

Strategic Use of Analytics in Government

John Hokkanen

Northwestern University

Predict 402, Section 63

Executive Summary

Modernity has seen the rise of big government tackling large tasks. These ventures include immensely difficult tasks like predicting the weather and immensely expensive tasks like providing government-funded healthcare. Government agencies now use analytics to further many of their missions like eliminating fraud by healthcare providers, planning revenue for legislation, protecting the public's safety, predicting a hurricane's path, and regulating companies subjected to antitrust regulations. These projects show the government's unique advantages such as the ability to invoke resources, data, and methods across multiple agencies as well as an ability to act boldly as a sovereign. However, the government also has unique challenges such as the loss of control over the budget and the possibility of a change in political leadership and agendas.

Introduction

Federal, state, and local governments spend about forty percent of the total gross domestic product in the United States. See Figure 1. Analytics commentators like Thomas Davenport observe that huge opportunities exist for governments to apply analytics to their goods and services as is done in the private sector. Davenport and Jarvenpaa (2008). In doing so, the government can accomplish goals that seem impossible, protect important programs, plan for the future, provide for the public's safety, and minimize burdens on business from mandatory reporting requirements. This paper seeks to explore examples of how different governmental agencies are tackling various kinds of problems. It will also examine the opportunities and challenges inherent in the government context. Finally, governments do not typically view themselves as competitors, at least not in the conventional sense, but, as will be shown, the Davenport and Harris model of analytical competition is useful in analyzing obstacles for these agencies.

To examine what advantages and challenges exist for government agencies seeking to implement analytics, a broad range of government projects were selected: 1) National Oceanic and Atmospheric Administration's state-of-the-art hurricane model; 2) New York Medicaid Office of Inspector General's analytics to stop healthcare fraud; 3) Congressional Budget Office's use of a long-term simulation model for calculating costs and risks of healthcare legislation; 4) Memphis Police Department's implementation of pre-crime predictive analytics for interdicting and preventing crime; and 5) the Federal Trade Commission/Department of Justice approach to corporate use of predictive analytics in preparing required submissions. By selecting both Federal and State government agencies, scientific and administrative programs,

and internal and external applications of the technologies, the paper examines a broad range of issues that confront governmental agencies.

In terms of obtaining critical information about the analytics efforts, there are unique advantages and problems in researching government projects. On the one hand, because the government typically holds no copyright on its works, agencies tend to share a great deal of details about their work on the Internet and at public meetings. Transparency and open government initiatives across all levels of government also promote, if not require, candid sharing of ideas, information, and data. This is counterbalanced by lack of access by news organizations into the agencies where investigators are typically directed to the political appointees who then serve as marketers rather than the providers of details. Fortunately, the multitude of reports issued by agencies offers some substitute, and one must be willing to live with informal, non-binding remarks because that may be all that one can obtain.

The pressure of diminished funding as well as the possibility of leadership and agenda changes pose different challenges than those found in private business. In the private sector, success of an initiative is usually rewarded, but in the government sector, changes can occur whether or not one the analytics project is building momentum or has already proven to be successful.

Achieving the Impossible

National Weather Service - Predicting Hurricanes

People have turned to the Federal government to tackle projects where the nation as a whole would benefit and where the project might be too long-term and investment too great for private commerce. Predicting the weather is such an example. The weather affects everything and everyone in the United States: elementary schools, baseball games, transportation corridors,

farmers, insurance companies, and satellite launches. And while forecasting the weather a few days later is important, predicting a hurricane's path is essential.

Hurricane Weather Research and Forecast System (HWRF)

Predicting where a hurricane will strike and the attendant rain and surge is critical for involving local authorities and saving lives and property. This became apparent after the devastating hurricane season of 1954. As a result, the National Oceanic and Atmospheric Administration (NOAA), through the National Weather Service (NWS), formed the National Hurricane Research Project which became the National Hurricane Research Laboratory. National Hurricane Research Project (2011). It operates the National Hurricane Center and its mission is to describe and predict the paths of hurricanes in the Atlantic and Pacific and to provide scientific estimates and warnings to the government of these hazards. National Hurricane Center (n.d.). NOAA Scientists in the 1950s began developing complex physical models in the Geophysical Fluid Dynamics Laboratory (GFDL). GFDL (2014). Scientists soon found that this sort of weather modeling and hurricane prediction was enormously challenging. Silver (2012). Scientists quickly discovered that small measurement or modeling errors would be magnified progressively over days so that the predictive ability of the models was limited. Silver (2012). The conclusion was that the problem required more data and simpler models. Silver (2012).

The result has been nothing less than remarkable. In 2007 the NWS has deployed its latest generation Hurricane Weather Research and Forecasting (HWRF) model for prediction. Environmental Modeling Center (n.d.) According to the NWS, "The HWRF™ is a high resolution coupled air-sea-land prediction model with a movable nested grid and advanced physics for high resolution" that has the "capability to address hurricane structure and rainfall

forecast problems in addition to advancing wave and storm surge forecasts.” Environmental Modeling Center (n.d.). The NWS intends for it to be leveraged by other models so that rain, drainage, and inland flooding can be predicted. Environmental Modeling Center (n.d.). The model not only uses data from the fixed reporting locations, it also takes real-time radar feeds from NOAA’s high altitude jets to create and maintain its models. Environmental Modeling Center (n.d.).

The approach used by the National Weather Service to create and manage the HWRF model and ultimately issue predictions shows an interesting approach to how governmental science can overcome what once appeared to be impossible. First, the responsibilities for the HWRF analytics system are distributed. Overall responsibilities lie with the NWS’s National Center for Environmental Prediction which hosts both the Environmental Modeling Center (EMC) which handles the day-to-day implementation and operation of the HWRF model, National Weather Service (n.d.), and the National Hurricane Center (NHC) which is responsible for interpreting and provisioning the model results to the public. National Weather Service (2009). Meanwhile, NOAA’s Earth System Research Laboratory oversees the GFDL which is responsible for modeling many of the core fluid dynamics algorithms. NOAA also sponsors the Development Testbed Center to link the weather research community with the governmental efforts. DTC (2014). By splitting responsibilities for various operational aspects of the HWRF model, NOAA creates focus and accountability.

In its initial years from 2002 to 2005, the HWRF model was assembled using core algorithms designed at the GFDL. Talapragada (2010). From 2005 to 2007, real-time data inputs were added to the algorithms. Talapragada (2010). Eventually the core algorithms, combined with real-time data updates, were released, though their forecasting strength was only

competitive with the core GFDL algorithms. Talapragada (2010). From 2008 to the present, the HWRF model has continued to improve but the GFDL algorithm continues to outperform it slightly, and both of these are substantially outperformed by the European model. Talapragada (2010); Masters (2014).

In 2009, NOAA stated in a document produced by the National Weather Service that the GFDL hurricane model would not be developed further, National Weather Service (2009). However, the GFDL continues to enhance its approach to hurricane prediction and its most recent 2013 upgrades have reduced bias, increased tracking accuracy, and improved the predictions of the most severe hurricanes. Bender et al (2013). Apart from the HWRF model, only the GFDL model provides specific intensity forecasts lacking in other models. Masters (2014). Masters ranks the GFDL model as better than the HWRF model and among the best in the world. But he concedes that “no other hurricane model access this wide of a range of meteorological information.” Masters (2014). Thus, NOAA’s statements that the GFDL algorithms would not receive additional funding appear to have been misspoken and NOAA appears to have ongoing support of this internal competition with the HWRF project.

NOAA’s work on the HWRF system has seen criticism. Specifically, in 2012 the NHS model predicted Hurricane Sandy’s impact on New Jersey only four days before it hit. Considering that the NHS only releases a five day forecast, this forecast might seem fairly good—until one considers that the European computer model predicted the event a full week before arrival. Vergano (2012). The Europeans leveraged satellite data from an aging U.S. polar orbiter in ways that the HWRF model did not. NOAA explained that it had planned to use a replacement satellite but the new device had gotten scrapped by the Obama administration. Vergano (2012).

In the final analysis, the NHC does not rely solely on the HWRF model but its predictions to the public are usually the average of the top models. Masters (2014). This is supported by the NWS's statement that the net result is "on average, NHC official forecasts usually have smaller errors than any of the individual models." NWS (2009). However, the NWS continues to improve the HWRF algorithm. In 2013, NHC asserts that its most recent findings (as shown in Figure 2) reveal that its newest models are more accurate than either the European model or the Global Forecast Model operated by the NWS. Brown, D. (2013).

Applying the Davenport/Harris model to NOAA is not easily done. The EMC which develops the HWRF is likely characterized as a stage four analytical company as its entire mission is built around predictive analytics. As an organization it has distinctive capabilities (technology, software models, and personnel) whose mission is to generate these tools. The NHC, as the public face of the HWRF, is also an analytic company as their employees are the ones who provide additional inputs into the model and ultimately leverage the model to provide guidance to the nation. Yet the problems revealed by Hurricane Sandy show that NOAA still has room to grow. Its failure to leverage the aging satellite data and to produce the best prediction possible--unlike the Europeans--runs contrary to the premise of analytical competition. NOAA's continued support of the GFDL work suggests that creating competition within the government may serve as important a function as competition in the private sector.

As a governmental organization, the NWS's hurricane-related aspirations may not extend beyond this analytical mission. Nevertheless, the NHC's web site shows a transparent agency committed to sharing its predictions with the public on their home web page. National Hurricane Center (n.d.). The Agency also reveals its prediction errors openly. National Hurricane Center Forecast Verification (n.d.). See Figure 3. Through the DTC, the HWRF algorithm is shared

with researchers in the US and abroad. DTC (2014). Finally, the NWS pushes its data out to other agencies. For example, the NHC provides real time updates of data directly to the Department of Energy. The DOE in turn uses the data in its Energy Disruptions web application to report storms that could affect energy infrastructure. Marshall (2013).

Achieving Better Outcomes

Stopping Fraud in Medicaid

Under the Affordable Care Act, twenty eight states have expanded their use of Medicaid. Kliff (2014). Now, approximately sixty million (60M) people have health insurance coverage through this federally-funded, state-administered health insurance program. FamiliesUSA (n.d.). Total expenditures are \$415B annually, with California's and New York's combined expenses totaling over \$100B. The Henry J. Kaiser Family Foundation (2013). A large distributed program requires oversight to prevent fraud and states are using a wide variety of analytics to protect these resources. Pew Charitable Trusts (2013b). Because New York is the largest recipient, this paper will examine some of that state's anti-fraud analytics.

First, it is crucial to understand what kind of fraud is at stake. For example, consumers who use Medicaid can do so fraudulently. New York State Office of the Medicaid Inspector General (2014). An enrollee can loan a Medicaid card to a friend or resell prescriptions paid for by the program. While there may be ways to prevent both of these, these sorts of offenses are beyond the scope of this paper.

The largest losses due to fraud is not committed by consumers but by providers who bill Medicaid and receive its reimbursement. According to the New York Medicaid Inspector General, the program's enforcement arm, this fraud falls into a few broad categories: billing Medicaid for services or tests not provided or not deemed necessary; selling prescriptions; and

providing money/gifts for the agreement to provide services. New York State Office of the Medicaid Inspector General (2014). The state warehouses all billings submitted by its providers and this data is a key asset that may be leveraged to identify improper acts.

The Problem that New York Faces

At fifty-three billion dollars of expense New York is the single largest provider of Medicaid. The Henry J. Kaiser Family Foundation (2013). The Office of Medicaid Inspector General (OMIG) is responsible for monitoring the delivery of Medicaid services. Medicaid Integrity Program (2010). With its 250+ auditors, the office audits billings to “look at the medical necessity and appropriateness of services billed to the State and the accuracy of payments.” Medicaid Integrity Program (2010). The auditors complete approximately 2000 audits annually and recover approximately \$20M per year. Medicaid Integrity Program (2010). In other words, the audited reviews discover a total of 0.24% of the total expenses as fraudulent. Yet a 2005 New York Times article cites former Medicaid investigators who suggest that ten percent of all claims were fraudulent. Levy and Luo (2005). At the same time, the U.S. government estimates a huge benefit to fraud investigation with the payoff for Medicaid fraud units to be eight to one for every dollar spent. DHHS Office of Inspector General (n.d.). It becomes apparent that potentially billions of taxpayer dollars are at stake which might be saved with analytics.

To do this, OMIG has engaged in a number of projects to do large scale data mining and data matching. “OMIG uses the State's Medicaid data warehouse, which stores five years of Medicaid claims with payments exceeding \$200 billion, to support audit initiatives by targeting problem provider behaviors, conducting pre-audit analyses, and selecting audit samples.” The Pew Charitable Trusts (2013). Using its knowledge of different provider categories and the

guidelines that govern their claim submissions, OMIG has developed algorithms to automatically identify potentially incorrect payments. Providers are given an opportunity to self-audit the billing and substantiate it or return the monies. The Pew Charitable Trusts (2013a).

Development of the Projects

The OMIG issues an annual report of its activities and these reports provide a way to examine what data warehouse analysis activities are being performed. Table 1 extracts the relevant sections on the analytics initiatives presented by the Inspector General's office from its annual reports from 2006 to 2012. This paper uses an examination of annual reports to determine what was occurring within the agency. In 2007, there had been extensive coverage of a wide variety of analytics projects. New York State Office of the Medicaid Inspector General (2007). That continued in 2008, but then had a precipitous fall-off in coverage. By 2012, the OMIG had reduced its discussion of analytics projects to a mere two sentences:

The Bureau of Business Intelligence (BBI) uses analytical tools and techniques, as well as knowledge of Medicaid program rules, to mine Medicaid claims data and identify improper claim conditions. These recovery projects are developed by BBI Claims Analysis and issued by BBI System Match and Recovery (SMR).

New York State Office of the Medicaid Inspector General (2012). To the extent that coverage in the annual report is any indicator, it appears that the early OMIG projects were likely discontinued, and OMIG is far less committed to its data mining efforts than five years before.

The CMA Corporation, which has the current contract to provide data warehouse services, presents a rosier image of New York's results. It asserts that New York "ranks #1 in the nation for fraud detection and recovery at a little over 1% of total." CMA (n.d.) Salient Corporation provides the front-end data visualization environment which allows one to perform exploratory data analysis. See Figure 4. Salient (n.d.). Salient suggests that their system easily allows investigators to discover outliers and thereby quickly identify abusers of the Medicaid

system. Salient (n.d.). In other marketing materials, Salient describes how one investigator discovered how 90% of a physician's prescriptions filled at one pharmacy were for a schedule 2 narcotic; the physician was arrested and indicted. Salient (n.d.). Another investigator examining home health aide services quickly discovered fraudulent activity among some of the outlier high billers. Salient (2011). Finally, a newspaper piece quotes the Salient as explaining that prior to its software, the OIG had few people who could extract data from the system whereas they now have plans to have 130 people trained. Times Union (2013).

Applying the Davenport/Harris model to this organization, it would be difficult to conclude that the OMIG has progressed beyond Stage 2 in having localized analytics. In order to track billings/payments, the Department of Health had to create a data warehouse to handle the massive reimbursement business which they are running. While this creates a key asset to be leveraged by OMIG, the evidence wavers on whether that office is doing so. Back in 2007 and 2008, the Inspector General was providing substantial leadership support of the analytics projects. Proof of concept projects, new technologies, and development of analytics staff reveal a commitment to use new approaches to uncover fraud.

However a change of Governors in 2010 resulted in a change in Inspector Generals, and this leadership change reveals a substantial shift away from the support of the analytics effort as shown in the annual reports. In terms of sponsorship, the analytics activities reported by Salient (and not the OMIG) appear to be functional and tactical, arising out of particular individuals' clever application of the Salient software. Without leadership, the analytical process appears disjointed and has a narrow recovery focus. The one bright spot is that Salient is seeking to train a great many people and that may move the use of the software from isolated pockets to something greater. The fact that the OMIG now credits the analytics software with recovery of

less than ½ of 1% of the total program expenditures says a great deal about how little is being done.

This example serves as a warning for those wanting to bring analytics to government problems. In this case, the size of the Medicaid program has grown and the reasons for analyzing payments of this size cannot have diminished since 2006. Rather than fraud going away, it was the political commitment to leverage these new tools that dissipated. This case study begs additional questions:

- Why does the NY OMIG, an organization that requires data to be submitted by providers in a standardized form, produce non-standardized reports where it is difficult to correlate activities from one year to the next?
- Is there a mechanism by the Federal Government under the Centers for Medicaid and Medicare (who match funds) to create accountability for the fraud detection and recovery by the states?
- Given the obvious reasons to expand the use of analytics, are external forces in play? Did Governor Cuomo receive political support from the health care institutions that submit large billings to Medicaid that may not want detailed statistically-based oversight?

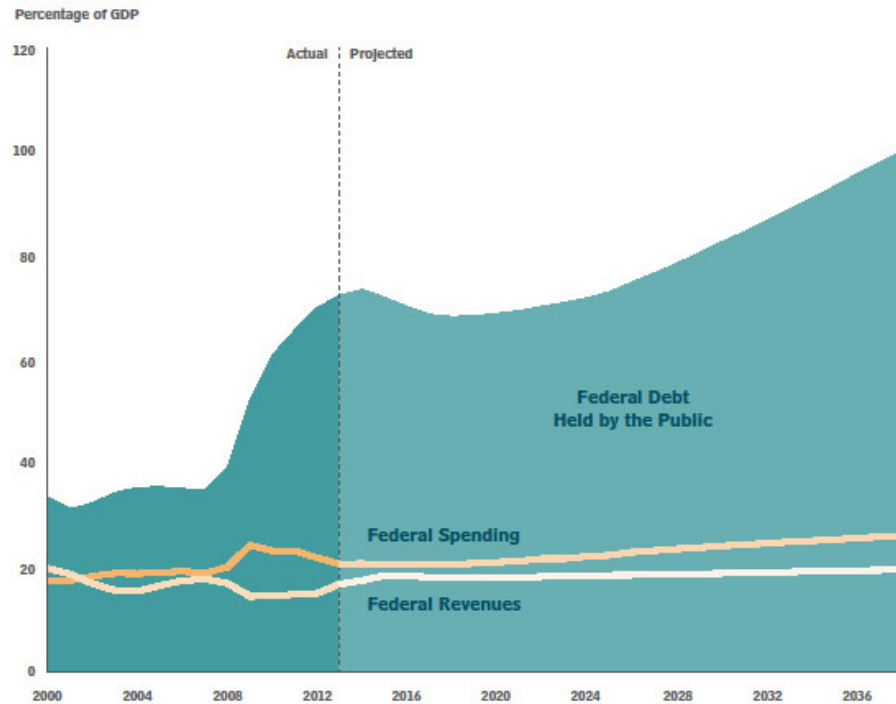
Further study of the analytics efforts of other major Medicaid recipients like California and Texas might offer additional insight. Such a comparison might serve as backdrop for understanding the requisite conditions for successful continuation of analytical projects.

Managing Our Money

The Congressional Budget Office

The Congressional Budget Office (CBO) is an arm of the Federal legislature and

produces “independent analyses of budgetary and economic issues to support the Congressional budget process.” Congressional Budget Office (n.d.) Overview. The need for this independent analysis becomes obvious when looking at the debt:



CBO's 2013 Long-Term Budget Outlook – Front Cover Graphic of The Federal Debt

With trillions of dollars of debt, the need to forecast and understand the impact of revenue collection and spending has become essential. The CBO's explicit purpose is providing impartial (i.e., nonpartisan) and objective analysis. It creates dozens of reports and hundreds of cost estimates ranging from analysis of the President's budget and scoring of enacted legislation to budget projections, economic forecasts, and budget options to decrease cost and increase revenue. Congressional Budget Office (n.d.). CBO | Our Products.

The CBO's existence is one based on the idea that analytical models help make for good governance. In generating its models, it has a number of key goals such as: “use existing evidence to make future projections; facilitate consistency and replication of methods for

estimates over time; enable timely responses to requests for estimates; and incorporate behavioral responses (if feasible).” Congressional Budget Office (2012). Modeling methods include spreadsheet models, regression models, micro simulation models, and a combination of the three. Congressional Budget Office (2012)

CBO’s Long Term Model and Healthcare

With a US deficit totaling in the trillions of dollars, one cannot overstate the importance of having accurate projection models that show the balance between revenues and expenses. In order to deliver meaningful recommendations based on well-reasoned estimates, the CBO developed “a sophisticated mathematical tool with which to analyze potential reforms to federal entitlement programs and quantify the nation’s long-term fiscal challenges.” Congressional Budget Office (2009). The CBO Long-Term Model (CBOLT) is a micro simulation model capable of providing 75 year estimates. Congressional Budget Office (2009). Because of the aging US population, increased involvement of the federal government in healthcare, longer lifespans, and an overall increase in healthcare costs, it becomes obvious why a long term model is essential for understanding healthcare issues. By 2023, healthcare spending by the federal government will reach almost 6% of GDP. Reimbursement Advisor (2014).

Technical Aspects of CBOLT

CBOLT began in 2001 as an actuarial, cell-based model which used average values for population groups classified by age, sex, and marital status. Congressional Budget Office (2009). By moving it to a micro simulation model, analysts could run thousands of simulations using different policies and assumptions. Congressional Budget Office (2009). For example, one might vary the age of retirement, deductibles for Medicare, or the lifespan of a particular age group. Consequently, with the ability to simulate the US population, economy, and federal

budget, the CBO innovated a new way of generating not only average outcomes but distributional outcomes. Congressional Budget Office (2009). For example, it could assess the impact of a policy upon an economic subgroup, women or the elderly.

The data warehouse that powers the simulation is drawn from large samples of social security data which contain some basic demographic data (date of birth, age, sex).

Congressional Budget Office (2009). Additional data is populated statistically using other data sources to provide statistical relationships (e.g., earnings data). Congressional Budget Office (2009). CBOLT has economic assumptions based on empirical data such as its use of a fifty year average historical data for productivity. Congressional Budget Office (2009).

The model then runs simulations in which it iteratively makes predictions about the behaviors of the hundreds of thousands of fictitious people under its control. People get married, have children, die, emigrate, become unemployed, and claim benefits. Congressional Budget Office (2009). Meanwhile, policy levers that can vary include parameters such as Social Security benefits, retirement age, tax rates, and changes in health cost growth for federal health programs. Congressional Budget Office (2009). Interestingly, CBOLT does not generate racial, geographic, or asset information for the people in its system. Congressional Budget Office (2009). The models are run hundreds of times allowing it to draw independently from a set of probability distributions based on time-series analyses. Congressional Budget Office (2009). By running the models numerous times, analysts are able to assess the distributions and identify where uncertainties lie. Congressional Budget Office (2009).

CBO's Analysis of Healthcare Issues

CBO in turn uses this model to generate recommendation on many topics like healthcare. Review of the CBO's analysis papers shows a refreshingly blunt style of writing with focus on

evidence-based conclusions. For example, in its recent report “Approaches to Reducing Federal Spending on Military Health Care” issued on January 16, 2014, the CBO asks the question “What is the Primary Cause of Growth in Military Health Care Cost?” Congressional Budget Office (2014). It answers that question with the following list: “1) TRICARE medical benefits expansion by lawmakers; 2) Increased utilization of Tricare due to low cost to the enrollee; and 3) Medical costs of recent wars.” It further states “the first two factors explain most of the growth...[while] the third has had a comparatively small effect.” Congressional Budget Office (2014).

The CBO’s analysis that follows for how to reduce these military health care costs is equally straightforward. For example, it analyzed multiple methods of increasing the share of costs borne by enrollees of TRICARE. Congressional Budget Office (2014). But the CBO notes that the assessment is complicated because some veterans exposed to high out-of-pocket costs might switch to other veteran health plans or others might switch to an employer plan which has a non-taxable component which would thus reduce federal tax revenues. Congressional Budget Office (2014).

The CBO has placed healthcare spending as a major section of its long-term budget outlook. The CBO states “the health care programs....are responsible for almost three-quarters of the rise in spending [of the largest programs].” Congressional Budget Office (2013). The CBO identifies three causes: 1) the aging population; 2) the rate of growth of health care costs vis-à-vis the rate of growth of GDP has resulted in excessive cost growth; and 3) expansion of Medicaid under the Affordable Care Act. Congressional Budget Office (2013). The CBO projects that the total cost in terms of percentage of GDP will rise from the current 4.5% to nearly 8% by 2035. Congressional Budget Office (2013). Of that amount, Medicare accounts

for over half of the cost. See Figure 5. In terms of financing this increase, the CBO models predict that Medicare payroll taxes which fund the Medicare portion will drop by over 10%. See Figure 6. Congressional Budget Office (2013). The CBO concludes that “the [Medicare] trust fund will be exhausted just beyond the coming decade.” Congressional Budget Office (2013). These dire predictions are made with the same dispassionate explanation as seen in any scientific experimental paper.

CBO as an Analytical Competitor

For years, commentators have asserted that the CBO’s forecasts are biased in one manner or another and untrustworthy. Reynolds (2001). Yet CBO makes great efforts to explain the basis for its forecasts and even publishes comparative data on its economic forecasting accuracy. Congressional Budget Office (2013). CBO’s processes establish an analytical framework for a fact-based budgeting process where revenues and costs are assessed realistically from established models. They explain their methods and models and all but a few of their studies are made public. This allows private organizations like the American Enterprise Institute to compete publicly with CBO’s conclusions in analyzing the legislative topics. To some extent, this writer is convinced that the allegations of bias reflect a desire by partisan elements to frame an issue one way or another to obtain a particular outcome.

In applying the Davenport/Harris model, reading the technical papers concerning CBOLT and CBO’s research papers show deep strategic insights and continuous improvement in their methods. CBO embodies what it means to bring analytics to policy. They have highly skilled analysts who use tools like SAS and STATA and, as shown by the CBOLT model, CBO desires to understand the uncertainties associated with its predictions. Their reports are so remarkably blunt that it is unimaginable that the CBO lacks any support from its top leadership to call the

problem exactly as it is seen. They clearly have developed a fact-based culture in a testing environment.

Protecting the Public

Crime Prediction

Quantitative crime analysis has a long history extending back to the 18th century. From shows like CSI (Crime Scene Investigation) the public is aware of how quantitative methods are used to prove guilt. But predictive analytic techniques are now employed by police not only to prosecute crime after it has occurred but to predict where crime will take place in the future to improve prevention and/or interdiction.

Memphis

Memphis became known for violence levels that far exceeded its size and by 2005 even the police department was admitting that “street crime had adversely affected the quality of life for our citizens.” Memphis Police Department (2011). The police department turned to technology as a tool and began developing a data driven street operation blending data, analytics, and traditional policing skills; the initiative was named Blue Crush. The police department describes the program as follows:

Operation Blue CRUSH™ has evolved from a pilot operation to a department-wide philosophy. The Police Services command staff has developed a comprehensive plan to significantly reduce street crime by deploying the Department’s resources in a sustained and proactive manner. The Department, now restructured (Felony Assault Unit, Criminal Apprehension Team, Safe Streets Task Force, etc.), will allow more effective integration of its manpower, resources and intelligence gathering to continue to reduce crime....The Blue CRUSH™ methodology can now be utilized in a sustained, integrated and efficient

manner with the emphasis on accomplishing our goal. Memphis Police Department (2011).

At the center of this blend is the police data warehouse. The warehouse captures relevant attributes of crime including type, date, day, time, and location. The warehouse also seeks to tie together disparate sorts of policing data like school truancy data. The result is a multi-layered GIS heat map based on the crime data. Memphis Police Department (2011).

But technology alone does not address crime. The Blue Crush approach is “a multi-faceted, data-driven strategy, requiring both enhanced use of statistics and a cultural and organizational shift within the department.” Janikowski (2011). Officers are responsible for getting crucial data into the system, and the department uses it to deploy officers when and where they are needed. Janikowski (2011).

Blue Results

From 2006 to 2010, Memphis saw a 26 percent decrease in property crimes and a 23 percent decrease in violent crimes by 2010. Janikowski (2011). By 2010 homicide rates dropped to the lowest the city had seen in 30 years. Janikowski (2011). Data and analysis that flowed from it could be applied in ways not originally considered. For example, call center data was combined with incident data and presented on maps. The analysis revealed how response boundaries of the various precincts could be adjusted to result in more effective response times. Law (2012).

But despite the police department’s aspirations and successful implementation, progress

was impeded. As occurred with the New York OMIG, the last few years have seen a change in the Memphis mayor as well as police commissioner. Overtime pay for the police department was taken away in 2011 and that resulted in a loss of data entry and a drop of 60% to 70% of the data analyzed by Blue Crush. Peck (2013). Serious crimes increased in 2012 by ten percent. Peck (2013). Online access to near real-time data for the public publicized in articles, Lamont (2009), no longer appears to be online. The use of Blue Crush as a primary enforcement strategy now seems to have come into question, with the new Police Commissioner wanting to focus less on Blue Crush and more on community policing. Peck (2013).

Analysis of the Memphis Police Department

In implementing Blue Crush, the police department has moved up in the Davenport and Harris model to a group that has analytical aspirations and may well be on the way to becoming an analytical agency. The Memphis Police Department integrated the approach at all levels from the cop on the beat to senior management, and have made changes to make the integration a success. For example, the data reports to senior management were standardized so that commanders could not cherry pick the data and generate misleading reports; accountability was important. IBM (2011). The approach has senior leadership support and is integrated into local law enforcement strategic plans. Memphis Shelby Crime Commission (2012, January). The police department is moving even further with their data integration by including data collected at a local university to add data layers like foreclosure and problem properties, socio-demographics, and school and community risk factors. Lamont (2009). At first blush, it would appear that the Memphis Police Department is on the move transitioning from one with analytical aspirations to being an analytical organization.

However, the recent change in mayors and police commissioners suggests a major retraction in the commitment to the analytical process. As is well documented by Davenport and Harris, an organization cannot advance along the analytics development cycle without strong support from leadership, and that leadership has wavered. Part of that issue relates to budget issues. Peck (2013). Thus, we have another instructive lesson for those wanting to promote the transformation of government with analytics. Not only are continuity of budgets essential, but programs need to survive the changes of leadership, and governmental leaders prefer to support initiatives that they have created rather than those of predecessors. Some thoughts about the ways that projects might be institutionalized within political organizations are addressed in the conclusions of this paper.

Minimizing the Heavy Hand of Government

Reducing the Corporate Load of Compliance

The Federal government regulates many industries, and large companies bear many burdens. One such burden is the Clayton Act and The Hart–Scott–Rodino Antitrust Improvements Act of 1976, 15 U.S.C. § 18a, which requires that companies notify the Federal Trade Commission and the Assistant Attorney General of the United States Department of Justice Antitrust Division of any contemplated mergers and acquisitions that exceed specified thresholds. Marjoras (2006). Such mergers and acquisitions require the tendering of documents to the FTC and DOJ to determine whether any of the antitrust laws will be violated by the proposed merger. Marjoras (2006). This is not a reporting burden similar to other reporting burdens. While companies may have to supply reports on a quarterly basis, these reports are due only upon the decision to pursue an acquisition or merger, and for that reason, the burden can be extraordinarily difficult.

In today's era of large companies with big data acquiring other large companies with big data, these regulations impose substantial burdens on the merger process. Eight years ago, in 2006, the FTC reported that it had rarely received document productions with a million documents, but in that year it received nine such productions. Marjoras (2006). Assuming a skilled review team of five can review a thousand documents per hour for relevancy, that team would take nearly a half a year to make it through the first round of review. Meanwhile, the merger must wait.

Machine Learning

Finding a few important documents in a large document set is an ideal type of problem to apply machine learning algorithms. The idea is simple: process the document corpus with the software, and let the computer act as a man-machine interface. Acosta (2012). The computer presents documents for review, the human reviewer deems them worthwhile or worthless, and the software takes notes and adjusts its algorithms. Acosta (2012). After a few thousand inputs from the human reviewer, the software typically has learned enough so that it can rank the entire million document collection by the likelihood of relevancy. Acosta (2012). Once an entire collection is ranked by most likely relevant to least likely relevant, the lawyers can sample from the pile that they intend to reject to get an idea of how many relevant documents may be in the reject pile. Acosta (2012). By consolidating the most important documents in the top ranks, the lawyers can potentially eliminate anywhere from 50 to 90% of the collection depending upon the nature of the collection.

Allowing Private Entities to Use Analytics for Submissions

The degree and manner to which the FTC and DOJ have allowed corporations seeking approval under the antitrust laws of a merger to use predictive coding for their submissions

shows a different aspect of the use of analytics in government. In this case, agency use and familiarity with these technologies involve external-facing relationships created by law. Allowing predictive coding is a theoretical win-win for both the corporations and the government. The corporations do not bear the cost of reviewing millions of documents and, consequently, the corporations need not seek to pass that burden to the government by producing a collection of theoretically responsive but irrelevant documents. The ultimate approval or rejection of the merger goes much more quickly, costs less, and business occurs.

In January 2012, the FTC proposed new rules relating to submissions. In the introductory notes to the new rules, one of the reasons for the new rules was “because ESI [(Electronically Stored Information)] is broadly dispersed and not always consistently organized by its custodians, searches, identification, and collection all require special skills and, if done properly, may utilize one or more search tools such as advanced key word searches, Boolean connectors, Bayesian logic, concept searches, predictive coding, and other advanced analytics.” Rules of Practice (2012). This led many in the legal field to think that the FTC would, in fact, explicitly recognize the use of predictive coding. Instead the finalized rules were silent on its use. One might then have expected the FTC to include some discussion of predictive coding in its “model” second request form; however, the FTC continues to promulgate a document from 2010 that remains silent on the subject. Federal Trade Commission (2010).

Meanwhile, the DOJ has not sought to amend its rules at all, but has instead released informal guidance. Specifically, its model request states, in pertinent part:

If the company or its agent uses or intends to use software or technology to identify or eliminate potentially responsive documents and information produced in response to this Request, including but not limited to search terms, predictive coding, near-deduplication,

deduplication, and email threading, the company must provide a detailed description of the method(s) used to conduct all or any part of the search....The Department strongly recommends that the company provide these items prior to conducting its collection of potentially responsive information and consult with the Department to avoid omissions that would cause the company's response to be deemed deficient. Department of Justice (n.d.).

Though more concerned with the use of search terms than predictive coding, by including the technology, the DOJ sends a signal that such technologies will be allowed provided that underlying technical details come with the production. Department of Justice (2012). In association with that release, DOJ posted a document entitled “Electronic Discovery at the Antitrust Division: An Update.” In it, DOJ states:

The promised benefit of new tools like predictive coding and concept searching is the notion that only the “really” relevant documents could be produced, instead of all the documents that are technically responsive to the Second Request. While the Division strives to be sensitive to the burden of a Second Request, the use of these tools raises obvious questions. Relevance, of course, is in the eye of the beholder. It is not at all uncommon for an investigative staff and counsel to have significant disagreement about whether a collection of documents are relevant to a particular investigation. Therefore, we believe that if the parties wish to use these platforms and tools to identify “more” relevant documents and information, as opposed to responsive documents and information, the producing party must seek a written modification of the Second Request.

To date, while the Division has discussed using these techniques on several occasions, it has never reached an agreement with a producing party to use one of these platforms or tools. I would think there is a significant opportunity to reduce the amount of material produced in response to the Division using one of these platforms or tools, however, I do believe it would require a great deal of cooperation, transparency, time, and hard work very early in the Second Request process. The first issue would be the means by which the Division could assure itself that the relevant/not relevant designations were consistent with its view of the pertinent issues. I would imagine that this could be accomplished by the producing party providing a statistically significant sample of both relevant and non-relevant documents. The Division would review those sets of documents and raise any concerns. In addition, there would have to be a mechanism by which, if an issue arose that had not been anticipated by the Division at the outset, e.g. , a natural experiment about a critical issue in the investigation, additional information relevant to that issue would be produced. As an initial framework, I believe a plan like this could serve as the basis for a meaningful negotiation between the Division and the producing party. Greer (2012).

To paraphrase, if a company wants to reduce its burden with predictive coding, then the company should be open about what it is doing, should share the results of its sampling and present the false negatives (i.e., documents that appeared in the sample that would be excluded) so that the DOJ can assess any impact before accepting the document production.

Since the publication of that statement, the DOJ approved the use of predictive coding for a merger inquiry. In 2013 Anheuser-Busch InBev and Grupo Modelo had announced a merger, and the government had sued the entities early in to prevent it. McKenna Long & Aldridge

(2013). After negotiations, the parties came to agreement that the companies would collect nearly a million documents and use predictive coding. The reported cost was half of what it would have been if a typical linear review had been performed. McKenna Long & Aldridge (2013). In published remarks, Deputy Assistant Attorney General Renata B. Hesse recently observed: “When it works well, predictive coding reduces the document review and production burden on parties while still providing the [antitrust] division with the documents it needs to fairly and fully analyze transactions and conduct.” Hesse (2013).

Justice Judged

Concluding that the DOJ or FTC has even analytical aspirations would be a stretch of imagination. There is little doubt that these agencies have smart people and localized capabilities to create analytical models to assess the transactions which they must approve. However, the lack of leadership, evidenced by the fact that DOJ publishes statements in which the authors express their own views (and not the official views of the Department) illustrate the agencies’ resistance to provide definitive guidance on these topics. Both the FTC and the DOJ are essentially handling the use of these technologies on an individual case-by-case method rather than providing a platform where business can be conducted efficiently. In either case, these Agencies could instead show leadership by selecting various software platforms, developing personnel who could assist on-site with the running of these applications at the corporate locations, and facilitating a partnership with the entities that require merger approval. A critical reader walks away thinking that both the FTC and DOJ have been dragged kicking and screaming into these new technologies.

Conclusions

This paper has explored a few of the many reasons why governments ought to integrate analytics into their broad range of services. When well executed, as in the case of the hurricane analytics platform or the Congressional Budget Office's long-term model, these tools allow the government to function at a very high level. Yet, despite huge value that could be delivered, as in the case of Medicaid oversight, predictive policing, or providing technical leadership in submissions mandated by law, some governmental organizations struggle. Part of this struggle may simply lie with the fact that the government has many stakeholders where Davenport and Harris's companies have shareholders. In the case of the government, some of those stakeholders may have interests contrary to the government's use of analytics, and may use their political leverage to stymie such projects. In the case of the FTC and DOJ, a least part of their lack of leadership arises from both a lack of internal personnel and budget to take a leadership role combined with concerns about uncertainties of committing to a particular path.

On the other hand, NOAA's hurricane prediction model and the CBO's development of long-term simulation models show that the government can not only succeed but achieve fabulous results. In NOAA's case, it has leveraged its mighty resources of personnel across multiple disciplines (e.g., fluid dynamics, modeling, and weather) as well as brought to bear a massive range of technologies (e.g., supercomputers, weather airplanes, and satellites). It has persisted with its mission over sixty years and has succeeded in developing highly reliable five day models, and, with the infrastructure that it has built for its HWRF model, it may now be able to extend its forecasts towards 8 days or longer. Moreover, the scientists at NHC use human judgment and adjust the model based on other predictions and all the information received.

Likewise, the CBO shows how effective a non-partisan, fact-based approach can work in government resource allocation. The CBO can draw upon large resources and knowledge of many agencies (e.g., SSA, DOL, Treasury) to supply them with not only the best data but access to some of the most experienced, thoughtful analysts in these various fields. By combining expertise with an independent mission accountable to neither political party, the CBO is able to generate substantial analysis which can lay the backdrop for meaningful bipartisan conversations.

The Memphis Police, NOAA, and CBO examples show how being a sovereign governmental agency provides a unique advantage in implementing these cutting edge technologies. As a governmental entity, the Memphis Police may take steps that might not be feasible for private entities. Specifically, because the government cannot typically be sued for money damages (referred to as sovereign immunity), the use of analytics for pre-crime may be possible where private entities would risk claims of discrimination. Being a permanent agency with reasonable expectations that it will not be dismantled anytime soon, NOAA can build a hurricane model for the future even if its current performance is outperformed by other models. As a government agency, it may account for long-term public goals instead of being driven only by the immediate, short-term performance goals as often occurs with private companies. And in the case of the CBO, as an independent, non-partisan agency, it may generate analysis that literally appeals to no one simply because of its nonpartisan nature. If it were a private business, it would need to establish a core set of customers and build its value proposition from that base, and as we see from “independent” think tanks like the American Enterprise Institute, they are usually associated with one side or the other.

The lessons of the Memphis Police Department and the New York Medicaid Office of Inspector General point to a very interesting question for analytics managers. How can one create projects that can both capitalize on strong leadership while minimize the downside risk of a change in leadership? Below are some initial thoughts for examination in a subsequent paper:

- Create brick and mortar capabilities with technologies, staff and processes that do not dissipate quickly simply because of changes in administration.
- Do not rely solely upon the leadership of one's own agency. Develop paths of support at sister agencies whose leadership may remain constant and who may be able to bring continued institutional support for the analytics efforts.
- Create an ecosystem of capabilities across multiple groups so that one may migrate to an alternative group to weather the change of leadership.
- Document and publicize the success of the initiatives as much as possible to thwart a rollback in policy.
- Analyze how projects may be de-personalized so that they are not strongly associated with the key leader even if he/she is essential to its support. Find and build collective identity in the project.

It would not be surprising if academic thought leaders have given consideration to the methods of institutionalizing technical projects within political organizations. Further research on this topic could yield interesting and valuable insights.

From a meta-forces perspective, it seems reasonable to say big data and governmental budgetary constraints to do more with less will force government agencies down the analytics road. Due to bureaucratic inertia it will likely take a long time for some agencies. Those

agencies that embrace technologies to become analytical agencies, if not analytical competitors, will govern better.

Figures

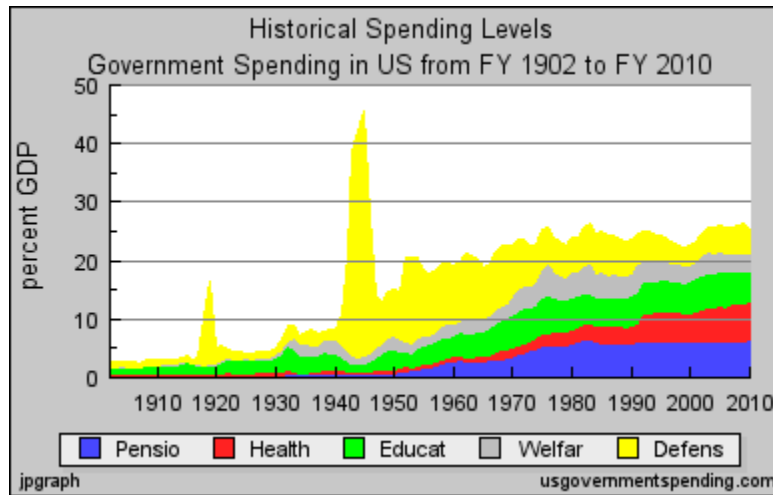


Figure 1, Government Spending As a Percentage of GDP. Government spending. (2014).



Preliminary Verification of 2012 NHC 6- and 7-day track forecasts

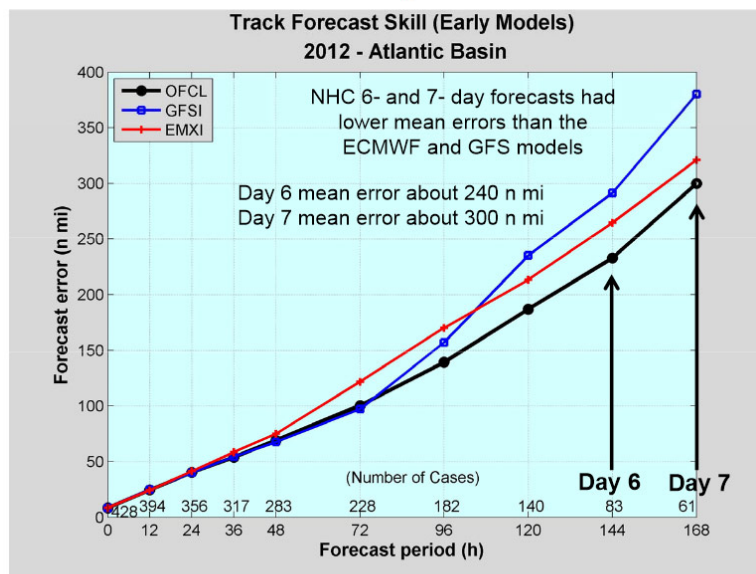


Figure 2. NHC's Claim that HWRF is now top performer. Brown, D. (2013).

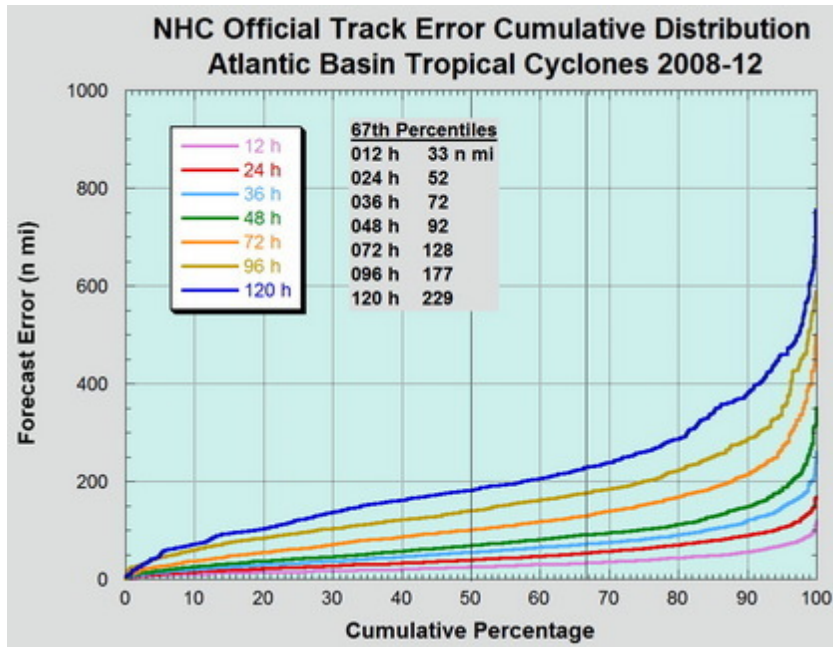


Figure 3. NHC's Five Year Prediction Error Distribution. National Hurricane Center Forecast Verification (n.d.).

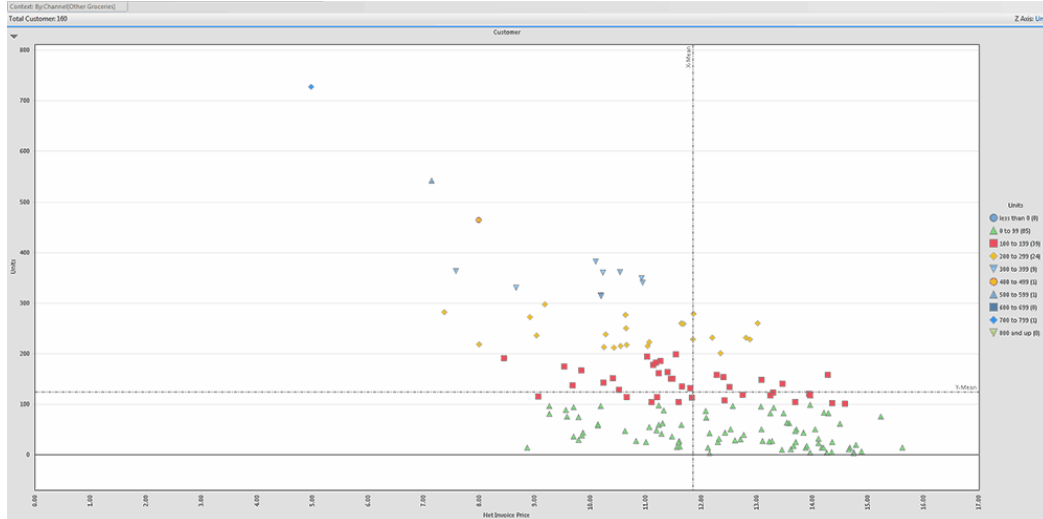
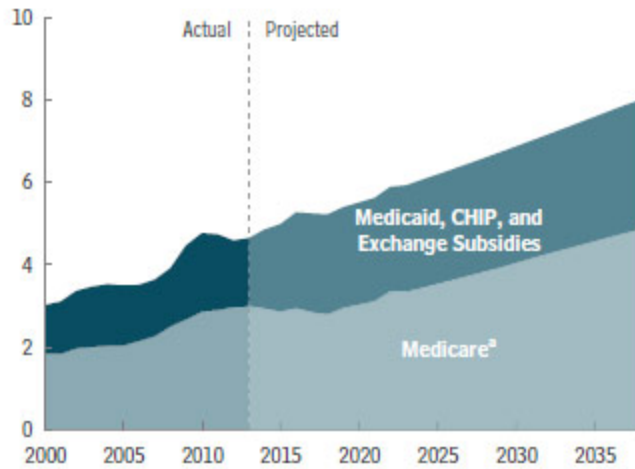


Figure 4. Salient's Data Exploration Interface. Salient (n.d.). Salient Charting Gallery.

Federal Spending on Major Health Care Programs, by Category, Under CBO's Extended Baseline

(Percentage of gross domestic product)



Source: Congressional Budget Office.

Figure 5. Medicare as a share of GDP

(Percentage of gross Medicare benefits)

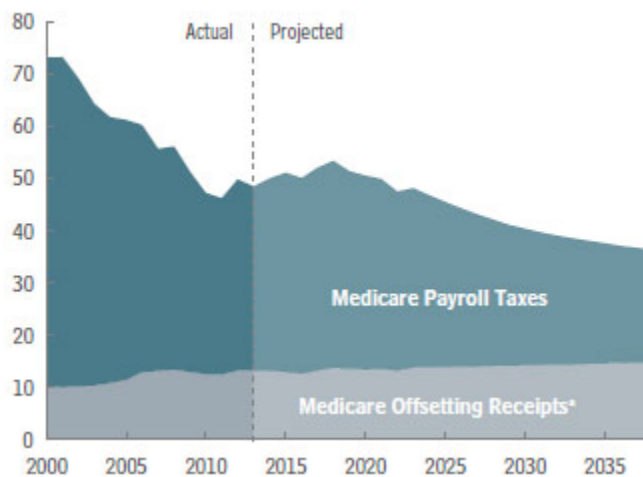


Figure 6. CBO Forecast for Medicare

Tables

The following materials are direct quotes taken from OMIG annual reports.

Year	Activities Reported
2006	<p>Phase 1</p> <ul style="list-style-type: none"> • “Evaluating current Fraud, Waste and Abuse activities, pre- and post-payment reviews, and technologies utilized through data mining and analysis • Identifying new initiatives to detect and prevent improper payments • Estimating possible savings, costs, and time • Implementing new improvements and initiatives approved by the State and assisting New York in changing current processes in order to improve recoveries and cost avoidance.” <p>Phase 2 involves “implementation of the new overpayment detection and recovery strategies resulting from the Phase I review and evaluation of the current fraud.”</p> <p>New York State Office of the Medicaid Inspector General (2006).</p>
2007	<p>Staffing and Organization</p> <p>Data mining success of noteworthy scale and consequence requires competent, dedicated staff with a myriad of skills. Dedicating the staff resources, and acquiring and synthesizing the different skill sets are the greatest challenges to creating a culture of data mining excellence. Recognizing this, the OMIG has taken several steps to improve our staffing and organization.</p> <p>The OMIG has formed a Business Intelligence Unit (BIU), currently staffed with 14 full-time employees, who service a spectrum of data needs to support the agency’s mission. Their tasks include targeting, conducting provider analysis, supporting targeting tools, creating data match algorithms and providing pre-audit analysis and audit samples. The BIU’s efforts are primarily focused on fee-for-service providers; however, an additional group with six full-time staff members provides similar data services for managed care and rate-based services.</p> <p>OMIG’s long term goal is to integrate data analysis tools, capabilities and data access into the work of every employee who performs audit, investigative and program integrity functions.</p> <p>In an effort to promote the creativity and field knowledge of the program staff while simultaneously creating a center of data mining activities and strategies, OMIG established a data mining task force to help steer data mining efforts.</p> <p>In addition to these efforts, the OMIG intends to augment the Federal Payment Error Rate Measurement (PERM) concept by perpetuating random sampling (every year, not just New York’s cycle years) to create a continuous information stream for measuring the success of program integrity efforts.</p> <p>The OMIG is continuing to invest in its pre-payment controls and monitoring. This includes legislative authorization to expand our Cardswipe and Post and Clear programs by adding 2,000 portable terminals across the state. The authorization includes increasing staff in the Cardswipe, Post and Clear and Pre-Payment Review (Edit 1141) units.</p>

Tools

New York State leads the nation in the development and use of current data mining tools to identify patterns of improper payments, provider concerns, and enrollee issues. New York's leadership is due in part to the size of the program and the volume of claims (over \$200 billion) stored in its data warehouse and available for analysis. Though New York leads the nation, the opportunities available based on analysis of this data, and data becoming available for integration from other sources, are just beginning. These opportunities extend beyond control of fraud and waste to disease management, medical error and unanticipated outcome detection, and assistance to patients in managing their medical conditions.

In the past year, the OMIG has received information from almost every leading organization in health care data mining about their capabilities, special uses of their tools and systems, and compatibility with existing New York systems in the past year. In addition, the OMIG has greatly expanded the staff and training for data mining within the organization.

In order to optimize the efficiency of the staff involved in data mining and promote increasingly complex data analysis, the OMIG has expanded the use of commercially available tools that specialize in areas of data mining. Examples of commercial products under active consideration include:

Desktop Graphical User Interface Tool. Following a successful joint pilot project, the OMIG and the Office for Health Insurance Programs are actively engaged in exploring procurement options for a data tool that presents ease-of-use through a graphical user interface, yet allows the user to make complex queries and effortlessly drill down into increasing levels of detail. This tool holds the promise of engaging a greater percentage of OMIG staff beyond the typical IT/power user audience.

IBM Entity Analytics Software (EAS). The OMIG recently conducted a pilot project with IBM to assess their EAS tool. EAS focuses on resolving entity relationships (i.e., identity attributes) from disparate data sources. The pilot demonstrated the power of the tool in a number of areas. Using a partial set of data, it uncovered numerous instances of duplicate recipients in our enrollment file. Based on the sample we estimate that more than 22,000 duplicates are on file. The OMIG is currently pursuing purchasing the tool and related integration services.

Fraud Abuse Management System (FAMS). The FAMS sensitively ranks providers in relation to their peers within specific geographic regions, specialties and subspecialties, and any other parameters users choose to define. Results are displayed in a three dimensional graph format that readily identifies providers who fall outside norms. Numerous features and attributes can be measured and weighted to build complex models that assess behaviors in a selected peer group. FAMS supports further investigation by enabling users to drill down into detailed information on providers' claiming practices.

The OMIG has owned IBM's FAMS for more than ten years. However, this tool has not always suited the agency's needs. Based on a review of usage and some of the issues that have contributed to dissatisfaction with the tool, OMIG engaged IBM to address the concerns and customize the FAMS to better meet the agency's needs. Through a formal work engagement with IBM, the OMIG has asked the company to include an upgrade of several versions of the product, a switch from IBM's DB2 to an Oracle database (OMIG database administrators can

now support it) and a more seamless ability to extract data from our data warehouse into the FAMS data format. In addition, we are making a more conscious effort to dedicate and train staff in the use of the product.

New York State Office of the Medicaid Inspector General (2007).

2008

New York Leads the Way

Data Mining

A cornerstone of the OMIG's strategy to detect and prevent fraud, waste and abuse in the Medicaid program, is to continually use technology to detect behaviors, control point of service transactions, review select claims and provide agency staff with critical support data.

Bureau of Business Intelligence

The OMIG continues to emphasize the creation of a center of expertise in the area of data mining and other data support functions. By consolidating with another unit, the Bureau of Business Intelligence (BBI) has grown to 26 staff members who provide support services to meet the agency's mission. Their tasks include targeting, conducting provider analysis, supporting targeting tools, creating data match algorithms and providing pre-audit analysis and audit samples. In addition, the BBI performs hundreds of desk audits annually. These audits (a/k/a system matches) are based on algorithms designed with specific knowledge of various provider types and the guidelines that govern their claim submissions.

OMIG's long term goal is to integrate data analysis tools, capabilities and data access into the work of every employee performing audit, investigative and program integrity functions.

In an effort to promote the creativity and field knowledge of the program staff while simultaneously creating a center of data mining activities and strategies, OMIG established a data mining task force to help steer data mining efforts. The key areas of OMIG's data mining focus over the past year are highlighted below.

Tools

Data Warehouse - New York State's Medicaid Data Warehouse continues to be our most valuable resource for data mining. The warehouse stores five years of Medicaid claims with payments exceeding \$200 billion. Tools inherent within the system include a graphical user interface which assists users in the compilation of queries. More sophisticated users have access to the data through the use of structured query language which allows for more complicated queries. As the OMIG has expanded and matured the capabilities of the BBI, the bureau's ability to leverage this important resource has grown correspondingly.

Desktop Graphical User Interface Tool - Following a successful joint pilot project, the OMIG and the Office of Health Insurance Programs are engaged in exploring procurement options for a new data tool. This tool allows ease-of-use through a graphical user interface, yet allows the user to make complex queries and effortlessly drill down into increasing levels of detail. This tool holds the promise of engaging a greater percentage of OMIG staff beyond the typical IT/power user audience. Though we have incurred delays in the procurement of this tool, at this writing, we have begun the formal procurement process and anticipate an implementation in late 2009 or early 2010.

IBM Entity Analytics Software (EAS) - The OMIG also conducted a pilot project with IBM to assess their EAS tool. EAS focuses on resolving entity relationships (link analysis) from disparate data sources. The pilot project demonstrated the power of the tool in a number of areas. Using a partial set of data, EAS uncovered numerous instances of duplicate recipients in our enrollment file. Based on the sample it is estimated that more than 22,000 duplicates are on file. The OMIG has purchased the tool and has an active procurement for expert services in progress to integrate the product and develop the data

sources and linking logic.

Data Sources

Though the New York State Medicaid data warehouse represents a huge investment and a powerful tool in support of data mining, it is essential that additional data sources be acquired in order to maximize the OMIG's ability to detect fraud and abuse.

Recent efforts included projects utilizing full state vital statistics data. The OMIG is working with the Office of Health Insurance Programs, using state vital statistics data to make substantial improvements in the accuracy and timeliness of the State's processes for matching Medicaid recipients and providers to vital statistics death data.

The BBI continues to do Medicaid-to-Medicare matching through the federally sponsored Medi-Medi project. Through this project BBI identifies duplicate payments between the two programs.

The Bureau of Payment Controls and Monitoring, Medicaid System Controls, is also working with the New York City Taxi and Limousine Commission to match their data, which provides global positioning system (GPS) coordinates for the geographic start and end points of ambulette trips.

Collaboration

A key challenge to maximizing data mining efforts is to ensure that a two-way exchange of support between data mining staff and field staff from our Divisions of Audits and Investigations exists. Some key examples of this type of collaboration are outlined below.

Customized Audit Samples - To support our field auditors, BBI staff routinely prepare audit packages consisting of the audit sample, universe and provider-specific support data. In some cases, it is advantageous for audit and BBI staff to discuss the characteristics of certain audit universes. In a number of instances, staff have discussed audits and with the assistance of a peer expert in statistics, developed stratified audits to ensure that any associated findings result in a full and accurate projection of the overpayments owed to the Medicaid program.

General Clinic Match Project - Field auditors spoke with data mining staff about findings they encountered while performing Outpatient Department (OPD) audits. Based on these discussions, a systems match project was initiated. The auditors noted many instances where the provider was submitting separate claims for services that should have been provided as part of the patient's clinic visit.

Data mining staff created algorithms to identify these potentially inappropriate services. During 2008 BBI identified approximately \$6 million of overpayments; of which, just over \$1 million was recovered during 2008.

Dental Matches - Data mining staff, in conjunction with the dental unit in OMIG's Division of Medicaid Audit, developed several algorithms to identify inappropriate claims. These matches identified services such as radiography, cleanings, fillings and extractions for recipients without teeth. Medicaid claims should not be submitted for relines, rebases and repairs of dentures within the first six months the recipient has them. These services are included in the initial fee for dentures. The recovery estimate for these inappropriate payments is approximately \$2.3 million.

Intensive Psychiatric Rehabilitation Treatment (IPRT) - The Office of Mental Health provides patients with a time limited intensive psychiatric rehabilitation treatment program, with active psychiatric rehabilitation designed to assist persons in forming and achieving mutually agreed upon goals in living, learning, working and social environments. Program guidelines limit reimbursement to 72 hours monthly or 720 hours yearly. Data mining staff

identified instances where these limits were exceeded. This project identified approximately \$150,000 in inappropriate claims.

Hospice Analysis - Audit staff requested analysis of the entire universe of Hospice billing. Data mining staff supplied data on multiple variables. From this analysis, a single Hospice provider was identified as an outlier in almost every dimension of analysis. Audit staff further collaborated with federal auditors to embark on a combined Medicaid/Medicare audit of that specific provider.

Net Applicable Monthly Income (NAMI) – OMIG’s Division of Medicaid Audit requested assistance with the analysis of a Bronx nursing home’s application of their residents’ NAMI. Data mining staff did an extensive analysis of this provider which resulted in a referral to the New York State Attorney General. The outcome of this case is pending. With the knowledge learned during this analysis, staff applied the logic to all providers of this type. Further cases are being prepared and field work is expected to begin in 2009.

Investigations

Unraveling the complexities within the Medicaid program that can lead to fraud, waste and abuse requires an interwoven system of investigation. All Division of Medicaid Investigation (DMI) units and projects focus on four main areas that address the integrity of the Medicaid program: (1) fraud, waste and abuse; (2) cooperation with other entities; (3) deterrence; and (4) quality of care. Several forward thinking and unique projects are in process within DMI that demonstrate how New York leads the way.

DMI has an outstanding Provider Surveillance and Utilization Review System (PSURS) staffed with a certified coder, who utilizes complex computerized queries to compare providers to their peers, and medically trained experts, such as nurses and a dental hygienist, that provide a unique skill set to detect fraud, waste, and abuse in providers that render and order services for Medicaid recipients. DMI’s PSURS Unit fulfills the requirements of 42CFR 456.3 which states the Medicaid agency must implement a statewide surveillance and utilization control program that safeguards against unnecessary or inappropriate use of Medicaid services and against excess payments; assesses the quality of those services; provides for the control of the utilization of all services provided under the plan; and provides for the control of the utilization of inpatient services.

The PSURS Unit identifies trends in the medical industry and specializes in off-label prescription use, learns about the latest trends in therapy that may not be approved, and detects quality of care issues. The PSURS staff members utilize their individual medical expertise and experience to look behind the billing and find providers who match certain criteria and may be more likely to commit fraud or provide poor quality of care. In 2008, 39 of these providers were referred to the DMI Provider Investigation Unit for further field work. Additionally, SURS referred three providers to the Attorney General’s MFCU.

In 2008, the acquisition of a dental hygienist expanded the range of the PSURS reviews. For the first time, dentists treating recipients in a private practice and/or dental clinics were studied by a trained dental hygienist. Aberrant billing practices were subjected to the same analysis and investigative measures as other providers. Dental records, including radiographs, were examined to ensure that recipients are treated well and the quality of care meets professional standards.

Undercover investigations are an excellent tool for discovering and confirming suspicions of fraud, poor quality of care, and billing problems. The Undercover Investigations Unit’s findings have resulted in arrests, prosecutions, exclusions, terminations, and penalties. Numerous groups, inside and outside the OMIG, rely on the Undercover Investigations Unit to assist in their investigations.

Undercover investigators seek services from Medicaid providers. They are equipped with pseudonyms, Medicaid cards to match the pseudonyms, and document the provider's conduct during the undercover operation (UCO). The provider's subsequent claims are reconciled with the investigator's written report. Differences between the evidence obtained by the undercover investigator and the provider's claims receive additional scrutiny.

Undercover operations are conducted based on information gleaned from various targeting tools such as Provider SURS, results of sending out explanation of medical benefits reports to recipients, telephone hotline calls, internet complaints, and anonymous reports. Information provided by the Enrollment Audit Review (EAR) Unit and Provider Investigations are also used to direct UCOs. During 2008, DMI undercover investigators conducted 2,192 UCOs.

During 2008, UCOs obtained evidence proving that some enrolled pharmacies and durable medical equipment (DME) providers wrongfully billed for services and products dispensed by non-enrolled pharmacies or DME providers. The act of claiming to have serviced recipients when another person or entity actually serviced the recipient is not permitted and is considered an unacceptable practice, false filing, and potential fraud. Both enrolled and non-enrolled providers who participate in such schemes face exclusion or termination from the Medicaid program and criminal sanctions.

The OMIG excludes providers who pose a risk to the integrity of the Medicaid program. Protecting recipients from health care professionals who fail to provide appropriate care or fail to maintain the necessary standards prompts the OMIG to exclude individuals and entities from participating in the Medicaid program.

The OMIG takes action against individuals and entities for program-related criminal indictments, convictions, patient abuse or neglect, licensing board disciplinary actions, or for engaging in any practices considered unacceptable under the Medicaid program. Under certain circumstances, a lesser action is censure. The OMIG utilizes a consistent and fair approach to ensure a just outcome based on the individual facts surrounding the case. Censured providers are monitored to ensure integrity in the Medicaid program. These providers are considered high-risk and the particular underlying issue that led to the censure is scrutinized along with other areas that are traditionally abused.

Various federal, state and local agencies provide information used to determine whether exclusion is warranted. The OMIG reacts rapidly to criminal indictments, convictions, and licensure actions to remove providers who defraud the system or provide poor quality of care. New York licensed providers who are excluded by other states are considered for exclusion from the New York State Medicaid program.

DMI identifies providers for potential exclusion by maintaining a cooperative relationship with district attorneys in New York State; conducting internet searches that reveal Medicaid and health care-related arrests; gathering information concerning investigations conducted by other agencies at various Health Care Task Force meetings and meetings with MFCU, OSC and other state agencies; and reviewing actions taken against health care professionals by the United States Attorneys.

When the Office of Professional Medical Conduct (OPMC) or the State Education Department (SED) takes an action, they provide the OMIG with consent orders on those cases. The OMIG obtains the full investigative file and considers the underlying facts to make an independent decision on whether exclusion is warranted. Clinical experts consult with the exclusion specialists as appropriate.

These and all other DMI programs combine to ensure the integrity of the Medicaid program,

protect the most vulnerable population and the investment of New York State's taxpayers.

2009 New York State Office of the Medicaid Inspector General (2008).
Data Warehouse - New York State's Medicaid Data Warehouse continues to be OMIG's most valuable resource for data mining. The warehouse stores five years of Medicaid claims with payments exceeding \$200 billion. Tools inherent within the system include a graphical user interface which assists users in the compilation of queries. More sophisticated users have access to the data through the use of Structured Query Language (SQL) which allows for more complicated queries. As OMIG has expanded and matured the capabilities of the BBI, OMIG's ability to leverage this important resource has grown correspondingly.

Desktop Graphical User Interface Tool – OMIG has completed the procurement of Salient Corporation's Medicaid MuniMinder Software and associated Data Center Hosting Services. This tool provides ease-of-use through a graphical user interface, yet allows the user to make complex queries and effortlessly drill down into increasing levels of detail. This tool holds the promise of engaging a greater percentage of OMIG staff beyond the typical IT/power user audience. OMIG production use of the tool began in July, 2010.

Link Analysis Software - Following a successful pilot of IBM's Entity Analytics Software (EAS), OMIG completed a purchase of the product as well as expert consulting services to install and implement our initial algorithms. The software specializes in resolving entity relationships (e.g. identity attributes) from disparate data sources. The initial implementation logic included the identification of duplicate Medicaid recipients, deceased recipients, deceased providers, providers who are recipients and connections between providers/business associates who have been sanctioned. Staff are currently analyzing the results and making necessary adjustments to the logic and match thresholds. Once completed, OMIG will address the initial match population and conduct periodic match runs for each of the described algorithms. Staff are also working on additional data feeds and matches.

2010 New York State Office of the Medicaid Inspector General (2009).
Innovative Data Mining Capabilities
A cornerstone of OMIG's strategy to detect and prevent fraud, waste, and abuse in the Medicaid program is to avoid costs by detecting behaviors, controlling point of service transactions and determining which claims to select for review. This is accomplished by effective data mining. In 2010, OMIG began using Salient's Medicaid MuniMinder data analysis tool. This tool allows OMIG to improve targeting methods used to further fraud detection and analysis. Some of the results highlighted in this report were the result of using of this new tool.

Staff within OMIG's Bureau of Business Intelligence provide data related services to support the agency's mission, but also engages in provider analysis through Systems Match and Recovery (SMR) activities. SMR activities identify overpayments made to providers by performing data match reviews and then collecting the overpayments identified by each match. During 2010, SMR initiated a total of 1,095 provider reviews with recovery activity totaling \$14,348,391.

2011 New York State Office of the Medicaid Inspector General (2010).
The Bureau of Business Intelligence (BBI) uses analytical tools and techniques, as well as knowledge of Medicaid program rules, to mine Medicaid claims data and identify improper claim conditions.

In conjunction with OMIG's Division of Medicaid Investigations (DMI), BBI staff provided

data analysis and data mining services using the Salient Interactive Miner. This solution assisted staff in preparing analysis directed at potential abuse of certain high cost medications. The analysis looked at analgesics, narcotics, and HIV medication; and, focused on both the providers ordering the drugs as well as the pharmacies dispensing the drugs. The resulting data was used to support the efforts of a multi-governmental task force.

The document identified \$4.3M in recovery from dental match and ambulatory care, and \$170M in improper pharmacy claims.

New York State Office of the Medicaid Inspector General (2011).

2012

The Bureau of Business Intelligence (BBI) uses analytical tools and techniques, as well as knowledge of Medicaid program rules, to mine Medicaid claims data and identify improper claim conditions. These recovery projects are developed by BBI Claims Analysis and issued by BBI System Match and Recovery (SMR).

New York State Office of the Medicaid Inspector General (2011).

The document identifies a total savings of about \$23.6M. New York State Office of the Medicaid Inspector General (2012).

References

- Acosta, A. (2012, October). Predictive Coding: The Beginning of a New E-Discovery Era, RES GESTAE, 8-9.
- Bender, M., Marchok, T., Ginis, I, Thomas, B., Yablonsky, R., Tuleya, R. (2013). Upgrades to the Operational GFDL/GFDN Hurricane Model. Retrieved from http://www.nhc.noaa.gov/jht/ihc_13/s5b-02-bender.pdf
- Brown, D. (2013) Possible Future Enhancements to NHC Forecast Products. Retrieved from http://www.nhc.noaa.gov/outreach/presentations/nhc2013_nhcFutureProducts.pdf
- CMA (n.d.). New York State Department of Health (DOH) – Medicaid Data Warehouse and OHIP Datamart. Retrieved from <http://www.cma.com/PDF/MDW%20Marketing%20Information%20%281%29.pdf>
- Congressional Budget Office (2007, October). CBO's Health Insurance Simulation Model: A Technical Description. Retrieved from <http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/87xx/doc8712/10-31-healthinsurmodel.pdf>
- Congressional Budget Office (2009, June). CBO's Long-Term Model: An Overview. Retrieved from <http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/103xx/doc10328/06-26-cbolt.pdf>
- Congressional Budget Office (2012, November 16) Using Analytic Models and Communicating Their Findings. Retrieved from <http://www.cbo.gov/sites/default/files/cbofiles/attachments/43739-MIDAS-11-2012.pdf>

Congressional Budget Office (2013). CBO's Economic Forecasting Record: 2013 Update.

Retrieved from <http://www.cbo.gov/sites/default/files/cbofiles/attachments/43846-ForecastingRecord.pdf>

Congressional Budget Office (2013, October 31). The 2013 Long-Term Budget Outlook.

Retrieved from <http://www.cbo.gov/publication/44521>

Congressional Budget Office (2014, January 16) Approaches to Reducing Federal Spending on

Military Health Care. Retrieved from <http://www.cbo.gov/publication/44993>

Congressional Budget Office (n.d.) Overview. Retrieved from [http://www.cbo.gov/](http://www.cbo.gov/about/overview)

[about/overview](http://www.cbo.gov/about/overview)

Congressional Budget Office (n.d.). CBO | Our Products. Retrieved from

<http://www.cbo.gov/about/our-products>

Crowley, C (2011, October 25). Taking pulse of Medicaid costs. Retrieved from

<http://www.timesunion.com/local/article/Taking-pulse-of-Medicaid-costs-2236429.php>

Davenport, T. and Jarvenpaa, S. (2008, April) Strategic Use of Analytics in Government.

Retrieved from <http://www.businessofgovernment.org/sites/default/files/StrategicAnalyticslidedeck.pdf>

Dennis Barry's Reimbursement Advisor (2014, January). CBO Ideas to Reduce Federal Health

Care Expenditures.

Department of Justice (n.d.). Model Request. Retrieved from [http://www.justice.gov/atr/](http://www.justice.gov/atr/public/220239.htm)

[public/220239.htm](http://www.justice.gov/atr/public/220239.htm)

Developmental Testbed Center (2014). DTC || Hurricane WRF Tutorial. Retrieved from

<http://www.dtcenter.org/HurrWRF/users/tutorial/tutorialannouncement.php>

DHHS Office of Inspector General (n.d.) Annual Summary of Grant Expenditures. Retrieved from http://oig.hhs.gov/fraud/medicaid-fraud-control-units-mfcu/expenditures_statistics/fy2011.asp

Environmental Modeling Center (n.d.). A Bit of Background on the HWRF. Retrieved from <http://www.emc.ncep.noaa.gov/?branch=hurricanes&tab=mis>

FamiliesUSA (n.d.) Medicaid. Retrieved from <http://familiesusa.org/issues/medicaid/>

Federal Trade Commission (2010, June). Model Request for Additional Information and Documentary Material. Retrieved from <http://www.ftc.gov/sites/default/files/attachments/merger-review/guide3.pdf>

GFDL (2014). Brief History of Global Atmospheric Modeling at GFDL. Retrieved from <http://www.gfdl.noaa.gov/brief-history-of-global-atmospheric-modeling-at-gfdl>

Government spending. (2014, January 31). In *Wikipedia, The Free Encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Government_spending&oldid=593303554

Greer, T. (2012). Electronic Discovery at the Antitrust Division: An Update. Retrieved from http://www.justice.gov/atr/public/electronic_discovery/281388.htm

Hesse, R. (2013, February 8). IP, Antitrust and Looking Back on the Last Four Years. Retrieved from <http://www.justice.gov/atr/public/speeches/292573.pdf>

IBM (2011, May 6). Memphis PD: Keeping ahead of criminals by finding the “hot spots” Retrieved from http://www-01.ibm.com/software/success/cssdb.nsf/CS/GREE-8F8M7J?OpenDocument&Site=default&cty=en_us

Janikowski, R. (2011, March). The American City & County 126.3.

- Kliff, S. (2014, January 29). The four most important states to watch on Obamacare's Medicaid expansion. Retrieved from <http://www.washingtonpost.com/blogs/wonkblog/wp/2014/01/29/the-four-most-important-states-to-watch-on-obamacares-medicaid-expansion/>
- Lamont, J. (2009, September). GIS helps government chart a forward course. KMWorld September 2009.
- Law, J. (2012, December 14). Memphis Police Department Station Realignment – 2013. Retrieved from <http://www.youtube.com/watch?v=nJWH3YnOfXU>
- Levy, C. and Luo, M. (2005, July 18) New York Medicaid Fraud May Reach Into Billions. Retrieved from <http://www.nytimes.com/2005/07/18/nyregion/18medicaid.html?pagewanted=all>
- Marjoras, D. (2006, February 16). Reforms to the Merger Review Process. Retrieved from <http://www.ftc.gov/sites/default/files/attachments/merger-review/mergerreviewprocess.pdf>
- Marshall, P. (2013, August 2) Real-time map mashup shows storm threats to energy sources. Retrieved from http://gcn.com/articles/2013/08/02/energy-information-administration-mashup.aspx?admgarea=TC_BigData
- Masters, J. (2014) Hurricane Forecast Computer Models. Retrieved from <http://www.wunderground.com/hurricane/models.asp>
- McKenna Long & Aldridge LLP (2013, December). Predictive Coding: A Cost-Effective Approach to Antitrust Compliance. Retrieved from <http://www.mckennalong.com/publications-advisories-3465.html>

Medicaid Integrity Program (2010, December) Final Report. Retrieved from

[http://www.cms.gov/Medicare-Medicaid-Coordination/Fraud-Prevention/
FraudAbuseforProfs/Downloads/NYfy10.pdf](http://www.cms.gov/Medicare-Medicaid-Coordination/Fraud-Prevention/FraudAbuseforProfs/Downloads/NYfy10.pdf)

Memphis Police Department (2011). BLUE C.R.U.S.H. Retrieved from

<http://www.memphispolice.org/blue%20crush.htm>

Memphis Shelby Crime Commission (2012, January). Action Agenda with Detailed Action

Plans 2012 – 2016. Retrieved from [http://operationsafecommunity.org/assets/
1294/osc_2012-2016_action_agenda_with_detailed_action_plans_final_102212.pdf](http://operationsafecommunity.org/assets/1294/osc_2012-2016_action_agenda_with_detailed_action_plans_final_102212.pdf)

National Hurricane Center (n.d.) NHC Mission & Vision. Retrieved from

<http://www.nhc.noaa.gov/mission.shtml>.

National Hurricane Center (n.d.). Home Page. Retrieved from <http://www.nhc.noaa.gov/>.

National Hurricane Center Forecast Verification (n.d.). Official five-year mean errors and distributions. Retrieved from <http://www.nhc.noaa.gov/verification/verify4.shtml>.

National Hurricane Research Project. (2011, December 28). In Wikipedia, The Free

Encyclopedia. Retrieved 08:18, February 3, 2014, from [http://en.wikipedia.org/w/
index.php?title=National_Hurricane_Research_Project&oldid=468079546](http://en.wikipedia.org/w/index.php?title=National_Hurricane_Research_Project&oldid=468079546)

National Weather Service (2009, July). Technical Summary of the National Hurricane Center

Track and Intensity Models. Retrieved from [http://www.nhc.noaa.gov/pdf/
model_summary_20090724.pdf](http://www.nhc.noaa.gov/pdf/model_summary_20090724.pdf)

National Weather Service (n.d.). Hurricane Modeling at EMC. Retrieved from

<http://www.emc.ncep.noaa.gov/index.php?branch=HWRF>

New York State Department of Health (2011) A Plan to Transform the Empire State's Medicaid Program. Retrieved from http://www.health.ny.gov/health_care/medicaid/redesign/docs/mrtfinalreport.pdf

New York State Office of the Medicaid Inspector General (2006). 2006 Annual Report. Retrieved from <http://www.omig.ny.gov/images/stories/annual%20report%202006.pdf>

New York State Office of the Medicaid Inspector General (2007). 2007 Annual Report. Retrieved from <http://www.omig.ny.gov/images/stories/annual%20report%202007.pdf>

New York State Office of the Medicaid Inspector General (2008). 2008 Annual Report. Retrieved from http://www.omig.ny.gov/images/stories/annual_report_2008.pdf

New York State Office of the Medicaid Inspector General (2009). 2009 Annual Report. Retrieved from http://www.omig.ny.gov/images/stories/2009_annual_report_10-19-10.pdf

New York State Office of the Medicaid Inspector General (2010). 2010 Annual Report. Retrieved from http://www.omig.ny.gov/images/stories/annual_report/annual_report_2010.pdf

New York State Office of the Medicaid Inspector General (2011). 2011 Annual Report. Retrieved from http://www.omig.ny.gov/images/stories/annual_report/2011_annual_report.pdf

New York State Office of the Medicaid Inspector General (2012). 2012 Annual Report. Retrieved from http://www.omig.ny.gov/images/stories/annual_report/2012_annual_report.pdf

New York Office of the Medicaid Inspector General (2014). Fraud/Abuse. Retrieved from <http://www.omig.ny.gov/fraud/fraud-abuse>

- Peck, C. (2013, January 27). Blue CRUSH controversy. Retrieved from <http://www.commercialappeal.com/news/2013/jan/27/blue-crush-controversy/?print=1>
- Reynolds, A. (2001). Don't Trust the CBO's Numbers. Retrieved from <http://www.cato.org/publications/commentary/dont-trust-cbos-numbers>
- Rules of Practice, 77 Fed. Reg. 3191 (January 23, 2012)
- Salient (n.d.). Like A Video Game That You Can't Lose. Retrieved from <http://www.salient.com/medicaid-program-integrity-omigcase-study-4/>
- Salient (n.d.) Salient Charting Gallery. Retrieved from <http://www.salient.com/our-technology/visual-data-mining/sim-charting-gallery/>
- Salient (2011). NYC Audit Team Finds 'Beauty' in Salient. Retrieved from http://www.salient.com/docs/case-study_OMIG-3.pdf
- Tallapragada, V. (2010, June 22). NCEP Operational Hurricane WRF (HWRF) Modeling System. Retrieved from http://www.mmm.ucar.edu/wrf/users/workshops/WS2010/presentations/session%202/2-2_vijay_wrf_users_workshop_2010.pdf
- The Henry J. Kaiser Family Foundation (2013). Total Medicaid Spending FY 2012. Retrieved from <http://kff.org/medicaid/state-indicator/total-medicaid-spending/>
- The Pew Charitable Trusts (2013a, January 31). Medicaid Anti-Fraud and Abuse Practices. Retrieved from <http://www.pewstates.org/research/data-visualizations/medicaid-anti-fraud-and-abuse-practices-85899446210> (online database)
- The Pew Charitable Trusts (2013b, March 6). Combating Medicaid Fraud and Abuse. Retrieved from <http://www.pewstates.org/research/reports/combating-medicaid-fraud-and-abuse-85899445796>

Vergano, D. (2012, October 30). U.S. forecasts late arrival stirs weather tempest. Retrieved from <http://www.usatoday.com/story/weather/2012/10/30/sandy-hurricane-models/1668867/w>