions

The mapping tool is a way to visualize a user's local County/District data in light of its region and the state. Figure 1 displays a county-level screenshot of unemployment rates and a school district screenshot of local property tax rates. Of note, the callout window displays a highlighted name, data point, and the ability to click on a "See Trend" link to display a line chart with 18 years of data for this particular location and variable.

Figure 1 - Map Tool Screenshots

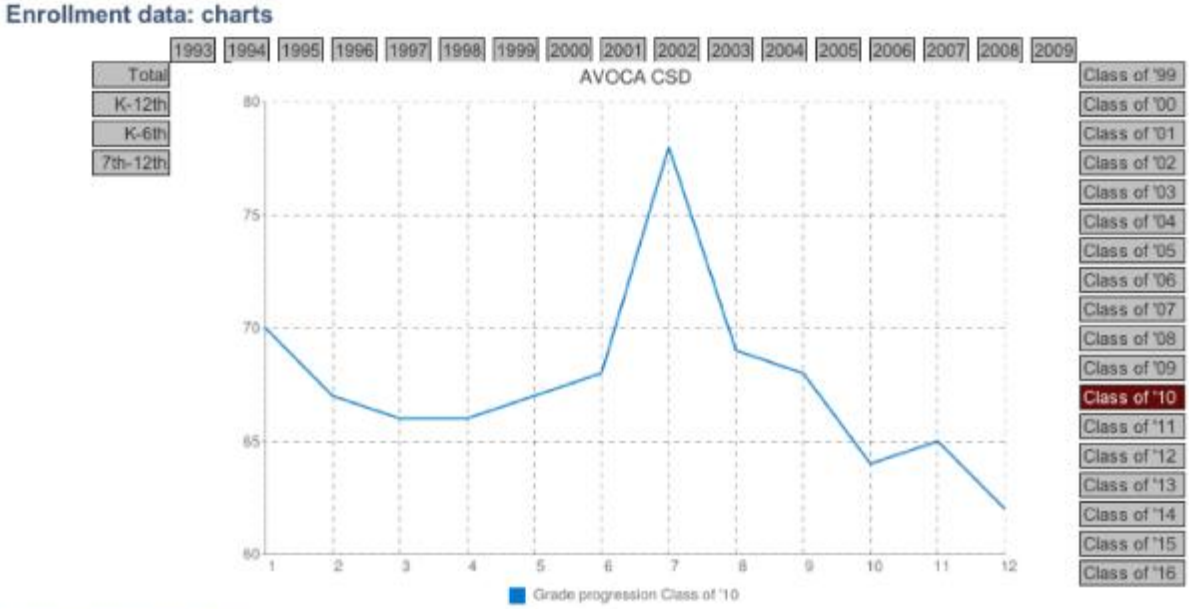


The Projection tool uses past data to project population and enrollments into the future. This can be done in a variety of ways including population pyramids and line charts for particular age groups. Figure 2 displays a population pyramid of a single county projected to 2030 and a line graph of the population decline of school age children. Depending on the end-user's question at hand, they may examine year-to-year changes, trends over time, or examine age-graded sub-groups, including graduation cohorts.

Figure 2 - Population Pyramids of a single county in 2000 vs 2030.

Our Enrollment Detail tool allows for simple inspection of enrollment trends, status and multiple views. The user can look at enrollment at each grade level for a particular year of data, or look at the cohort size over time (see Figure 3). Figure 3 also displays the grade progression details below that can be downloaded by the reader.

Figure 3 - Enrollment Details Tool Screenshot

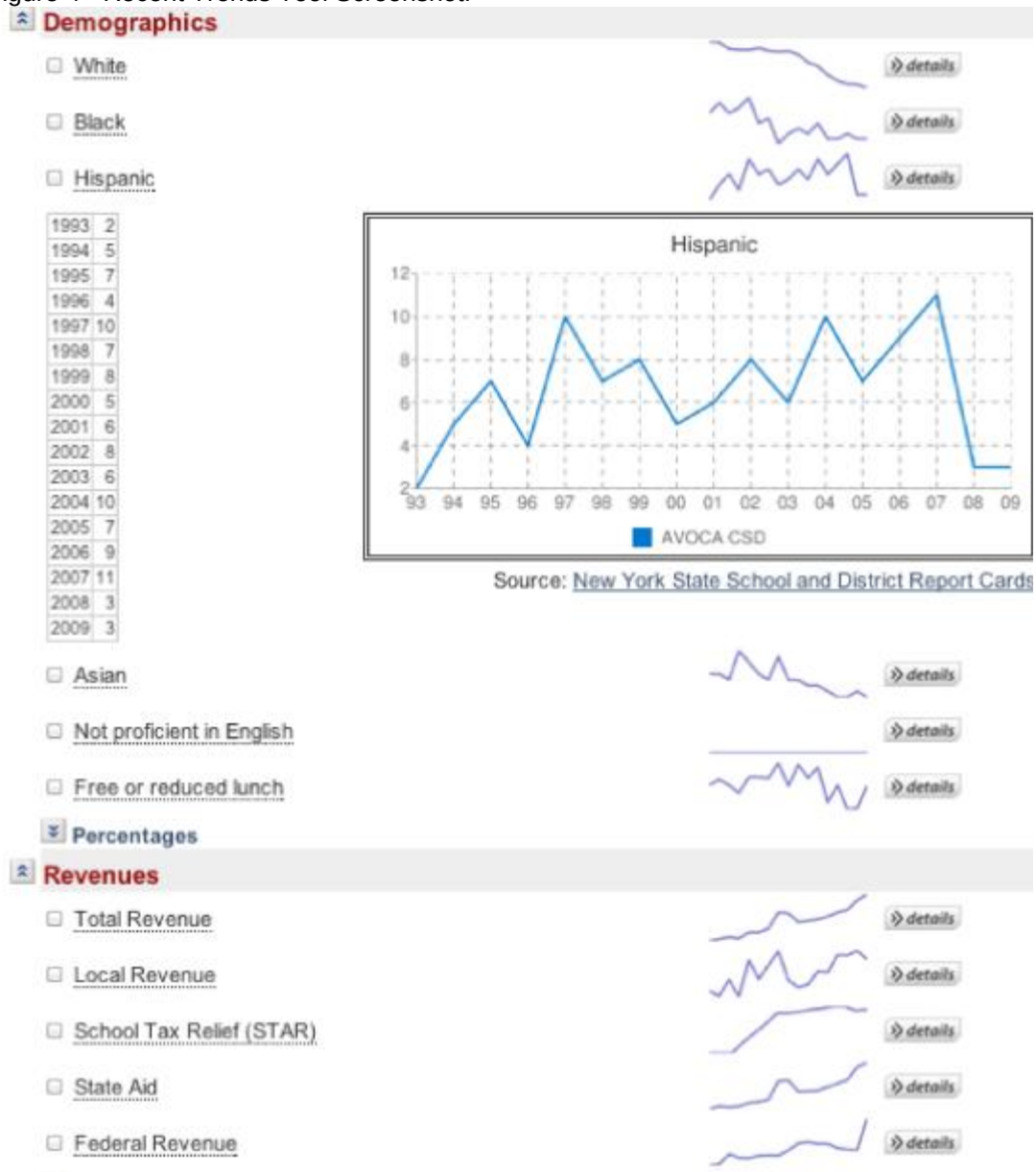


Enrollment data: table

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Total	738	739	713	714	719	717	707	700	703	694	657	652	607	586	577	596	589
K-12th	738	739	713	714	719	717	707	700	703	694	657	652	607	586	577	573	561
K-6th	438	434	414	423	429	426	409	413	410	387	335	317	289	264	264	263	270
7th-12th	300	305	299	291	290	291	298	287	293	307	322	335	318	322	313	310	291
Pre-K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	28
K'garten	51	65	58	70	69	67	49	50	44	42	38	42	39	33	44	37	35
1st grade	53	56	61	52	69	70	65	44	50	36	37	35	44	33	29	49	42
2nd grade	65	50	51	61	55	62	67	61	45	53	35	41	35	41	39	29	46
3rd grade	69	64	49	51	61	53	55	66	58	46	50	34	41	39	40	36	34
4th grade	57	70	69	51	42	68	54	59	66	58	47	52	31	37	38	39	37

The Recent Trends tool allows the user to quickly view trends in school district demographics (e.g., race, language proficiency, revenues and expenditures). Sparklines on the right hand side show a quick view, and clicking on the sparkline reveals a larger line graph (see Figure 4). The financial data can be adjusted for inflation or left in nominal dollars. Check boxes on the left of each variable can be selected for data download, mapping, or to graph multiple variables on the same line chart.

Figure 4 - Recent Trends Tool Screenshot.



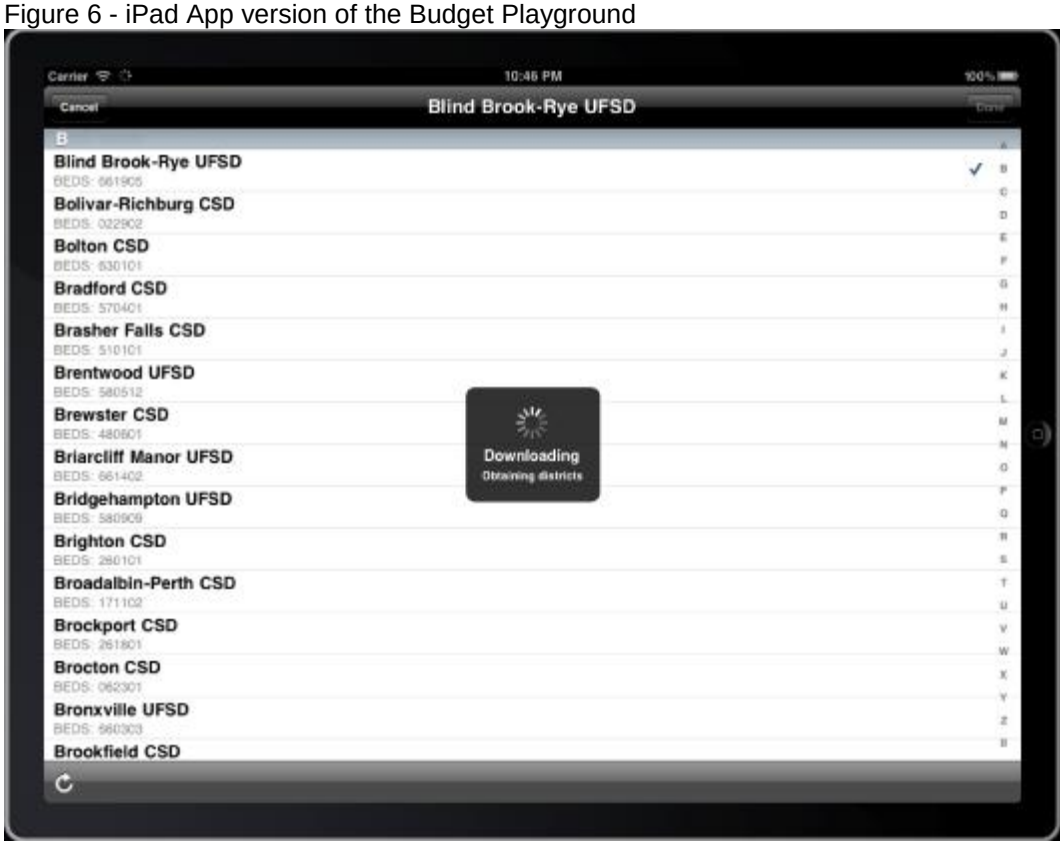
Our most innovative initiative is a financial scenario-building tool, called the Budget Playground. This tool, first built as a web app by the Program in Applied Demographics, and then enhanced in the development of an iOS app, requires more user input but also provides a powerful capability in an easy-to-use app. Drawing on the archived and audited financial data for every school district in the state, the iApp uses a REST service to call on data hosted at Cornell University. As the data is updated by the state, the service automatically draws on the new data seamlessly and in real time. Both apps have sparklines to display past trends and future projections. Both apps allow the user to save scenarios under different names. The iPad app allows for comparison of two scenarios and the web app displays a summary visualization of revenues and expenditures as well as the Fund Balance and year over year balance.

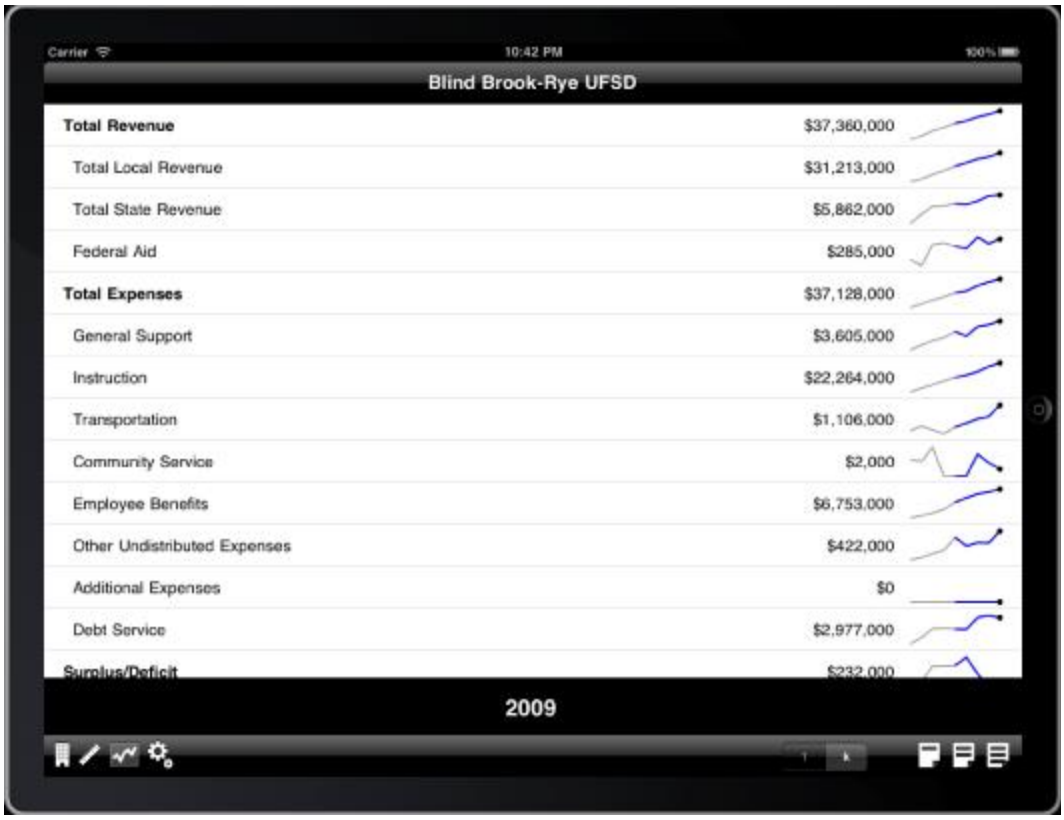
The web app allows for imputation of current data yet unavailable from the state and then the ability to insert future budget assumptions as year-to-year percentage increases or in actual dollar amounts. The balance is automatically recalculated as each data point is inserted. The user can also work at various levels of detail, depending on their interests and knowledge of the data. The user can adjust the future expenditures at the most aggregate level or select from the pull down menu a more detailed examination with more than 20 expenditure categories. From the revenue sider of the ledger, the local revenues can be adjusted by changing tax rate and/or property values. The apps provide as much detail or aggregation as the user desires.

Figure 5 - Web App version of the Budget Playground

	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015
Total Revenues	10515324	12061517	13155962	14166967	15341288	16629021	18041496	19591193
Total Expenses	10084592	11526044	12464660	13789363	15425265	17516066	20216085	23735283
Surplus/Deficit	430732	535473	691302	377604	-83977	-887045	-2174589	-4144090
Surplus/Deficit (% of budget)	4.3%	4.6%	5.3%	2.7%	-0.5%	-5.1%	-10.8%	-17.5%
Unreserved Fund Balance	519213	658342	649224	704656	764821	830123	901001	977931
Unreserved Fund Balance (% of budget)	5.1%	5.7%	5.2%	5.1%	5.0%	4.7%	4.5%	4.1%
Enrollment	673	650	600	567	541	528	525	525
Expenses per Pupil	14985	17732	20824	24320	28513	33174	38507	45210

Figure 6 - iPad App version of the Budget Playground





8 CONCLUSION AND NEXT STEPS

This paper provides a brief introduction to a set of web apps that are attempting to build the capacity of local decision makers (formally appointed and elected officials, local taxpayers and interested citizens, and members of the media). We saw a problem with lack of access to government data and have worked for three years to create and refine usable tools to meet defined needs of local decision makers. Our project is ongoing with the data updated annually as new government data becomes available, periodic enhancements to the apps themselves, and the publishing of a REST service that allows anyone in the world to access the data and use it in the development of their own web or mobile app.

The next step is to continue interactions with field users and learn how the tools are used and what changes/additions can be made to enhance the utility for the user. We have plans to do this evaluation much more formally than the organic process we have used to date. We have begun the process of promoting the use of our open database. An initial attempt was made by a group of Information Science graduate students who created a prototype app for examining school district performance using linked open data standards. The prototype is an application that links related data from disparate sources, each of which uses standardized semantics to describe different aspects the school districts. Similar to the success of the UK Edubase initiative aforementioned in this paper, the prototype illustrates the kind of utility that can be achieved by adopting standardized schemas, and has the potential to be seminal to new innovations related to making education data publicly available, discoverable, and usable in the state of New York.