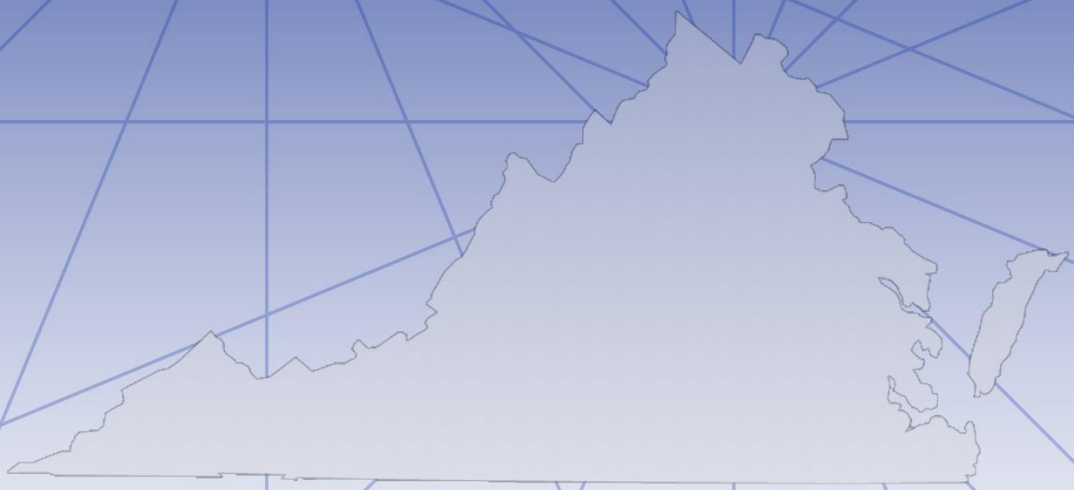
Virginia Information Technologies Agency

Executive Directive 7: *Leveraging the Use of Shared Data and Analytics*

Final Report



January 25, 2017

www.vita.virginia.gov

Table of Contents

Executive Directive 7 Final Report
Executive Summary

In May 2016, Governor McAuliffe issued Executive Directive 7: *Leveraging the Use of Shared Data and Analytics* ("ED7") to promote greater utility and accessibility of data assets maintained by state agencies. ED7 lays out the following strategic objectives linked to agency data sharing, governance, and analytics:

- Enhancing government transparency
- Streamlining business processes
- Increasing operational efficiency and effectiveness
- Minimizing duplication and overlap of current and future systems development

Barriers to Data Sharing

Analysis conducted under ED7, which leveraged analysis from Executive Directive 6: *Expanding Cyber-Related Risk Management Activities* (2015), revealed only 22.2% of the 1,686 enterprise data assets – high value data assets with demand for sharing – held by state agencies responding to the survey were shared outside of the source agency, and that attempts to increase data sharing to meet the ED7 objectives remain blocked by a complex array of federal and state laws, regulations, program rules, and related policies.

For example, several provisions in the *Code of Virginia* place general restrictions on agencies seeking to share data. State statutes also significantly limit the use of shared data in some contexts, such as provisions in the law that prevent agencies from using data on citizens except for the purpose for which the data was collected.

Apart from these general statutes, an extensive array of laws, regulations, and policies have been established to govern the sharing of specific types of data. These vary greatly and their application depends, in part, on the type of data being shared, who is sharing the data, and with whom the data is being shared.

State agencies, faced with such legal complexity and the associated risk of sharing data in a noncompliant manner, the penalties for which may include civil or criminal penalties, have developed a risk-averse culture. In addition, agencies often struggle to find the necessary resources – technical, financial, and personnel – to sustain data sharing relationships.

Regardless of these restrictions, the Commonwealth has a wide range of opportunities to pursue the currently un-shared data.

Data Sharing Hotspots

The ED7 analysis identified five (5) data sharing "hotspots" at the Secretariat level – with a hotspot being defined as a Secretariat whose data sharing activity makes up roughly 10 percent or more of the state government total shared assets. These five (5) hotspot clusters accounted for 71.7% of the state's entire amount of data sharing activities.

Data sharing observed in the ED7 analysis consisted of three types: data sharing between state agencies, data shared at the agency's prerogative as open data, and data shared upon formal request from the public, such as under the Freedom of Information Act (FOIA). Agencies have experienced an increased demand for data sharing to drive enhanced agency efficiency, reduce data and infrastructure costs, and build capacity for informed decision making.

Secretariat (Hotspots in *Italics*)	Total Data Assets	Data Assets Shared	% Data Assets Shared	% State Govt. Total
Health & Human Resources	*517*	*80*	*15.5%*	*21.3%*
Transportation	*230*	*59*	*25.7%*	*15.7%*
Natural Resources	*114*	*57*	*50.0%*	*15.2%*
Public Safety & Homeland Security	*142*	*37*	*26.1%*	*9.9%*
Finance	*316*	*36*	*11.4%*	*9.6%*
Agriculture & Forestry	62	33	53.2%	8.8%
Administration	63	30	47.6%	8.0%
Education	112	27	24.1%	7.2%
Commerce & Trade	74	9	12.2%	2.4%
Technology	50	7	14.0%	1.9%
Veterans & Defense Affairs	6	--	0.0%	0.0%
Total	**1,686**	**375**	**--**	**100.0%**

Most Used Data Analytics Tools

Secretariats using the most analytics tools included Public Safety and Homeland Security with 28 unique tools, Health and Human Resources with 24, Education with 18, and Natural Resources and Transportation each with 13. The primary uses of the tools tended to be for statistical analysis, data visualization, and business intelligence.

Secretariat	Number of Unique Analytic Tools
Public Safety & Homeland Security	28
Health & Human Resources	24
Education	18
Natural Resources	13
Transportation	13
Agriculture & Forestry	9
Commerce & Trade	9
Finance	9
Administration	6
Technology	6
Veterans & Defense Affairs	5

A majority of state agencies reported using Microsoft Excel as their primary analytic tool, with a frequency of use at 26 instances. This confirms that all agencies have, at a minimum, at least one data analytics asset at their disposal, given that Excel comes standard in the Microsoft Office suite of applications. Nearly all Excel users said the application remained a "strategic" asset for their analytics capabilities, and more than half of the Excel users rated themselves at the "mastery" or "advanced" level of expertise.

Best Practices

The ED7 analysis identified several best practices in the areas of data sharing and data analytics. The most salient best practices reported by agencies were as follows:

Best Practices – Data Sharing

- Engage legal counsel to review and document compliance requirements
- Enforce compliance through audits on source and downstream data systems
- Establish compliant requirements for physical and logical access controls
- Provide regular training and technical support relating to compliance
- Adopt standardized templates or trust frameworks for data sharing agreements
- Implement restricted use agreements to control use of shared data
- Require compliant electronic authentication for data access [§ 59.1-550 et seq.]
- Design data sharing interfaces to conform with external data exchange standards

Best Practices – Data Analytics

- Follow data analytics methods, standards, and established techniques
- Inspect data quality, integrity, values and constraints
- Implement established methodologies for analytics
- Adopt methods, formats, and data visualization techniques aligned with requirements
- Align data analytics requirements with data governance models

Recommendations

Recommendation 1: Open Data, Data Accessibility, and Data Utilization by State Agencies

Recommendation 1.1. Dedicate OAG legal support to agencies to assist in determining whether data may be classified as "open" data
Recommendation 1.2. Invest in the Virginia Open Data Portal to enhance accessibility, ease of use, and capacity
Recommendation 1.3. Improve discovery and access to high value open datasets for state agencies and the public
Recommendation 1.4. Invest in state-level licensing for data analytics, business intelligence, and data anonymization applications

Recommendation 2: Data Sharing Governance, Ethical Use, and Authority

Recommendation 2.1. Continue to support the state government's enterprise data governance program and explore the advantages in creating a senior enterprise data leader position
Recommendation 2.2. Adopt a policy that defines the role of the Data Owner and establishes the obligations for data sharing and governance
Recommendation 2.3. Perform ongoing Data Management Maturity (DMM) assessments for agencies across domains of the state government
Recommendation 2.4. Publish the results from the state-wide data asset inventory in a searchable repository to promote discovery and accessibility

Recommendation 3: Data and Analytics Projects to Promote the New Virginia Economy

Recommendation 3.1. Establish a process to identify potential projects for business case development that align with the Governor's Policy Priorities
Recommendation 3.2. Require agencies to incorporate a "Data Plan" into their Information Technology Strategic Plans
Recommendation 3.3. Projects recommended for future consideration pursuant to the Executive Directive

Projects recommended in this report for future consideration have been highlighted due to their potential for realizing value of data and analytics, generating potential cost savings, aligning with the Governor's Policy Priorities, and supporting the vision of a New Virginia Economy, as required by the Executive Directive. VITA identified the projects through input from the Office of the Secretary of Technology, the VITA Executive Leadership Team, and agency representatives during the stakeholder focus groups. The projects were not chosen through a formal selection process, nor were they scored using objective criteria. Review of future data analytics projects should be led by the Commonwealth's enterprise data governance office, as stated in this report under Recommendation 2.1.

Section 1. Background, Report Scope, and Statutory Authority

"Increasing the use of shared data and analytics among Virginia agencies through a comprehensive and coordinated effort will improve the provision of services and outcomes, maximize the use of resources, and increase the return on investment of our citizens' tax dollars in their government."

Governor Terence R. McAuliffe
Executive Directive 7 (2016)

1.1 Background

In May 2016, Governor McAuliffe issued Executive Directive 7: *Leveraging the Use of Shared Data and Analytics* ("the Executive Directive") to promote greater utility and accessibility of information assets collected and maintained by state agencies. The Executive Directive encouraged state agencies to take a more systematic approach to using shared data and analytics as a means of improving services and outcomes, maximizing agency resources, and increasing return on investment for citizen tax dollars. The Executive Directive's overarching goal centered on continuing the Commonwealth's advancement toward a New Virginia Economy.[1] A copy of the Executive Directive has been provided as **Appendix 1**.

The Executive Directive set out broad strategic goals to be accomplished through an increased capacity by state agencies for data sharing, correlation, and analysis. These included (1) achieving efficiencies in the administration of state programs and services, and (2) allowing state government to more efficiently and effectively address public health, public safety, education, and quality of life outcomes. Tied to these strategic goals were four primary objectives linked to agency data collection, data sharing, and analytics:

- Enhancing government transparency
- Streamlining business processes
- Increasing operational efficiency and effectiveness
- Minimizing duplication and overlap of current and future systems development

While the Executive Directive set a framework for data utility and accessibility, the Governor's action recognized that state agencies must take active, persistent measures to protect the privacy and security of citizen-centric information. "State government shall continue to protect individual privacy, adhere to applicable state and federal regulations, and cybersecurity best practices during any activity involving the collection of sensitive information," Executive Directive 7 (2016). The Executive Directive also established that state agencies must ensure the ethical use of data, regardless of the degree of sensitivity.

1.2 Report Scope

This report has been prepared by the Virginia Information Technologies Agency (VITA) under the direction of the Chief Information Officer (CIO) of the Commonwealth, acting on behalf of the Secretary of Technology and the Secretary of Finance, as called for in the Executive Directive. The report summarizes findings from VITA's analysis of information submitted by more than 300 stakeholders, representing all 63 executive branch agencies, three (3) statutory committees, and seven (7) institutions of higher education. The report also offers recommendations across the core requirements stated in the Executive Directive: data sharing, data utility and accessibility, and data analytics to drive the New Virginia Economy.

[1] Governor McAuliffe's policy statement on the "New Virginia Economy" may be accessed at https://governor.virginia.gov/media/3501/new-virginia-economy-12052014.pdf

VITA implemented multiple data collection methods as part of its research program to gather information for this report. These included two (2) separate structured survey instruments: one focused on data assets and sharing, the second on data analytics, implemented to executive branch agencies; a series of focus groups with primary stakeholders, including executive branch agencies from across Secretariats and institutions of higher education; work sessions and briefings with statutory committees responsible for advising the Secretary of Technology and the CIO of the Commonwealth; and direct agency engagement through VITA's Customer Account Managers (CAMs). Details on VITA's data collection methodology and review process have been provided in **Appendix 2**.

1.3 Previous Data-Related Activities

The report builds upon a data analytics and governance framework that has been under development within the state government since 2011. The following initiatives reflect a portion of the existing framework and the state government's accomplishments, to date. Artifacts from these milestones and deliverables have been provided in **Appendix 3**.

- Secretarial Committee on Data Sharing: Committee formed in September 2011 by the Secretaries of Technology and Health and Human Resources to explore opportunities and constraints for an enterprise data-sharing agreement for state agencies, built on a trust framework governance model.

- Commonwealth Enterprise Information Architecture (EIA) Strategy: The Secretary of Technology in August 2013 adopted an enterprise data strategy, developed with input from agency leaders, business managers and technical leads. Strategic goal areas: Data governance, data asset management, data standards, and data sharing.

- Data Exchange Standards for Interoperability: The Secretary of Technology and CIO of the Commonwealth, to date, have adopted more than 130 data exchange standards to promote interoperability and sharing of data in a compliant, standardized manner. Standards cover administrative data for core operations of state government, as required by the 2008 Appropriation Act; health information, on recommendation from the Commonwealth's Health IT Standards Advisory Committee (HITSAC) pursuant to § 2.2-2699.7; and the National Information Exchange Model (NIEM) for citizen-centric data, to meet requirements under Item 427 of the 2012 Appropriation Act.

- Data Stewards Groups: In February 2014, the Commonwealth inaugurated three data steward groups – Executive, Functional (Business), and Technical Data Stewards – to support ongoing agency engagement and direction for implementation of the EIA Strategy and related data governance activities.

- Governor's Data Internship Program (GDIP): The Office of the Governor and the Secretary of Technology in the fall semester of 2014 implemented the internship program to pair interns from state universities with state agencies to perform advanced analytics on "real-world" problems.

- Next Generation (NextGen) Analytics Pilot Program: The Commonwealth in November 2015 established contracts with 11 vendors to supply next-generation analytics services at zero cost to state agencies. These services cover both products and the resources needed to utilize those products.

- Commonwealth's Open Data Portal: The Secretary of Technology has worked with the Library of Virginia and VITA to develop a public-facing open data portal, located at http://www.data.virginia.gov/. The open data portal supports the discovery, accessibility, and utilization of the state's open data assets.

- Governor's Datathon: The Office of the Governor and the Secretary of Technology in 2014 hosted the first annual "Datathon" challenge to promote the use of open data and data analytics. Teams representing state agencies, local governments, universities, and private industry have competed to build applications and analytics toolsets aligned with the Governor's Policy Priorities.

The data analytics and governance framework resulting from these milestones and deliverables not only helped to inform the agency engagement and analysis presented in this report, the framework may serve as a potential starting point for the state government to act on the report recommendations.

1.4 Statutory Authority

The following sections in the *Code of Virginia* establish the statutory authority for the entities responsible for this report, and for submitting to the Governor the recommendations required under the Executive Directive. References to statutes in this document shall be to the *Code of Virginia*, unless otherwise specified.

Secretary of Technology
§ 2.2-225. Position established; agencies for which responsible; additional powers
http://law.lis.virginia.gov/vacode/title2.2/chapter2/section2.2-225/

Secretary of Finance
§ 2.2-211. Position established; agencies for which responsible; additional powers
http://law.lis.virginia.gov/vacode/title2.2/chapter2/section2.2-211/

Chief Information Officer (CIO) of the Commonwealth
§ 2.2-2007. Powers of the CIO
http://law.lis.virginia.gov/vacode/2.2-2007/

Virginia Information Technologies Agency
Chapter 20.1. Virginia Information Technologies Agency
http://law.lis.virginia.gov/vacode/title2.2/chapter20.1/

Section 2. Inventory of Data Analytics Assets, Data Assets, and Data Sharing

"[C]reate an inventory of state agencies' data analytics assets, capabilities, best practices, and data sharing activities…generate a common data sharing lexicon and terminology to eliminate friction and confusion among state agencies."

Governor Terence R. McAuliffe
Executive Directive 7 (2016)

The Executive Directive tasked the Secretary of Technology, the Secretary of Finance, and the CIO of the Commonwealth to compile an inventory of state agency data analytics assets, data assets, and data sharing activities, and to develop a lexicon of data sharing terminology to promote closer collaboration for data sharing by state agencies. The following section documents the findings from VITA's analysis of the inventory and related data submitted by state agency representatives in the survey instrument and stakeholder focus groups. The section has been organized to first present key findings on agency data assets and data sharing activities, followed by a summary of the data analytics assets being used by state agencies, levels of analytics expertise reported by state agencies, and references to source documents supporting a data sharing lexicon.

2.1 Inventory of Data Assets and Data Sharing Activities

VITA pulled from two primary sources of information to compile the inventory of agency data assets and data sharing activities. First, VITA reviewed details relating to agency data assets collected by the Commonwealth Security and Risk Management Directorate under Governor McAuliffe's Executive Directive 6: *Expanding Cyber-Related Risk Management Activities* (2015).[2] The following goals established in Executive Directive 6 aligned with the data collection strategy required under Executive Directive 7:

- Identify data assets used by state agencies
- Identify sensitivity and integrity of the data
- Identify classifications of the data (e.g., protected health information, etc.)
- Prioritize risk of each system based on their data assets
- Identify risk-based approach to protect systems and data assets

VITA observed 2,074 data assets reported by state agencies under Executive Directive 6, but removed 335 data assets from the analysis based on the data owner's classification of the asset as "retired," "not a dataset," or "duplicate." VITA also excluded 53 data assets since no information on data sharing had been provided. The analysis resulted in the determination of 1,686 "significant" data assets presently maintained by state agencies and for which details on data sharing had been submitted by the data owner.

The second source of information consisted of data collected from state agencies using a structured survey instrument. The survey requested agency data owners to update the information submitted for Executive Directive 6 and provide additional details on data sharing activities. VITA, with a 99% response rate, found that, of the 1,686 significant data assets logged in the inventory, 77.8% (1,311) of the assets did not support data sharing activities; 22.2% (375) did support data sharing activities.

The 375 agency data assets that supported data sharing tended to be distributed in five (5) high activity areas across the state government. A high activity area was defined as a Secretariat whose data sharing makes up roughly 10 percent or more of the state government total. These five (5) high activity areas accounted for 71.7% of the state's entire inventory of shared data assets.

[2] Executive Directive 6: *Expanding Cyber-Related Risk Management Activities* may be accessed at https://governor.virginia.gov/media/4398/executive-directive-6ada.pdf

VITA observed the highest concentration under the Secretary of Health and Human Resources, with 21.3% of the state total (80 data assets), followed by Transportation with 15.7% (59), Natural Resources with 15.2% (59), Public Safety and Homeland Security with 9.9% (37), and Finance with 9.6% (36). **Table 1** shows the breakout of high activity areas by Secretariat, ranked by the Secretariat's percentage of the state total.

Table 1. High Activity Data Sharing Areas by Secretariat

Secretariat	Total Data Assets	Data Assets Shared	% Data Assets Shared	% State Govt. Total
Health & Human Resources	517	80	15.5%	21.3%
Transportation	230	59	25.7%	15.7%
Natural Resources	114	57	50.0%	15.2%
Public Safety & Homeland Security	142	37	26.1%	9.9%
Finance	316	36	11.4%	9.6%
Agriculture & Forestry	62	33	53.2%	8.8%
Administration	63	30	47.6%	8.0%
Education	112	27	24.1%	7.2%
Commerce & Trade	74	9	12.2%	2.4%
Technology	50	7	14.0%	1.9%
Veterans & Defense Affairs	6	--	0.0%	0.0%
Total	**1,686**	**375**	**--**	**100.0%**

The inventory of data assets and data sharing activities revealed that a large majority of data collected, maintained, and used by state agencies remains within the host agency, not shared with other agencies. This is particularly true of the two Secretariats with the largest inventory of data assets – Health and Human Resources (517) and Finance (316) – which shared only 15.5% and 11.4% of their data assets, respectively. However, the inventory did highlight several "hotspots" for data sharing. Insights from executive leaders, business leads, and data stewards from agencies in these high activity areas may help to inform future data sharing opportunities.

2.2 Key Factors for Data Sharing

To better understand the business-related factors of agency data sharing, VITA presented a series of questions to focus group participants targeting primary drivers and requirements underlying existing data sharing relationships. First, VITA found that besides the sharing of data by state agencies with their federal, state, and local partners to support regular program administration and reporting, most of the data sharing activities could be tied directly to business-driven use cases. Examples of these relationships include the Virginia Longitudinal Data System, hosted by the Department of Education, and exchanges of geospatial information by the Virginia Geographic Information Network.

Second, VITA observed during the focus groups that, even when agencies had established a compelling business case for sharing data, they had to meet often rigorous prerequisites before onboarding to, or establishing, the data sharing relationship.
Some of the core prerequisites centered on an agency's capacity for the following:

- Technical requirements and specifications for required data exchange interfaces
- Multiple tiers of policies and standards for security, privacy, and governance
- Data exchange standards required by the exchange to promote interoperability
- Security and risk management protocols to prevent sharing of sensitive data

The survey and focus group results showed most agencies had a definite demand for data sharing. However, meeting this demand will require understanding the legal constraints to data sharing and having a mechanism to support discovery of high-value data assets available at other agencies. The Commonwealth is in the process of procuring new messaging services, developing a migration plan, and structuring appropriate governance to support Gmail and the complimentary suite of products including Google Docs for document sharing. This represents an important future-state opportunity which should be pursued if the appropriate security controls and governance can be implemented to protect commonwealth data and control costs.

2.3 Best Practices for Data Sharing

VITA identified best practices for data sharing in the survey and focus group responses. The best practices originated primarily from (1) federal/state policies, standards, guidelines, or program rules; (2) guidance from professional associations; and (3) lessons learned by the agency or members of the data sharing partnership. **Table 2** ranks best practices observed by VITA, based on the frequency of the theme being expressed in agency responses.

Table 2. Best Practices for Data Sharing

Rank	Category	Best Practices
1	Information Security & Compliance	• Engage legal counsel to review and document compliance requirements prior to formalizing data sharing relationships • Coordinate with Commonwealth Security and Risk Management on applicable data sharing requirements • Enforce compliance requirements through regular audits on source and downstream data systems
2	Physical & Logical Access Controls	• Establish compliant requirements for physical and logical access controls as part of data sharing agreements • Incorporate access control logs and metrics in audit protocols for participant (source and downstream) systems • Provide regular training and technical support to data sharing participants on physical and logical access controls
3	Agreements & Relationships	• Develop data sharing relationships based on business-driven use cases and agreed-upon purpose statements • Adopt standardized templates and/or a trust framework governance model for data sharing agreements • Implement restricted use agreements to control downstream and future use of shared data
4	Governance & Metadata Documentation	• Build trust-based governance models with clearly stated business, legal, and technical requirements • Enable discovery and interoperability of shared data through published metadata for each data asset • Document data definitions and specifications for data elements to be included in the data sharing relationship
5	Information Systems, Authentication, & Interoperability	• Document performance and service specifications for the information systems involved in data sharing • Require compliant electronic authentication and identity management protocols for data access [§ 59.1-550 et seq.] • Design data sharing interfaces to conform with external data exchange standards to promote interoperability

2.4 Inventory of Data Analytics Assets

VITA collected information using a structured survey instrument to compile the inventory of data analytics assets currently being used by state agencies. The survey results showed agencies favored 12 analytics toolsets, with a total frequency of 119 reported instances of the tools. **Table 3** shows the analytics tools and frequency of use by agencies.

Table 3. Data Analytics Tools Used Most Frequently by Agencies

Data Analytics Tools	Number of Instances	% State Govt. Total
Microsoft Excel	26	21.8%
Microsoft Access	12	10.1%
IBM (SPSS, Cognos)	12	10.1%
Esri ArcGIS	11	9.2%
LogiXML/Logi Analytics	10	8.4%
Microsoft PowerBI/SQL Server Analysis Services	9	7.6%
Microsoft SQL Server	9	7.6%
SAS	9	7.6%
Oracle (OBIEE, Exalytics)	6	5.0%
Crystal Reports	5	4.2%
Tableau	5	4.2%
Google Analytics	5	4.2%
Total	**119**	**100.0%**

A majority of state agency respondents reported using Microsoft Excel, with a frequency of use at 26 instances. This confirms that all agencies have, at a minimum, at least one data analytics asset at their disposal, given that Excel comes standard in the Microsoft Office suite of applications. Almost all of the Excel users said the application remained a "strategic" asset for their analytics capabilities, and more than half of the Excel users rated themselves at the "mastery" or "advanced" level of expertise. Other high-ranking analytics toolsets currently in place included Microsoft Access and IBM SPSS or Cognos at 12 instances, Esri's ArcGIS geographic information system (GIS) at 11 instances, and LogiXML/Logi Analytics at 10 instances.

Secretariats using the most analytics tools included Public Safety and Homeland Security with 28 unique tools, Health and Human Resources with 24, Education with 18, and Natural Resources and Transportation each with 13. The primary uses of the analytics tools were for statistical analysis, data visualization, and business intelligence. **Table 4** breaks out the number of analytics tools used at the Secretariat level.

Table 4. Data Analytics Tools by Secretariat

Secretariat	Number of Unique Analytic Tools	Top 4 "Strategic" Analytics Tools
Public Safety & Homeland Security	28	MS Excel, MS Access, IBM SPSS, ArcGIS
Health & Human Resources	24	MS Excel, IBM SPSS, MS PowerBI, SAS
Education	18	MSPowerBI, LogiXML, SAS, Google
Natural Resources	13	MS Excel, MS Access, ArcGIS, SQL Server
Transportation	13	MS Excel, ArcGIS, DQTools, Splunk
Agriculture & Forestry	9	MS Excel, SAS, Business Objects, Informatica
Commerce & Trade	9	MS Excel, MS Access, ArcGIS, AutoCAD
Finance	9	MS Excel, ArcGIS, MS SSRS, SigmaPlot
Administration	6	ArcGIS, MS Excel, MS Access, SQL Server
Technology	6	LogiXML, R, Google Analytics, AiMIQ
Veterans & Defense Affairs	5	MS Excel, Google, Facebook, Twitter Analytics

2.5 Best Practices and Skill Assessment for Data Analytics

VITA reviewed survey and focus group results to identify best practices in the area of data analytics. Most of the analytics-related best practices centered on the importance of taking a standards-based approach to data analysis, data quality, and data modeling. Other best practices revealed the importance of aligning data analytics program requirements with the underlying governance model for data sharing. **Table 5** ranks the data analytics best practices, as reported by agency respondents.

Table 5. Best Practices for Data Analytics

Rank	Category	Best Practices
1	Data Analysis	• Follow data analytics methods, standards, and established techniques to ensure validity and reliability in the analytic results
2	Data Quality	• Inspect data quality, integrity, values and constraints to ensure accuracy, precision, validity, and reliability in the data analytic results
3	Data Modeling	• Implement established methodologies for building, applying, and diagnosing statistical, predictive, and spatial analytic models
4	Reporting & Business Intelligence	• Adopt methodologies, formats, and data visualization techniques to meet reporting, business intelligence, and outcome measurement objectives
5	Analytics Program Governance	• Align data analytics requirements with data governance models underlying the analytics program and associated data sharing relationships

VITA then assessed agency reported levels of analytics expertise, focusing on current state and priority sets for a desired state. VITA found current levels of expertise generally ranged across survey response categories: Limited, Basic, Intermediate, Advanced, Master.

As for priority areas to enhance analytics capacity, the highest prioritization hierarchy: (1) business intelligence, (2) statistical analysis, and (3) data visualization, with 34 agencies favoring this category. The next favored hierarchy, as reported by agencies: (1) business intelligence, (2) data visualization, and (3) predictive analytics. However, VITA also observed gaps in the current and desired level of expertise. The gaps tended to be in the areas of (1) business intelligence, (2) data visualization, and (3) predictive analytics.

2.6 Data Sharing Terminology

The Executive Directive called for a "common data sharing lexicon and terminology to eliminate friction and confusion among state agencies." The following terms have been identified from adopted sources in the state government's information technology policies, standards, and guidelines, specifically the Commonwealth EIA Strategy (COV EIA Strategy) and the Information Technology Resource Management Glossary (ITRM Glossary):

Data Assets – An enterprise's data and information resources viewed as having a measurable value and used to achieve business objectives. [COV EIA Strategy]

Data Breach – The unauthorized access and acquisition of unredacted computerized data that compromises the security or confidentiality of personal information. Good faith acquisition of personal information by an employee or agent of an individual or entity for the purposes of the individual or entity that is authorized to view the data is not a breach of the security of the system, provided that the personal information is not used for a purpose other than a lawful purpose of the individual or entity or subject to further unauthorized disclosure. [ITRM Glossary]

Data Classification – A process of categorizing data according to its sensitivity (see definition for Sensitivity). [ITRM Glossary]

Data Custodian – An individual or organization in physical or logical possession of data for Data Owners. Data Custodians are responsible for protecting the data in their possession from unauthorized access, alteration, destruction, or usage and for providing and administering general controls, such as back-up and recovery systems. [ITRM Glossary]

Data Dictionary – A centralized repository of information about data such as meaning, relationships to other data, origin, usage and format. A data dictionary includes such items as complete and accurate definitions of entities and attributes, attribute domains, valid values, synonyms or aliases, default values, data type and length, required/not required constraints and other information. [ITRM Glossary]

Data Management – Development and execution of architectures, policies, practices, and procedures for managing the full data lifecycle: define, obtain/create, store/maintain, use, share, archive, and destroy. [COV EIA Strategy]

Data Owner – An individual, who defines, manages and controls the use of data and ensures compliance with adopted standards within an agency. The Agency Head or designee designates the Agency Data Owner(s) for the functional/subject areas within their jurisdictional control or authority and ensures adequate resources for Agency Data Owner(s) to develop and maintain their respective functional subject areas in support of the Commonwealth's Data Management Program. [ITRM Glossary]

Data Standards – Mutually accepted agreements governing the data elements, representations, formats, and definitions of common or shared data. [COV EIA Strategy]

Data Steward – An individual assigned by an agency to represent the agency's interagency data needs and ensure that proposed standards meets those needs. Agency Data Steward(s) work on behalf of their Agency Data Owner(s) and should have a broad understanding of the agency's data, be able to research data usage, be empowered to obtain agreement from Data Owner(s) and have the requisite authority to address data issues for the agency. [ITRM Glossary]

External (Data) Standard – A standard defined and maintained by a Standards Development Organization to improve the ability to share electronic data and ensure semantic interoperability. Generally may apply to services, documents, vocabularies (i.e., reference terminologies), and/or messages. Includes extending (e.g., adding data elements or codes to) an existing external standard to accommodate requirements specific to the Commonwealth. [ITRM Glossary]

Interoperability – Ability of diverse information systems to share or exchange data regardless of differences in applications or system platforms. [COV EIA Strategy]

Metadata – A set of data that describes and gives information about an agency's data assets. [COV EIA Strategy]

Sensitive Data – Any data of which the compromise with respect to confidentiality, integrity, and/or availability could adversely affect Commonwealth of Virginia interests, the conduct of Agency programs, or the privacy to which individuals are entitled. [ITRM Glossary]

Sensitivity – A measurement of adverse effect on COV interests, the conduct of agency programs, and/or the privacy to which individuals are entitled that compromise information systems and data with respect to confidentiality, integrity, and/or availability. Information systems and data are sensitive in direct proportion to the materiality of the adverse effect caused by their compromise. [ITRM Glossary]

Trust Framework – A formal agreement and supporting policies and procedures executed among agencies or other organizational entities that enforces the requirements, specifications, and permitted purposes for the participants to exchange, view, access, or otherwise share data. [COV EIA Strategy]

In addition to these terms from the state government's adopted sources, the National Institute of Standards and Technology (NIST) Interagency/Internal Report 7298, Release 2, Glossary of Key Information Security Terms, offers standards-based definitions and reference terminology relating to information security, data sharing, and data governance. Links to the Commonwealth and NIST documents have been provided in **Appendix 4**.

Section 3. Review of Data Sharing Concerns

> *"A comprehensive review of all legal, privacy, and governance concerns as they relate to data sharing"*
>
> Governor Terence R. McAuliffe
> Executive Directive 7 (2016)

The Executive Directive called for the Secretary of Technology, the Secretary of Finance, and the CIO of the Commonwealth to conduct a comprehensive review of the legal, privacy, and governance concerns relating to data sharing. The following section presents findings from the legal review compiled by VITA's Legal and Legislative Services Directorate, with assistance from the Office of the Attorney General (OAG). The section also highlights the primary concerns and formal constraints relating to privacy, security, and governance articulated by state agency representatives in their responses to the VITA survey instrument and discussed in the stakeholder focus groups.

3.1 Legal Review

The legal review conducted pursuant to the Executive Directive assumed the term "data sharing" to mean a state agency providing data it owns, or has custody of, to other state agencies. Agency activities to make data available to the general public, or "open data," has been discussed below in the Recommendations section of this report.

The legal review also assumed the only restrictions on agency data sharing were those codified in the data protection and privacy laws discussed below in this section of the report. However, it must be noted additional restrictions may exist based on how, and from whom, the state agency acquired the data (*e.g.*, a state agency may have a contractual obligation not to share data received from a third-party data supplier).

The principal finding from the legal review was the Government Data Collection and Dissemination Practices Act (GDCDPA) and other Virginia statutes, discussed in more detail below in this document, place certain restrictions upon agencies that affect their ability to share data. In fact, these statutes significantly limit the usefulness of shared data in some contexts (e.g., the GDCDPA's requirement for a state agency to only use personal information for the purpose for which it was collected).

Neither the United States nor the Commonwealth has a comprehensive data sharing law. Rather, the evolution of data protection and privacy laws has led to an extensive array of narrowly tailored, state and federal laws, regulations, rules and policies ("formal constraints") that govern the collection, maintenance, use, and dissemination of data. The vast majority of these formal constraints were designed to address a particular issue and relate to a specific industry or subject matter.[3]

While the Executive Directive called for a comprehensive review of all legal and privacy concerns related to data sharing, these state and federal laws, regulations, and policies are too diverse to address in this report. Accordingly, the report should be viewed as high level guidance, and agencies should engage their counsel at the OAG for specific legal advice. It may often be the case that compliance with federal law does not mean compliance with applicable Virginia law, and vice-versa.

[3] *See, e.g.*, I.R.C. § 6103 (federal tax information); IRS Publication 1075 (federal tax information); Va. Code § 58.1-3 (state tax information); Va. Code § 63.2-102 (public assistance programs and child support enforcement information); 20 U.S.C. § 1232g (student education records); 45 CFR §§ 160.101-.552, 164.102-.106, 164.500-.534 (protected health information); Va. Code § 32.1-271 (vital records); Va. Code §60.2-623 (unemployment benefits information); Va. Code § 19.2-389 (criminal history record information); and Va. Code § 55-210.24:2 (information furnished to the Division of Unclaimed Property).

In fact, the *Code of Virginia* in some instances prescribes specific penalties when data sharing does not comply with statutory restrictions. For example, § 58.1-3, which governs secrecy of tax information, specifies improper sharing of tax information shall be considered a Class 1 misdemeanor. Likewise, violations of § 63.2-104, which governs confidential information concerning social services, also constitute a Class 1 misdemeanor.

Although a comprehensive review of all formal constraints at the state and federal level is not feasible, this report summarizes several statutes of general applicability that together provide the *Code of Virginia*'s legal framework for sharing of most of the Commonwealth's information not governed by data, or partner-specific, laws. These include the GDCDPA, the Protection of Social Security Numbers Act, and the Virginia Freedom of Information Act.

The Government Data Collection and Dissemination Practices Act (GDCDPA)

After extensive study by the Virginia Advisory Legislative Council (VALC), the General Assembly adopted the Privacy Protection Act of 1976. Subsequently, in 2001 the General Assembly replaced the Privacy Protection Act with the GDCDPA. The primary purpose of the GDCDPA is to "ensure safeguards for personal privacy" and "preserve the rights guaranteed a citizen in a free society" by "establish[ing] procedures to govern information systems containing records on individuals."

Review of the provisions of the GDCDPA must begin with its definition of "personal information." Section 2.2-3801 states "'[p]ersonal information' means all information that (i) describes, locates, or indexes anything about an individual … or (ii) affords a basis for inferring personal characteristics."

Some examples of personal information include, but are not limited to, name, personal identification number (such as passport number, social security number, driver's license number, student identification number), medical history, financial transactions, and biometric data. Recognize that information that does not identify an individual when considered in isolation may nonetheless identify an individual when combined with other information.[4]

This statute thus recognizes extensive collection, maintenance, use, and dissemination of personal information directly affects an individual's privacy. The GDCDPA establishes certain principles of information practice guiding government agencies in the collection, maintenance, and use of personal information. Va. Code § 2.2-3800. The Supreme Court of Virginia described these principles as follows:

> [N]o secret personal information system shall be established; the need to collect the information must be clearly established in advance; information must be relevant to the purpose for which it has been collected; it should not be used unless accurate; the individual should be able to learn the purpose for which it is collected and particulars about its use and dissemination; the individual should be permitted to correct or erase inaccurate or obsolete information; and any agency maintaining such data should assure its reliability and prevent its misuse.
> *Hinderliter v. Humphries*, 224 Va. 439, 443 (1982).

[4] *Id.* Guide to Protecting the Confidentiality of Personally Identifiable Information (PII), NIST SP 800-122, at 2-1 ("[A] list containing only credit scores without any additional information concerning the individuals to whom they relate does not provide sufficient information to distinguish a specific individual. If the list of credit scores were to be supplemented with information, such as age, address, and gender, it is probable that this additional information would render the individuals identifiable.").

In accordance with these principles, the GDCDPA provides that a state agency shall "[c]ollect, maintain, use, and disseminate only that personal information permitted or required by law to be so collected, maintained, used, or disseminated, or necessary to accomplish a proper purpose of the agency." Va. Code § 2.2-3803(A)(1).

Despite this apparent limitation on the collection and dissemination of personal information, the Supreme Court of Virginia, in analyzing the predecessor to the GDCDPA, stated that "the [Privacy] Act does not render personal information confidential. Indeed, the [Privacy] Act does not generally prohibit the dissemination of information. Instead, the enactment requires certain procedural steps … to be taken in the collection, use, and dissemination of such data." *Hinderliter*, 224 Va. At 447.

While section 2.2-3803(A)(1)'s apparent limitation on the collection and dissemination of personal information does not hinder data sharing between state agencies, the GDCDPA's other principles of information practice and requirements for state agencies maintaining personal information systems limit interagency sharing of personal information. Several such principles and requirements are discussed below:

- Section 2.2-3800(C)(9) provides that "[t]here shall be a clearly prescribed procedure to prevent personal information collected for one purpose from being used for another purpose." Indirectly, this principle mandates that a state agency may only use personal information for the purpose for which it was collected. This principle affects both the transferring and receiving state agencies. A transferring state agency sharing data with knowledge that the receiving state agency plans to use the shared data for some purpose unrelated to the purpose for which it was collected may violate this principle. Furthermore, a receiving state agency may only use the shared data for the same purpose for which the transferring state agency collected it. For example, where a transferring state agency shares personal information collected to determine a citizen's eligibility for a benefits program, the receiving state agency cannot use the shared data to evaluate the same citizen's qualifications for a professional license. In effect, this principle limits both data sharing and the use of shared data.

- Section 2.2-3800(C)(2) provides that "[i]nformation shall not be collected unless the need for it has been clearly established in advance." This principle does not necessarily limit data sharing, but may affect how and when data can be shared. For example, a receiving state agency could not proactively request and collect all of a transferring state agency's data for future use. Rather, the receiving state agency must wait until a particular problem arises where the transferring state agency's data would be useful, and then request the data.[5]

- Section 2.2-3800(C)(5) provides that "[i]nformation shall not be used unless it is accurate and current." While not directly limiting data sharing, this principle could limit a receiving agency's ability to use shared data after it may no longer be current, at least without taking further steps to verify that the information is accurate and current.

- Section 2.2-3800(C)(8) provides that "[a]ny agency holding personal information shall assure its reliability and take precautions to prevent its misuse." This principle does not directly limit data sharing, but imposes additional responsibility on an agency receiving shared data. While section 2.2-3800(C)(5) limits a receiving agency's ability to use shared data, this principle affirmatively obligates the receiving agency to make certain that the data is dependable, or likely to be true or correct.

[5] *See, e.g.*, 2013 Op. Att'y Gen. Va. 7 (finding that the Virginia State Police's collection of data from an automated license plate reader "for potential future use if a need … arises respecting criminal or terroristic activities [(*i.e.*, "passive" data collection)] … does not comport with the [GDCDPA]'s strictures and prohibitions, and may not lawfully be done").

- Section 2.2-3803(A)(2) provides that "[an] agency maintaining an information system that includes person information shall collect information to the greatest extent feasible from the data subject directly." In essence, state agencies must, if at all possible, collect data directly from the individual rather than from some other source that already has the data. What makes such collection from an individual infeasible is likely a fact-sensitive question that varies depending upon the data being shared and perhaps evolving standards of feasibility, and it is unclear whether a state agency must make this determination on an individual-by-individual basis or if the General Assembly intended to permit a group determination of feasibility.

- Section 2.2-3803(A)(5) provides that "[a]n agency maintaining an information system that includes personal information shall make no dissemination to another system without (i) specifying requirements for security and usage including limitations on access thereto, and (ii) receiving reasonable assurances that those requirements and limitations will be observed." In other words, a state agency may only share data after establishing standards to govern the receiving state agency's use of the data and receiving a commitment of compliance therewith.

- Section 2.2-3806(A)(2) provides that "[a]n agency maintaining personal information shall give notice to a data subject of the possible dissemination of part or all of this information to another agency, nongovernmental organization or system not having regular access authority, and indicate the use for which it is intended, and the specific consequences for the individual, which are known to the agency, of providing or not providing the information." In other words, the data subject has the right to know for what purpose and with whom their personal information is being shared.

- Sections 2.2-3803(A)(6)-(9) impose requirements on a state agency that maintains an information system that includes personal information, these requirements include maintaining a list of all persons with regular access to the system and a complete and accurate record of every access to personal information for up to three years. While these requirements do not directly limit data sharing, the receiving state agency may incur additional costs and responsibilities based on their receipt of shared data.

Although the GDCDPA does not categorically prohibit the collection and dissemination of personal information, its principles of information practice and various obligations imposed on state agencies that maintain information systems containing personal information may hinder data sharing, particularly a state agency's ability to receive and use shared data.

Furthermore, a receiving state agency may acquire additional administrative duties and costs based on its receipt of shared data and failure to comply with these requirements exposes the state agencies involved to injunctive or monetary relief and exposes the officer or employee to civil penalties. Finally, although the GDCDPA's principles of information practice and requirements generally apply to all state agencies maintaining an information system that includes personal information, a particular information system or agency may be exempt from the GDCDPA based on a statutorily provided exemption or the Appropriations Act.[6]

[6] *See e.g.,* Va. Code § 2.2-3802(1) (providing an exemption from the GDCDPA for personal information systems maintained by any court of the Commonwealth); 2016 Va. Acts. Ch. 780, Item 4-9.01(b)(2)(a) ("Notwithstanding § 2.2-3800 of the Code of Virginia, the Virginia Department of Education, State Council of Higher Education for Virginia, Virginia Community College System, and the Virginia Employment Commission may collect, use, share, and maintain de-identified student data to improve student and program performance including those for career readiness.").

The Protection of Social Security Numbers Act (SSN Act)

While essentially every definition of personal information would include an individual's social security number, the General Assembly enacted the SSN Act in 2009 to specifically address access and disclosure of records containing social security numbers. The SSN Act provides that "the first five digits of a social security number contained in a public record shall be confidential and exempt from disclosure under the Freedom of Information Act (§ 2.2-3700 et seq.)."

Va. Code § 2.2-3815(A). Section 2.2-3815(B) of the SSN Act, however, contains a number of exceptions that permit "the release of a social security number." For example, section 2.2-3815(B)(3) allows an agency to share data that contains a social security number with another agency, either in Virginia or another state, when such data is requested "in connection with … the performance of such agency's official duties," and section 2.2-3815(B)(6) allows an agency to release a social security number "to a person or entity when necessary to administer any program of the agency [or] to perform a service or function of the agency." Thus, there may be instances where state agencies can share public records despite the fact that they contain social security numbers.

The Virginia Freedom of Information Act (FOIA)[7]

To provide openness in government activities, FOIA directs that all nonexempt official records be open to inspection and copying within five working days of the request, except as otherwise specifically provided by law. FOIA itself does not prohibit disclosure of exempt records or make exempt records confidential. Instead, FOIA provides that the custodian has the discretion to release exempt records, except where other law prohibits such disclosure. (Note: The term "custodian" used in this legal review has parallel meaning with "data owner" in the Commonwealth's ITRM policies, standards, and guidelines.)

Having said that, we must consider certain FOIA-related issues created by an environment where state agencies routinely share data, information, and records with other state agencies. Virginia law is unsettled regarding instances where a state agency shares data with another state agency, and whether the receiving state agency becomes a custodian of the data for purposes of responding to FOIA requests. Instead, the only settled area involves Section 2.2-3704(J), which is limited only to instances where an agency transfers records to another agency (such as VITA) for the limited purpose of storage, maintenance, or archiving.

In light of FOIA's broad definition of public records, one could argue that the receiving state agency becomes a custodian of the shared data. Consequently, any effort to foster an environment where both the transferring and receiving state agencies use the records in the transaction of public business must account for the FOIA responsibilities and exemptions applicable to both agencies.[8] If we assume the receiving state agency becomes an additional custodian of shared data, a crafty requester may be able to "custodian shop" and avoid a FOIA exemption only applicable to a particular state agency (*i.e.*, forcing disclosure of otherwise exempt public records). A transferring state agency with an agency-specific exemption, rather than a data-specific exemption, should carefully consider the potential consequences of sharing such data with another state agency.

The interaction of FOIA and other statues, such as the GDCDPA, must also be considered. On one hand, FOIA provides that "[e]xcept as otherwise specifically provided by law, all public records shall be open to inspection and copying." Va. Code § 2.2-3704(A). At the same time, however, the GDCDPA requires certain procedural steps to be taken in the

[7] The Secretary of Finance recommends having a second legal review of data sharing restrictions faced by state agencies to be conducted by the Virginia Freedom of Information Act Advisory Council or the Joint Commission on Technology and Science.

[8] *See* Va. Code § 2.2-3701 (defining public records as "all writings and recordings … in the possession of a public body or its officers, employees or agents in the transaction of public business").

dissemination of personal information (e.g., notice to a data subject of possible dissemination). This also seems to be an unsettled area.

In general, in order to harmonize and give meaning to both statutes, agencies may have to comply with the requirements of both statutes when disseminating public records that contain personal information pursuant to a FOIA request. It is likely that a similar balancing must also occur when data sharing involving personal information occurs, or when personal information is made available to the public as "open" records.

Other *Code of Virginia* Provisions with Indirect Impacts on Data Sharing

Another statute of general applicability, the Virginia Public Records Act (VPRA)[9], may have indirect impacts on data sharing. The VPRA establishes uniform procedures for the management and preservation of public records. Va. Code § 42.1-76. State agencies may only destroy or discard a public record in accordance with an approved retention and disposition schedule and after the expiration of the record's retention period. Va. Code § 42.1-86(A).

For records created after July 1, 2006, the VPRA requires the destruction of records upon expiration of the retention period. Va. Code § 42.1-86.1(C). In an environment where state agencies share data, information, and records with other state agencies, it is unclear whether the transferring state agency must ensure the statutorily required destruction of shared data after the expiration of its retention period, or whether a receiving state agency may preserve the data for a longer period under its own schedules.

In conclusion, the GDCDPA and other statutes inject certain restrictions associated with data sharing. In fact, these statutes significantly limit the usefulness of shared data in some contexts (*e.g.*, the GDCDPA's requirement that a state agency only use personal information for the purpose for which it was collected).

Apart from these statutes of general applicability, an extensive array of laws, regulations, and policies have been established to govern the sharing of specific types of data. These laws, regulations, and policies vary greatly and their application depends, in part, on the type of data being shared, who is sharing the data, and with whom the data is being shared. In many cases these variations likely require careful legal analysis on a case-by-case basis to determine the parameters of permissible data sharing.

As a general rule, the ability of agencies to share specific kinds of data (e.g., data on persons) is restricted, unless statutory authority has been granted to share those data with specific recipients and for specific purposes. This general rule may be illustrated by three recent examples:

- The Governor's Restoration of Rights Policy represents an opportunity to enable future data sharing between the Secretary of the Commonwealth and Department of Corrections, which will support determination of eligibility for restoration of voting rights.

- Chapter 235 of the *2013 Acts of Assembly* (HB 2148) allows the Department of Corrections to share medical and mental health data with the Department for Aging and Rehabilitative Services and the Department of Social Services "for the purposes of reentry planning and post-incarceration placement and services."

- Chapter 118 of the *2015 Acts of Assembly* (SB 817) authorizes the Virginia Department of Health Professions to provide Prescription Monitoring Program records to local law enforcement officers for the purpose of investigation, supervision, or monitoring of a specific recipient.

[9] Va. Code §§ 42.1-76 through -91.

3.2 Review of Data Sharing Concerns and Formal Constraints

The VITA/OAG legal review presented above documents, at a high level, state statutes and case law governing data sharing by state agencies. However, the Executive Directive also required an analysis of informal concerns and other formal constraints that impact an agency's ability to share data in a compliant manner. VITA targeted measures focused on these topics in the survey instrument and focus group questionnaire. **Table 6** summarizes VITA's findings from these measures.

VITA noted the first set of concerns and formal constraints as dealing with the complex regulatory environment over all phases of the data lifecycle. Primary areas of concern included (1) multiple tiers of regulation at the federal, state, and local level governing security, privacy, confidentiality, and consent of person-centric and other sensitive data; (2) lack of clarity within the laws and regulations as to if, when, and for what purpose data may be shared in a compliant manner; and (3) risks, including civil and criminal penalties, for agency staff who violate the legal or regulatory requirements.

Table 6. Data Sharing Concerns and Constraints

Rank	Category	Primary Concerns & Constraints
1	Information Security	• Maintain security and privacy requirements for sensitive person data, while at rest and in motion • Implement search constraints on ad hoc queries to restrict access to data not appropriate for sharing or public release • Address concerns and formal constraints for de-identification, anonymization, and confidentiality
2	Government Laws & Regulations	• Comply with the Freedom of Information Act and other laws and regulations that affect data sharing • Navigate complex array of legal and regulatory requirements for protecting person data • Extend legal and regulatory safeguards and compliance authority to downstream recipients of data once shared
3	Data Ownership & Control	• Maintain chain of authority for data once shared beyond the source agency • Implement audit processes and protocols to ensure compliance by receiving agencies
4	Cost & Resources	• Build and maintain technical infrastructure to support data sharing requirements • Invest in storage capacity to meet increased requirements for data management and sharing • Recruit and retaining staff with requisite knowledge and skill-sets for data sharing and analytics
5	Transparency & Citizen Consent	• Ensure government transparency on the purpose of data collected, how it will be used, and whether it will be shared • Gain citizen consent for the collection, maintenance, use and sharing of data in a compliant manner
6	Data Quality	• Implement business and technical measures to ensure data accuracy, quality and integrity • Document and publishing descriptive information about data assets, including data limitations and constraints

Agency representatives shared multiple examples of these concerns, ranging from risks identified by the Department of Social Services associated with using Federal Tax Information (FTI) to determine eligibility for public programs, eligibility data that otherwise would not be classified as FTI, due to the fact that this could be classified as "comingling" and require the agency to comply with the rigorous Internal Revenue Service (IRS) Publication 1075 for information security; to the criminal penalties cited by the Virginia State Police that attach in the law enforcement domain to releasing certain data in a non-compliant manner to external entities.

The second set of concerns identified by VITA dealt with agency ownership requirements, pursuant to Commonwealth Information Security Standard 501 (SEC501), over the data and questionable chain of authority for data shared to external partners. The concerns centered on downstream misuse, misinterpretation, ethical use, and the general integrity of the data once it has left the source agency's span of control.

As the Department of Taxation noted in a 1991 study, *A Study of the Secrecy of Tax Information Provisions Under Title 58.1*, while authorized disclosures of data have benefits, "these benefits are not without costs. The costs to the Department of Taxation for providing this information and the costs to the recipient of the information, associated with maintaining the confidentiality of the information, are other important considerations."

More recently, agencies in the Public Safety and Homeland Security Secretariat indicated that de-identified or anonymized data – data in which all information enabling the identification of a person has been removed – on person entities could be matched with data elements in datasets published by an array of other sources, within and outside of the state government, allowing perpetrators to identify individual citizens.

In addition to these concerns, which occur when ownership of the data resides with a state agency, other issues arise when other levels of government have legal standing as data owners but share the data with state agencies to meet business objectives. Cadastral data (e.g., property boundaries) and administrative boundaries for localities, for example, fall under the jurisdiction of local governments, but the Virginia Geographic Information Network must work with localities to ensure consistency and interoperability in the geospatial datasets submitted for state maps.

Additional concerns observed by VITA centered on costs and the "siloed" nature of funding streams supporting agency data sharing activities. Cost issues focused on the infrastructure, technical architecture, and human resources needed to (1) develop and implement a data sharing interface (i.e., website, web service, electronic interface, etc.); (2) prepare the data for sharing in a manner that ensures accuracy, quality, and integrity; and (3) support the data sharing interface through its lifecycle. Funding-related constraints tended to stem from rules attached to the source program, and consequently the data systems built to support the program, which in turn limit agency use and sharing capacity of the program data.

Interestingly, VITA did not observe instances of "territorialism" among state agencies with regard to data sharing. A hypothesis going into the research held that agencies cultivated a feudal sense of data ownership, viewing data as part of the agency fiefdom, and therefore rejected any data sharing requests. VITA's findings failed to support this hypothesis. Agency representatives instead reported greater openness to sharing data, driven by increased business demand, and a more collaborative relationship between data owners, information security officers, and other stakeholders for making data available to other agencies. This suggests a shift occurring in the state government from the feudalistic, proprietary paradigm toward transparency, stewardship, and willingness to share.

3.3 Review of Governance Concerns for Data Sharing

The third area of review required by the Executive Directive focused on governance-related concerns associated with agency data sharing. VITA concentrated its analysis on lessons learned by those agency representatives who have been substantively involved in ongoing data sharing relationships. Focus groups and follow-up discussions with these representatives yielded valuable insights on the principal business drivers of data governance; strategies for building successful governance models within, and across, state agencies; and the critical role of legal counsel and comprehensive data sharing agreements to structure interagency partnerships. **Table 7** documents the key governance-related concerns observed in the analysis.

State agencies have come to recognize the business value of data governance to support their data sharing activities. Primary business drivers cited by agency representatives responding to the survey and in focus groups ranged from the need for solid governance models to enable analytics and business intelligence to specific requirements for data standardization, as a means of promoting interoperability for data sharing. The latter remained consistent across respondents, who cited problems in sharing information when agencies had different definitions and specifications for the data.

Agency representatives reported that successful governance begins with a strong governance model. Data sharing depends on participating agencies having in place the necessary agreements and governance mechanisms to enforce the business, legal, and technical requirements of the relationship. A frequently noted example of a strong governance model currently in place within the state government was the Virginia Longitudinal Data System (VLDS), hosted by the Department of Education. VLDS features participation by multiple state agencies and has a solid, trust-based governance model with oversight by a central coordinating committee.

Also, agency respondents expressed the importance of engaging legal counsel for guidance on data sharing agreements; legal review to ensure data sharing could be done in a lawful, compliant manner; and regular compliance reviews to mitigate liability and risk. Additional concerns and observations included the need for the governance models to address the ethical use of data by incorporating confidentiality, non-disclosure, and code-of-conduct agreements for staff to execute. Multiple agencies, including Virginia Department of Health, the Library of Virginia, and the State Council on Higher Education for Virginia, reported having such internal agreements in place for their staff.

Table 7. Governance Concerns for Data Sharing

Rank	Category	Primary Concerns & Constraints
1	Business Drivers	• Integrating data-driven decision making into business processes to improve agency performance • Building intelligence through analytics on data collected and maintained by state agencies • Promoting interoperability among state agencies through data exchange standards • Meeting demand for consistency in data definitions, quality, and security
2	Governance Models	• Developing a trust-based governance model enforced by comprehensive data sharing agreements • Establishing a governance body comprised of business and executive leaders to provide direction and oversite • Designing the governance model to address the business, legal, and technical requirements of data sharing • Sustaining relationships between participating agencies and within the governing body
3	Legal Counsel	• Engaging legal counsel to maintain the governance model and data sharing agreements • Conducting legal and compliance reviews to ensure data continues to be shared in a lawful, compliant manner • Handling FOIA requests and other inquiries relating to data sharing agreements and governance models • Developing standardized templates for data sharing agreements to mitigate liability and risk
4	Data Sharing Agreements	• Ensuring consistency in business, legal, and technical requirements across data sharing partners • Reducing risk associated with disparate point-to-point agreements by using standard templates • Engaging legal counsel to review agreements and modifications to agreements over time • Addressing requirements set by laws, regulations, and program rules in standard agreement
5	Data Exchange Standards & Interoperability	• Implementing data exchange standards to promote interoperability among data sharing partners and systems • Adopting external standards published and maintained by standards development organizations • Defining requirements for data exchange standards to ensure consistency in data definitions and specifications • Documenting descriptive metadata about the shared data, including level of sensitivity and applicable standards
6	Ethical Use of Data & Authority	• Requiring agency staff to sign internal agreements with defined penalties to ensure ethical use of data • Designing internal agency agreements to address confidentiality, non-disclosure, and ethical use of data • Conducting regular security audits to monitor compliance with employee agreements • Providing ongoing training to agency staff on security, privacy, confidentiality, and ethical use of data

Section 4. Recommendations

The Executive Directive called for the Secretary of Technology, the Secretary of Finance, and the CIO of the Commonwealth to submit to the Office of the Governor recommendations designed to enable state agencies to use shared data and analytics more systematically. The following section documents the recommendations and rationale for each, based on VITA's findings from its research under the Executive Directive.

Recommendation 1: Open Data, Data Accessibility, and Data Utilization by State Agencies

"Recommendations on how to make data generated by state agencies more accessible and usable by state government and the public as 'open' data"

Governor Terence R. McAuliffe
Executive Directive 7 (2016)

The Commonwealth's ITRM Glossary does not establish a definition for "open" data, and VITA's research revealed substantial disparity in meanings of the term across the state government. The Open Data Handbook defines open data as "data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share alike." Extending the term based on the Open Definition, published by Open Knowledge International, the Open Data Handbook states the following principles:[10]

- Availability and Access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.

- Re-use and Redistribution: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.

- Universal Participation: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavor or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

Recommendation 1.1. Dedicate OAG legal support to agencies to assist in determining whether data may be classified as "open" data

State agencies have an array of data assets that may be eligible for classification as open data. However, the formal constraints identified above raise the risk for agency staff to make such a determination. Having dedicated OAG resource to address open data questions for agencies will be necessary to help them understand whether the data could be released in a lawful, compliant manner. OAG support also will help guide agencies in establishing a formal governance process for making open data determinations.

Model Innovation: The State of Illinois leveraged legal guidance from agency general counsels to form an enterprise governance model involving seven (7) agencies. The multi-agency collaborative enables the sharing of data, information systems, and business processes to support 60 health and human service programs. Similarly, the State of Ohio established an enterprise governance model built on comprehensive data sharing agreements, facilitated by the Governor's Office of Health Transformation. Government Technology Article: http://www.govtech.com/health/Overcoming-Data-Governance-Challenges-in-HHS.html

[10] The Open Data Handbook's definition for "open" data may be accessed at http://opendatahandbook.org/guide/en/what-is-open-data/

Recommendation 1.2. Invest in the Virginia Open Data Portal to enhance accessibility, ease of use, and capacity

Virginia's Open Data Portal, which is hosted by the Library of Virginia with infrastructure support from VITA, provides a centralized platform for the state government's open data assets. However, investments in the portal will be required to improve the portal's accessibility, user experience, and infrastructure capacity. Dedicated program management for the portal also will be necessary to ensure conformance with industry standards for open data, metadata, and data exchange formats.

Virginia Open Data Portal: http://www.data.virginia.gov/

Model Innovation: The Obama Administration in March 2016 launched the "Opportunity Project," which places data and technology in the hands of civic leaders, community organizations, and families to build more equitable and thriving communities. Since its inception, the Opportunity Project has yielded dozens of new digital tools designed to help meet critical needs in communities, such as finding affordable housing near jobs and transportation, advocating for broader access to opportunity in neighborhoods, and making data-driven investments to increase economic mobility. Opportunity Project Site: http://opportunity.census.gov

Recommendation 1.3. Improve discovery and access to high value open datasets for state agencies and the public

Discovery of open data assets maintained by state agencies, those non-sensitive data assets which may be made openly available to other agencies or to the public, presents a significant problem for agency users. In most cases, agencies struggle to identify what data assets may be available and understand the process required to have those data assets published on the open data portal. Improving tools and methods for discovery would enable state agencies to identify high value data assets in most demand, enabling them to leverage open data to achieve targeted outcomes.

Recommendation 1.4. Invest in state level licensing for data analytics, business intelligence, and data anonymization applications

State agencies across domains of government have launched a variety of initiatives in data analytics, data visualization, and performance dashboards. However, access to advanced toolsets to support these initiatives creates a barrier to entry for agencies seeking to advance their analytics capacity. Providing a state level licensing agreement with vendors identified in the data analytics asset inventory will help to extend the availability of analytics and visualization toolsets. However, it should be noted that any toolset for data anonymization or de-identification should take into account the need to, and complexity of, fully removing personally identifiable information prior to making such data "open."

Recommendation 2: Data Sharing Governance, Ethical Use, and Authority

"Recommendations for data sharing governance, ethical use, and authority"
Governor Terence R. McAuliffe
Executive Directive 7 (2016)

Recommendation 2.1. Continue to support the state government's enterprise data governance program and explore the advantages in creating a senior enterprise data leader position

The Secretary of Technology in August 2013 adopted the Commonwealth EIA Strategy and directed the state's data governance program within VITA to oversee its implementation. The program will continue to sustain the state's effort under the Executive Directive to build an enterprise approach to data sharing and governance. Core activities for the program will be to define the process for, and lead the review of, proposed data analytics projects, maintain existing data steward groups, identify new data sharing use cases, and support agency level governance activities. This recommendation aligns with the Commonwealth's EIA Strategy Goal 1. Data Governance. There may be advantages to the Commonwealth in establishing a more senior enterprise data governance role (i.e. Chief Data Officer), reporting directly to the CIO of the Commonwealth. The commonwealth should explore the breadth and scope of responsibilities for this position, and pursue if attractive.

Model Innovation: In 2007, the State of North Carolina's General Assembly created the Government Data Analytics Center (GDAC) to provide enterprise-level support for data sharing and analytics activities across state agencies. Government Technology Article: http://www.govtech.com/data/North-Carolina-IT-Takes-Control-of-State-Data.html

Recommendation 2.2. Adopt a policy that defines the role of the data owner and establishes the obligations for data sharing and governance

Data owners, pursuant to SEC501, play a central role in data sharing and governance activities. However, having a clear, consistent definition and statement of the obligations, liability, and compliance requirements for data owners remains a critical concern across state agencies. The Office of the Secretary of Technology and the CIO of the Commonwealth will need to engage data stewards groups to establish an enterprise policy that addresses these concerns in a systematic, standards-based manner. Key provisions for policy statement include data ownership principles and authority, practices associated with open data, chain of custody for shared data, system-to-system security, and related governance issues. This recommendation aligns with Objective 1.2 of the Commonwealth EIA Strategy.

Recommendation 2.3. Perform ongoing Data Management Maturity (DMM) assessments for agencies across domains of the state government

State agencies have expressed a need for ongoing training and agency assessments on the level of maturity in their data management, data governance, and data sharing practices. The state government completed its first wave of DMM assessments in 2015, with support from researchers at Virginia Commonwealth University. Continuation of the DMM assessments will provide valuable information to state agencies to continue growth along their targeted maturity curve for data management. Continuing to advance the data management maturity of state agencies adheres to the principles, goals, and objectives of the adopted EIA Strategy.

Model Innovation: Many Commonwealth agencies have started conducting systematic assessments of their data management capabilities. One of the tools used for this purpose is the CMMI Institute's Data Management Maturity (DMM) Model Assessment Framework. The tool allows state agencies to analyze their current state, identify gaps between the current state and desired future state, and design strategies for closing the gaps.
CMMI Institute DMM Model Overview and Resources: http://cmmiinstitute.com/data-management-maturity

Recommendation 2.4. Publish the results from the state wide data asset inventory in a searchable repository to promote discovery and accessibility

The data asset inventory initiated under Executive Directive 6, and continued under Executive Directive 7, cited above in this report aligns with Goal 3 of the adopted Commonwealth EIA Strategy, which focuses on data asset management. Specifically, Objective 3.1 encourages the state government to "Complete an inventory of enterprise data assets across the Commonwealth and compile metadata on each enterprise asset." Publishing the results of the inventory compiled for the purpose of this report would accomplish Objective 3.1 and provide a valuable resource for state agencies for discovery of enterprise data assets.

Recommendation 3: Data and Analytics Projects to Promote the New Virginia Economy

"Recommendations of key projects providing the highest likelihood of realizing value of data and analytics in new ways that will demonstrate cost savings and support the New Virginia Economy"

Governor Terence R. McAuliffe
Executive Directive 7 (2016)

Recommendation 3.1. Establish a process to identify potential projects for business case development that align with the Governor's Policy Priorities

State agencies, through their relationship with other agencies, academic institutions, and research centers, often identify potential high-value data analytics projects that do not, at the time, have designated agency sponsorship or funding stream. The enterprise level data governance office recommended above as Recommendation 2.1 could play an important role in shepherding these types of conceptual process through the Commonwealth's IT Investment Management business case development and proposal process.

Recommendation 3.2. Require agencies to incorporate a "Data Plan" into their Information Technology Strategic Plans

Agency Information Technology Strategic Plans (ITSPs) align the agency's mission, goals, objectives, and strategies to targeted investments in information technology, systems, and services. Given the value of data and analytics toolsets as agency assets, each ITSP should include a Data Plan documenting how the agency will (1) maintain an inventory of its enterprise data assets; (2) develop agency wide data related goals; (3) identify gaps between current state and one year, three year, and five year data related goals; and (3) increase agency capacity for data analytics and associated capabilities to achieve strategic outcomes. The Data Plan also should identify specific data-centric projects proposed for business case development and review pursuant to the Commonwealth's Information Technology Investment Management (ITIM) Standard (CPM 516-01). Moving forward, the CIO of the Commonwealth may include data analytics projects in the annual Recommended Technology Investment Project (RTIP) Report to the Governor and General Assembly.

Recommendation 3.3. Projects recommended for future consideration pursuant to the Executive Directive

Projects recommended in this report for future consideration have been highlighted due to their potential for realizing value of data and analytics, generating potential cost savings, aligning with the Governor's Policy Priorities, and supporting the vision of a New Virginia Economy, as required by the Executive Directive. The projects were not chosen through a formal selection process, nor were they scored using objective criteria. The project list consists of those initiatives currently in flight that align with the strategic intent, goals, and objectives of the Executive Directive. VITA identified the projects through input from the Office of the Secretary of Technology, the VITA Executive Leadership Team, and agency representatives during the stakeholder focus groups. For the purpose of this report, the term "project" assumes initiatives involving, or proposed by, state agencies and institutions of higher education that may not have been recognized as formal IT projects or investments, as defined by the Commonwealth's ITIM Standard (CPM 516-01) and Project Management Standard (CPM 112-03.3). Formal selection of forthcoming data analytics projects, as well as all requirements for their implementation, will be led by the Commonwealth's enterprise data governance office and conducted pursuant to adopted Commonwealth ITRM policies, standards, and guidelines.

VITA investigated the potential of Cyber Security data analytics projects. However, the environment of the current infrastructure contract, and its pending replacement procurement, was not conducive to finding viable candidate projects.

Workforce

Workforce Development – Open Data/Open Jobs
Agency Sponsors: Council on Virginia's Future, Virginia Tech, Virginia Community College System
Value Proposition: Providing access to open data with the goal of increasing the match between job seekers and job opportunities in the New Virginia Economy
Project Description: This project creates an open data set of job postings in Virginia that can be used by the academic, public, and private sectors for research and analysis that will allow the Commonwealth to identify employer needs for talent, new and emerging skills for the education and training community, and better connect job seekers to labor market opportunities. This project was supported by Virginia Tech through the Governor's Data Internship Program, referenced above.

Educational Outcomes for High School Equivalency Graduates
Agency Sponsors: Department of Education, George Washington University
Value Proposition: Expanding employment opportunities for Virginians with GEDs
Project Description: The project applies analytics and data visualization on data from the Virginia Longitudinal Data System to answer key adult education questions regarding post GED outcomes. The project supports the Executive Order 23 workforce initiatives.

Education

Exceptional Student Analytics Program
Agency Sponsor: Richard Bland College
Value Proposition: Increasing student retention in post-secondary education
Project Description: The project will implement advanced analytics using a state of the art predictive modeling solution, giving students GPS-like guidance through their college journey and helping them remain on track in their academic programs.

Government & Citizens

Virginia Open Data Portal Expansion
Agency Sponsors: Office of the Secretary of Technology, Library of Virginia, VITA
Value Proposition: Maximizing access to open data maintained by the state government
Project Description: The project will increase sharing of public data among state agencies to create greater efficiencies, and making public data more accessible to citizens as a means of increasing civic engagement, transparency, and use of the Commonwealth's data assets.

Arterial Roadway Bottleneck Analysis
Agency Sponsor: Virginia Department of Transportation
Value Proposition: Improving government services and infrastructure
Project Description: The project involves development of a traffic "bottleneck" model to give VDOT a better understanding of the natural and man-made influences affecting traffic flow. The project has been selected under the Commonwealth's NextGen Data Analytics Program, referenced above.

Health & Family

State-Local Government Data Exchange for Human Services
Agency Sponsors: Department of Social Services, Department of Juvenile Justice, Local Governments
Value Proposition: Promoting efficiency and effectiveness in human service programs
Project Description: The project will enable evidence based decision making by state human service agencies and their local partners, with the goals of cost savings and program improvement.

Healthcare Outcome Analysis
Agency Sponsors: Department for Aging and Rehabilitative Services, University of Virginia
Value Proposition: Improving healthcare services and outcomes for Virginians
Project Description: The project employs data analytics and visualization assets to explore population health outcomes. The project has been supported by the Governor's Data Internship Program, referenced above.

Social Services Master Data Management-Based Program Analysis
Agency Sponsor: Department of Social Services
Value Proposition: Enhancing program outcomes while reducing fraud, waste, and abuse.
Project Description: The project will employ a master data management service to match records for those individuals receiving services under the Supplemental Nutrition Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF) program, Medicaid, Foster Care, and Child Protective Services. The improved matching capability will enable more effective monitoring of program recipients over time and across program areas, with the goal of significant cost savings. The project has been selected under the Commonwealth's NextGen Data Analytics Program, referenced above.

Model Innovation: The Michigan Enterprise Information Management (EIM) Team working with the Office of the Secretary for Health and Human Services, established guidelines, processes and supporting technology to develop a similar "Identity Master" system incorporating multiple data sources to focus on Medicaid fraud, waste and abuse.
http://www.michigan.gov/dtmb/0,5552,7-150-56345_56351-336646--,00.html

Public Safety & Homeland Security

Cross-Agency Data and Analytics to Reduce Opiate/Opioid Addiction and Related Health Risks
Agency Sponsor: Secretary of Health and Human Resources
Project Description: The project features a partnership between state institutions to assemble data from multiple agencies to explore a set of urgent health issues, notably the opioid crisis but also including other forms of substance abuse, mental health, violence, and the deaths they cause. The goal of this data project will be to generate information that policymakers need and can put to practical use through actionable policy on matters of urgency to the Commonwealth. This project can support the Governor's Taskforce on Prescription Drug and Heroin Abuse. https://www.dhp.virginia.gov/taskforce/

Analytics on Whole Genome Sequences of Microorganisms (Genomics)
Agency Sponsors: Department of General Services/Division of Consolidated Laboratory Services, University of Virginia
Value Proposition: Reducing risk associated with foodborne illnesses and bioterrorism
Project Description: The project features advanced genomic analysis on whole genome sequences of microorganisms. The project has been supported by the Governor's Data Internship Program, referenced above.

Section 6. Conclusion

VITA's analysis for the Executive Directive revealed relatively low levels of data sharing activity currently among state agencies due to legal restrictions and risk associated with data privacy and security, as well as an array of resource constraints. However, agencies expressed a sincere desire to share data as a means of reducing costs and improving government performance.

Other states have developed innovative models for sharing data. These models tend to feature three primary components: (1) an enterprise approach for enabling technologies, programs, and governance; (2) dedicated legal guidance for data sharing agreements, governance model, and legislative programs; and (3) business-driven use cases to inform the state government's data-related activities.

This report recommends for the Commonwealth of Virginia to develop an enterprise-wide approach for achieving the strategic goals and objectives set by the Executive Directive. These recommendations will challenge the state government to (1) design new models for data sharing, analytics, and governance; (2) make the necessary investments for enhancing data management capacity at the enterprise and agency level; and (3) promote continuous improvement and assessment for advancing long-term data management maturity.

VITA believes there is significant opportunity for the Commonwealth in realizing higher levels of data sharing and analytics. Further action should be taken to facilitate the environment to realize this opportunity.

Appendices

Appendix 1: Executive Directive 7

COMMONWEALTH of VIRGINIA

Executive Department

Executive Directive 7 (2016)

LEVERAGING THE USE OF SHARED DATA AND ANALYTICS

Importance of the Issue

In order to continue the Commonwealth's advancement towards a New Virginia Economy that draws on all of the Commonwealth's vast resources, it is important that state agencies have access to all information necessary to better provide services to our citizens. Increasing the use of shared data and analytics among Virginia agencies through a comprehensive and coordinated effort will improve the provision of services and outcomes, maximize the use of resources, and increase the return on investment of our citizens' tax dollars in their government;

Increasing data sharing, correlation, and analysis capacity will enable the state to achieve efficiencies in the administration of state programs and services, and allow state government to more efficiently and effectively address issues related to public health, public safety, education, and quality of life.

But just as important as improving the flow of information among government agencies is the respect that state agencies must show for individuals' privacy interests. State government shall continue to protect individual privacy, adhere to applicable state and federal regulations, and cyber-security best practices during any activity involving the collection of sensitive information.

Commonwealth data collection and analysis activities shall focus on enhancing government transparency, streamlining business processes, increasing operational efficiency and effectiveness, and minimizing duplication and overlap of current and future systems development.

Accountability

Now therefore, I, Terence R. McAuliffe, by virtue of the power vested in me, do hereby direct the Secretaries of Technology and Finance and the Commonwealth's Chief Information Officer (CIO) to review all Commonwealth systems, practices, processes, policies, applicable laws and regulations governing the sharing of data across agencies and create an inventory of state agencies' data analytics assets, capabilities, best practices, and data sharing activities. As part of this

effort, the Secretaries and the CIO shall generate a common data sharing lexicon and terminology to eliminate friction and confusion among state agencies.

The Secretaries and the CIO shall provide a report of their findings to me no later than October 15, 2016. The report shall specifically include the following:

- A comprehensive review of all legal, privacy, and governance concerns as they relate to data sharing
- Recommendations on how to make data generated by state agencies more accessible and usable by state government and the public as "open" data
- Recommendations for data sharing governance, ethical use, and authority
- Recommendations of key projects providing the highest likelihood of realizing value of data and analytics in new ways that will demonstrate cost savings and support the New Virginia Economy

All executive branch agencies shall cooperate with the Secretaries and CIO to complete the review process and provide any information requested.

Effective Date of the Executive Directive

This Executive Directive shall be effective upon its signing and shall remain in force and effect unless amended or rescinded by further executive order.

Given under my hand and under the Seal of the Commonwealth of Virginia, this 23rd day of May, 2016.

.

_____

Terence R. McAuliffe, Governor

Attest:

_____

Kelly Thomasson, Secretary of the Commonwealth

Appendix 2: Data Collection Methods and Report Review Process

Data Collection Methods

VITA implemented multiple data collection methods, in phases, as part of its research program to gather information for this report.

- Data Collection Survey Instruments: Two (2) structured survey instruments in spreadsheet format submitted to Executive Branch agencies to capture quantitative data regarding data analytics tools, data assets, and sharing activity (300+ stakeholders)

- Stakeholder Focus Groups: Twelve (12) cross-Secretariat sessions to gather qualitative information on significant concepts, approaches, issues, and opportunities associated with data sharing, governance, and analytics faced by state agencies (70+ stakeholders)

- Project Checkpoints: Presentations and feedback sessions with the Customer Advisory Council (CAC), Agency IT Resources (AITRs), Information Security Officers Advisory Group (ISOAG), Information Technology Advisory Council (ITAC), VITA Customer Account Managers (CAMs), and the VITA Executive Team.

- IMSAC & HITSAC: Facilitated information gathering sessions and project updates to the Commonwealth's Identity Management Standards Advisory Council (IMSAC) and the Health IT Standards Advisory Council (HITSAC).

Report Review Process

VITA engaged multiple stakeholder groups to complete a comprehensive review of previous draft versions of this report.

- VITA ED7 Core Team: Members of VITA's ED7 Core Team contributed to the drafting and initial review of the report. Representation on the team included subject matter experts and leadership from the following VITA Directorates: Relationship Management and Governance, Legal and Legislative Services, Commonwealth Security and Risk Management, and the Office of the CIO of the Commonwealth.

- Office of the Secretary of Technology: The Secretary of Technology and Deputy Secretary of Technology contributed to the drafting, organization, and review of the report. The Office of the Secretary, in conjunction with the CIO of the Commonwealth, also directed VITA staff on the approach, analysis, and strategic direction in the response to the Executive Directive.

- Office of the Secretary of Finance: The Secretary of Finance and Deputy Secretary of Finance conducted a policy level review of the report to ensure alignment with the state's finance, accounting, and budget priorities. The Office of the Secretary also assisted in the shaping of the proposed amendment to the Appropriation Act to enable more systematic data sharing by state agencies.

- Office of the Secretary of Health and Human Resources: The Office of the Secretary of Health and Human Resources was invited to review the report to evaluate alignment between the findings and recommendations under Executive Directive 7 and the HHR Secretary's report to the General Assembly required by Item 284 C of the 2016 Appropriation Act.

Agency Engagement

The following Executive Branch agencies, boards, and institutions of higher education submitted information to support VITA's analysis under the Executive Directive:

Alcoholic Beverage Control
Attorney General and Department of Law
Board of Accountancy
Center for Innovative Technology
Commonwealth Attorneys' Services Council
Compensation Board
Department for Aging and Rehabilitative Services
Department for the Deaf and Hard of Hearing
Department of Accounts
Department of Agriculture and Consumer Services
Department of Aviation
Department of Behavioral Health and Developmental Services
Department of Conservation and Recreation
Department of Corrections, Central Activities
Department of Criminal Justice Services
Department of Education - Central Office Operations
Department of Elections
Department of Emergency Management
Department of Environmental Quality
Department of Fire Programs
Department of Forensic Science
Department of Forestry
Department of Game and Inland Fisheries
Department of General Services
Department of Health
Department of Health Professions
Department of Historic Resources
Department of Housing and Community Development
Department of Human Resource Management
Department of Juvenile Justice
Department of Labor and Industry
Department of Medical Assistance Services
Department of Mines, Minerals and Energy
Department of Motor Vehicles
Department of Planning and Budget
Department of Professional and Occupational Regulation
Department of Rail and Public Transportation
Department of Small Business and Supplier Diversity
Department of Social Services
Department of State Police
Department of Taxation
Department of The Treasury
Department of Transportation
Department of Veterans Services
Frontier Culture Museum of Virginia
George Mason University
George Washington University
Gunston Hall
Jamestown-Yorktown Foundation
Library of Virginia
Marine Resources Commission

Motor Vehicle Dealer Board
Norfolk State University
Office of Children's Services (Children's Services Act)
Office of the Governor
Office of the State Inspector General
Richard Bland College
Science Museum of Virginia
State Corporation Commission
State Council of Higher Education for Virginia
University of Mary Washington
Virginia Board for People with Disabilities
Virginia Commission for the Arts
Virginia Commonwealth University
Virginia Community College System
Virginia Department for The Blind and Vision Impaired
Virginia Employment Commission
Virginia Information Technologies Agency
Virginia Museum of Fine Arts
Virginia Museum of Natural History
Virginia Outdoors Foundation
Virginia Racing Commission
Virginia Resources Authority
Virginia School for the Deaf and Blind
Virginia State University
Wilson Workforce and Rehabilitation Center

Appendix 3: Artifacts from Previous Data-Related Activities

- Secretarial Committee on Data Sharing: Committee formed in September 2011 by the Secretaries of Technology and Health and Human Resources to explore opportunities and constraints for an enterprise data-sharing agreement for state agencies, built on a trust framework governance model.
  http://www.vita.virginia.gov/oversight/default.aspx?id=6442470188

- Commonwealth Enterprise Information Architecture (EIA) Strategy: The Secretary of Technology in August 2013 adopted an enterprise data strategy, developed with input from agency leaders, business managers and technical leads. Strategic goal areas: Data governance, data asset management, data standards, and data sharing.
  http://www.vita.virginia.gov/uploadedFiles/VITA_Main_Public/Oversight/EA/Data_Management_Group/Commonwealth_EIA_Strategy_FINAL.pdf

- ITRM Data Exchange Standards: The Secretary of Technology, to date, has adopted more than 130 data exchange standards to promote interoperability and the sharing of data in a compliant, standardized manner. Standards cover health information technology and the National Information Exchange Model (NIEM) for citizen-centric data.
  http://www.vita.virginia.gov/oversight/dm/default.aspx?id=12422

  NIEM Adoption Strategy:
  http://www.vita.virginia.gov/oversight/DM/default.aspx?id=6442473684

  Adopted Health IT Standards:
  http://www.vita.virginia.gov/uploadedFiles/VITA_Main_Public/ITAC/HITSAC/COV_Health_IT_Standards.pdf

- Data Stewards Groups: In February 2014, the Commonwealth inaugurated three data steward groups – Executive, Functional (Business), and Technical Data Stewards – to support ongoing agency engagement and direction for implementation of the EIA Strategy and related data governance activities.
  http://www.vita.virginia.gov/oversight/DM/default.aspx?id=6442472432

- Governor's Data Internship Program (GDIP): The Office of the Governor and the Secretary of Technology in the fall semester of 2014 implemented the internship program to pair interns from state universities with state agencies to perform advanced analytics on "real-world" problems.
  http://www.govtech.com/data/Virginia-Launches-Open-Data-Open-Jobs-Initiative.html

- Next Generation (NextGen) Analytics Pilot Program: The Commonwealth in November 2015 established contracts with 11 vendors to supply next-generation analytics services at zero cost to state agencies. These services cover both products and the resources needed to utilize those products.

Appendix 4: Links to Reference Documents for Data Sharing Terminology

Commonwealth Enterprise Information Architecture (EIA) Strategy:
http://www.vita.virginia.gov/uploadedFiles/VITA_Main_Public/Oversight/EA/Data_Management_Group/Commonwealth_EIA_Strategy_FINAL.pdf

Commonwealth Information Security Standard 501 (SEC501):
http://www.vita.virginia.gov/uploadedFiles/VITA_Main_Public/Library/PSGs/Information_Security_Standard_SEC501.pdf

Commonwealth Information Technology Resource Management (ITRM) Glossary:
http://www.vita.virginia.gov/uploadedFiles/VITA_Main_Public/Library/PSGs/PSG_Sections/COV_ITRM_Glossary.pdf

National Institute of Standards and Technology Interagency or Internal Report 7298, Release 2, Glossary of Key Information Security Terms:
http://nvlpubs.nist.gov/nistpubs/ir/2013/NIST.IR.7298r2.pdf