
MEMORANDUM

TO: VERMONT LEGISLATIVE JOINT FISCAL OFFICE
FROM: DANIEL SMITH / P&C SOFTWARE SERVICES, LLC
SUBJECT: IT REVIEW – STATEWIDE DATA INTEGRATION (“BIG DATA”)
DATE: JANUARY 22, 2019

“Big Data” has been a frequent topic of discussion within the JFO, the Agency of Digital Services, and the legislature. While the term may not be completely appropriate (since it is more commonly used for commercial applications), the understanding is that we are referring to the integration of various existing and potential state data sources that when combined can lead to better outcomes for state citizens. This memo will provide background on the subject area, a review of related activities in Vermont and other states, and recommendations for additional actions that can be undertaken by both Executive Branch agencies and the Legislature itself. Although “Data Integration and Analysis” may be a more accurate term, for the remainder of this memo I will use the term “Big Data” in the interest of brevity.

1. Subject Overview

“Big Data” is a term that is similar to “Artificial Intelligence” in that it can mean different things to different people, and the meaning can change over time. However, it is generally accepted that the term refers to the creation, gathering, and analysis of data from disparate sources that when combined yield new and beneficial insights. Big Data is characterized by high levels of Volume, Variety, and Velocity: Volume reflects the amount of data, Variety reflects the different types (text, visual, audio, etc.) and sources of the data, and Velocity reflects the frequency of changes in the data. As an example, a company such as Walmart may collect data on millions of in store transactions daily, and combine it with other sources such as surveillance video, online purchases, and others to develop information on collective and individual shopping and purchasing patterns. This analysis is then used to improve the company’s performance in areas such as sales and customer satisfaction.

Within the realm of state government the levels of Volume, Variety, and Velocity are generally much lower than in the commercial world. While there are numerous data sources that may be relatively large, the actual sizes and rate of change do not compare to the commercial environment. However, the same goals exist in both: combining disparate data sources and analyzing the aggregation in order to improve organizational effectiveness and service. For this reason, while the more accurate term for state level activities may be “Statewide Data Integration”, the term Big Data may still be used without it being too far wrong.

A term that is closely related to Big Data is “Data Governance”. This concept is defined as those organizational activities and responsibilities involved in maintaining data quality and consistency. Such activities include defining and documenting data ownership, availability, security, and management. Before any large scale Big Data effort is undertaken, proper data governance must be established to answer such questions as:

- How do we verify the accuracy and trustworthiness of data to be aggregated?
- Who does the aggregated data belong to, the originator or the aggregator?
- What security measures are needed to protect the aggregated data, and who is responsible for them?
- How will the data originator achieve benefits from the aggregation of their data?

A comprehensive, documented Data Governance policy must be viewed as a prerequisite to any Big Data activities that may be undertaken.

2. Activities in Other States

A review of Big Data activities in other state governments shows differing levels of focus and achievement. Given the variability in the definition of Big Data, some states may also be engaging in that type of activity but not reporting it as such. However, most of the states researched, including Vermont, have some level of large data integration efforts underway. In each case the activities undertaken reflect the basic purpose of Big Data analysis: the aggregation of data from multiple sources to yield beneficial insights that improve service and efficiency.

A listing of the states and related Big Data activities is available at the end of this memo, however the following are a few noteworthy successes:

Multi-State:

- Alabama, Florida, Georgia, Louisiana and Mississippi: National Accuracy Clearinghouse. This multi-state effort reduces dual participation in the Supplemental Nutrition Assistance Program (SNAP) through the pooling of individual state registration data. If used nationwide, it is anticipated that the activity could save up to \$193.4 million annually.

Single State:

- Vermont, Florida, Iowa, and Utah: Big Data Transportation Analysis. These states use big data to anticipate infrastructure deterioration ((highways, bridges, etc.), improve crash analysis, and improve traffic flow.
- Alabama: Prenatal Opioid Exposure. This partnership between Blue Cross and Blue Shield and AxialHealthcare of Tennessee will “help identify women at risk of opioid abuse before they become pregnant and deliver babies that may require medical treatment for dependence”.
- Alaska: In partnership with LexisNexis this effort will detect possible fraud in state oil dividend payouts by identifying those claiming to be Alaskans but with signs of residency in other states.
- California: A central repository for clinical data from all six UC health systems will bring benefits such as:
 - The ability to predict who may not respond well to metformin, the usual first line treatment for type 2 diabetes
 - Better screenings of candidates for revision surgery following a joint replacement
 - Assessments of heart failure risk among HIV positive individuals with chronic liver disease
- Connecticut: HealthIT and Business Intelligence (BI) framework will provide information that will assist the Department of Social Services in designing programs and delivering care matched to the individual person.
- Maine: The state's Health Information Exchange has positioned itself as “a big data analytics expert, providing population health management and data exchange services to all of the state’s hospitals and a growing number of its ambulatory providers”.
- Missouri: Investments in big data and population health tools “has helped providers in the state identify the needs of a notoriously vulnerable patient group and better monitor other chronic comorbidities”.
- Nevada: Nevada Governor’s Office of Economic Development (GOED) and the Desert Research Institute (DRI) in Reno have partnered with IBM to use big data analytics and advanced research applications to improve the higher education system and improve upon scientific research capabilities.

- New Hampshire: Data Analytics Platform for the Opioid Crisis. This activity will compile “pre-existing data across multiple state agencies to make educated policy decisions to handle this crisis and help effected residents”.
- New Jersey: “The NJHA Center for Health Analytics, Research, and Transformation (CHART) will apply predictive modeling and analytics to multiple sources of data to better understand underlying socioeconomic and community issues that may impact patient access to care and long-term outcomes”.

3. Desirable Outcomes

When done correctly, Big Data activities should result in measurable improvements in effectiveness, efficiency, and service. Examples of these outcomes include:

- Data Integration: Provide better service through the development of complete customer perspectives. Integrating data from various sources allows service personnel to see the client’s big picture. This may include health, mental health, child care and wellbeing, education, disabilities, etc.
- Predictive Analytics: Use integrated data beyond the individual level to identify patterns and predict trends, allowing proactive efforts to minimize future problems.
- Cost Reduction: Use integrated data and analysis to identify waste, duplication, and fraud in supported activities.

4. Risks and Issues

As with any large activity, the primary risks with pursuing a Big Data effort are failure to achieve the desired results and a waste of time and dollars. However, this type of activity also brings into play other issues that must be considered. The three primary issues are:

- Data protection: As data from disparate sources is integrated and analyzed, the protection and security on this consolidated data must be at least as great as the most secure source data.
- Individual privacy – As data from disparate sources in integrated and analyzed, the collected data may present more of a picture of a client than that person would like. Should individuals be given access to the consolidated data to approve or deny use?

- Data ownership – As data from disparate sources is integrated, who owns the consolidated data? It may be the source agency, the collecting agency, or the persons involved.

All of these issues indicate that Data Governance is a key concern, and a statewide data governance plan should be developed and approved prior to making any significant efforts in the Big Data subject area.

5. Vermont Activities, Achievements, and Plans

Within Vermont state government data integration and analysis at the Big Data level has generally been initiated by, and limited to, single agencies. For example, the Agency of Transportation is using Big Data in conjunction with Artificial Intelligence to predict ways of improving infrastructure maintenance and traffic flow. While there is some cross-agency integration of large data sets (between the Agency of Education and the Agency of Human Services, for example), this is more the exception than the rule. In most cases separate databases within an agency are combined and analyzed to support specific activities, and not to develop new capabilities or analysis. That said, the benefits of Big Data type activities are generally recognized and sought after, even if they have not been actively pursued or achieved at the statewide level. For example, for years there has been a vision within AHS for a system that allows case workers to obtain a holistic view of supported citizens, combining health, dependent, disability, and corrections data into a single record. This type of integration could potentially provide better service to individuals, while also supporting analysis activities that may identify trends, strengths, and weaknesses in service. While this vision has not yet been achieved, it is included as part of the cornerstone goals for evolving efforts such as the Integrated Eligibility and Enrollment system.

With regards to Data Governance, the need for this has been established for several years. However, while efforts in this area are ongoing, at this time there is no documented and approved policy regarding data governance at the agency level or above.

6. Agency of Digital Services Recommendations

The State CIO and the Chief Data Officer have the primary responsibility for initiating and executing any large, statewide Big Data activities. The following recommendations are provided as a starting point for approaching a statewide data integration and analysis effort.

- Inventory: Using internal and external resources create an inventory of all potential data sources. This would include state-maintained databases, transaction and financial information, web usage, and anything else that

- **Assessment:** Make use of experienced external resources to develop and assessment of how Vermont stands in terms of data management. Such an assessment would investigate and report on Vermont’s capabilities with regards to data management, data governance, data warehousing, business intelligence, and data analytics.
- **Vision:** Based on the inventory and assessment completed previously, develop a “wish list” of what could be accomplished in terms of statewide data integration and analysis. This should be highly imaginative, and involve input from a variety of interested parties. At this stage technical and budgetary constraints should not be considered; instead assume that magic is possible.
- **Plans:** Take the vision of what could be and apply realistic constraints to determine what is feasible in the near and term. This would result in a prioritized list of what we could really hope to achieve, and after applying resource constraints (time, dollars, people) develop the plans that detail how to move from potential to actual benefits. As with any other large project, the first step in the planning cycle would be to develop a business case that includes a cost-benefit analysis for each proposed activity.

7. Legislative Recommendations

A common theme that has emerged in the review and analysis of State IT projects is while Vermont technical staff may be adequate to support day to day operations and small scale projects, large and complex projects generally cannot not be accomplished internally without exceeding the acceptable risks. Big Data / Statewide Data Integration activities will probably also fall into this category, and without external help may not be successful. If this subject area is deemed important to the Legislature, the following near term actions are recommended for the Legislature:

- Provide clear guidance to the Agency of Digital Services that this is a subject area that the Legislature is interested in;
- Provide additional funding for ADS to effectively initiate activities (inventory, assessments, and vision);
- Provide both oversight and support to ADS in determining the plan of action to transition the vision to reality.

8. Potential Measures of Success

The following measures of success are suggested if and when Vermont elects to undertake work in the Big Data area:

1. The creation of an inventory of all State databases and other data sources that might be used for Statewide Integration and Analysis;
2. The completion of a statewide data assessment that reports on Vermont's capabilities with regards to data management, data governance, data warehousing, business intelligence, and data analytics;
3. The existence of a comprehensive, documented, and approved plan that details the reason for undertaking specific activities, including how the task will be approached, what the costs/benefits are, the initial timeline, and the key performance indicators for the activity;
4. The completion of at least one large, cross-Agency integration and analysis effort that yields measurable benefits in either cost or client support.

Further Reading

[NASCIO: Is Big Data a Big Deal for State Governments?](#)

[NASCIO: Data Strategy for State Governments](#)

[Big Data Can Help States Create Good Policy](#)

[Cross Agency Data Sharing](#)

Big Data Activities by State

The following table lists Big Data activities in the various states. It is intended to be representative, not comprehensive, and includes activities that are ongoing or projected as well as completed efforts.

Search	Big Data Activities	Other Activities or Related Links
Alabama	Prenatal Screening	National Accuracy Clearinghouse
Alaska	Oil Dividend Payout Fraud	
Arizona		Access Barriers to Big Data
Arkansas	Commission on Economic Competitiveness	
California	UC Medical Center Patient Care	
Colorado	Colorado Information Marketplace	
Connecticut	Department of Social Services Analysis	
Delaware		
Florida	National Accuracy Clearinghouse	Gartner Data & Analytics Summit 2019
Georgia	Georgia Tech Medical Imaging Study	National Accuracy Clearinghouse
Hawaii		
Idaho		
Illinois		
Indiana		
Iowa	DOT Smart System	
Kansas		
Kentucky		
Louisiana	National Accuracy Clearinghouse	
Maine	Health Information Exchange Analytics	
Maryland		The Ethics of Big Data
Massachusetts		
Michigan	Big Data in Patient Care	
Minnesota		
Mississippi	National Accuracy Clearinghouse	
Missouri	Tracking Behavioral Health Needs	
Montana		
Nebraska		
Nevada	Big Data and Economic Development	
New Hampshire	Big Data and the Opiate Crisis	
New Jersey	Big Data Health Care Analytics Center	
New Mexico		
New York		
North Carolina		

North Dakota		
Ohio	Cleveland Health Disparities	
Oklahoma		
Oregon		
Pennsylvania		
Rhode Island		
South Carolina		
South Dakota		
Tennessee		
Texas		
Utah	Big Data and Highway Safety	
Vermont	Predictive Transportation Improvement	
Virginia		
Washington		
West Virginia		
Wisconsin		
Wyoming		