Prince William County, Virginia
Internal Audit Report: Data Governance

May 5, 2024

TABLE OF CONTENTS

TRANSMITTAL LETTER

May 5, 2024

The Board Audit Committee of
Prince William County, Virginia
1 County Complex Court
Prince William, Virginia 22192

Pursuant to the internal audit plan for calendar year ("CY") 2023 for Prince William County, Virginia ("County" / "PWC"), approved by the Board of County Supervisors ("BOCS"), we hereby submit the internal audit of data governance. We will be presenting this report to the Board Audit Committee of Prince William County at the next scheduled meeting on September 17, 2024.

Our report is organized into the following sections:

Executive Summary	This provides a high-level overview and summary of the observations noted in our internal audit over data governance and protection process(es).
Background	This provides an overview of the function, as well as relevant background information.
Objectives and Approach	The internal audit objectives are expanded upon in this section, as well as a review of the various phases of our approach.
Observations Matrix	This section includes a description of the observations noted during our internal audit and recommended actions, as well as Management response, including responsible party, and estimated completion date.
Appendix	This section provides additional information, documentation, and detailed explanations of supplemental results which support identified observations.

We would like to thank the staff and all those involved in assisting our firm with this internal audit.

Respectfully Submitted,

RSM US LLP

RSM US LLP

EXECUTIVE SUMMARY

Background

Data governance and management involve developing, executing, and supervising plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data assets throughout their lifecycles. This comprehensive approach supports data's strategic value while maintaining integrity and security. Effective data management encompasses both technical and non-technical roles, requiring collaboration between departments and the Department of Information Technology ("DoIT") to support high-quality data that meets the County's strategic needs.

The County leverages various digital applications and systems managed by DoIT to enhance operational efficiency. DoIT's responsibilities include developing, maintaining, and securing these applications while aligning technological infrastructure with strategic objectives, industry standards, and regulatory compliance. DoIT also plays a crucial role in managing information technology ("IT") risks, safeguarding data integrity, and supporting the continuous availability and reliability of essential services across the County's departments.

Currently, each department manages its data independently, with no centralized data governance function at the County level. Reliable, complete, accurate, and timely data are essential for County employees to serve constituents effectively, making robust data governance and management routines critical.

Data governance is a component of the broader discipline known as Enterprise Information Management. Within this context, the Data Management Association (DAMA) has defined a comprehensive framework encompassing numerous principles. While there are more than 12 guiding principles in total, our internal audit focused on the 12 principles most relevant to the County's objectives. These principles provide processes and controls so that data is effectively governed, protected, and leveraged to support the County's goals. Specifically, they validate that data is managed in a way that promotes accuracy, security, accessibility, and strategic value, thereby enhancing the County's ability to meet its operational and strategic objectives.

Overall Summary / Highlights

The observations identified during our assessment are detailed in the pages that follow. We have assigned relative risk or value factors to each observation identified. Risk ratings are the evaluation of the severity of the concern and the potential impact on the operations of each item. There are many areas of risk to consider in determining the relative risk rating of an observation, including financial, operational, and/or compliance, as well as public perception or 'brand' risk.

Fieldwork was performed December 2023 through May 2024.

Objectives and Scope

The purpose of this engagement was to assess the design, control structure, and overall maturity of the County's data governance framework across 12 primary data governance principles. This assessment aimed to identify strengths and weaknesses within the current data governance practices, guiding targeted improvements to enhance data quality and compliance. We collaborated with DoIT to identify the areas of the County most reliant on data. Our review focused on the following departments and the supporting infrastructure they utilize:

- DoIT
- Finance
- Social Services
- Human Resources

This review assessed data governance principles and process areas including:

- Governance and Oversight
- Data Management Strategy
- Data Owners and Stewards
- Data Labeling and Classification
- Reference and Master Data Management
- Data Architecture and Lineage
- Quality and Issues Management
- Data Catalog Management
- Data Retention and Destruction
- Asset Management
- Interface Management
- Report Management

Our audit period was at the point of time of our fieldwork.

Summary of Observation Ratings (See page 4 for definitions)			
	High	Moderate	Low
Data Governance	1	-	-

EXECUTIVE SUMMARY (CONTINUED)

Observations Summary

Below is a summary listing of the observations that were identified during this internal audit. Detailed observations are included in the observations matrix section of the report. See the following page for risk rating definitions. In this internal audit, we have prioritized a primary observation concerning the absence of an enterprise-wide data governance strategy, as establishing this foundational framework is essential for effective and consistent data management across the County. We have identified 11 additional findings, which are detailed in the appendix, as these issues should be addressed within the context of a comprehensive strategy.

Summary of Observations	
Observations	**Rating**
1. Absence of an Enterprise-Wide Data Governance Strategy	**High**

Supplemental Results (See Appendix A)
1. Centralized Asset Listing for Data-Processing Applications
2. Data Governance Policies
3. Formal Data Governance Strategy
4. Formally Established Data Owners and Stewards
5. Data Classification and Labeling Practices
6. Data Architecture and Lineage Documentation
7. Formal Data Quality Requirements
8. Inventory for Critical/Key Reports
9. Data Cataloging Procedures
10. Formal Data Retention and Destruction Program
11. Master Data Management Program

EXECUTIVE SUMMARY (CONTINUED)

Observations Summary (Continued)

Provided below are the observation risk rating definitions for the detailed observations.

Observation Risk Rating Definitions	
Rating	**Definition**
Low	Observation presents a low risk (i.e., impact on financial statements, internal control environment, or business operations) to the organization for the topic reviewed and/or is of low importance to business success/achievement of goals. Action should be taken within 12 months (if related to external financial reporting, must mitigate financial risk within two months unless otherwise agreed upon).
Moderate	Observation presents a moderate risk (i.e., impact on financial statements, internal control environment, or business operations) to the organization for the topic reviewed and/or is of moderate importance to business success/achievement of goals. Action should be taken within nine months (if related to external financial reporting, must mitigate financial risk within two months).
High	Observation presents a high risk (i.e., impact on financial statements, internal control environment, or business operations) to the organization for the topic reviewed and/or is of high importance to business success/achievement of goals. Action should be taken immediately, but in no case should implementation exceed six months (if related to external financial reporting, must mitigate financial risk within two months).

BACKGROUND

Overview

The County relies on various digital applications and technological solutions to facilitate and streamline its operations. The Department of Information Technology ("DoIT") is responsible for developing, maintaining, and securing these applications and systems. DoIT plays a key role in aligning the technological infrastructure with the County's strategic objectives, adherence to industry standards, and compliance with regulatory requirements. Additionally, the DoIT plays a pivotal role in managing the IT risk landscape, safeguarding data integrity, and facilitating the continuous availability and reliability of essential services across the County's various departments. Each department is currently responsible for data management within its area, and no centralized function has been established to govern data at the County level. We selected four departments to evaluate the existing data management practices listed below. In addition to the DoIT, this review considered the practices in place for the applications used by the Finance, Social Services, and Human Resources departments, for which the County is solely responsible.

Data is a crucial asset for the functions of the County. Without data that is reliable, complete, accurate, and timely, employees of the County will have insufficient information to properly serve its constituents. In order to validate that data is available to meet this need, it is critical that the appropriate data governance/management routines are in place.

Data governance/management is defined by the Data Management Association ("DAMA") as the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles. Data management activities can include everything from the ability to make consistent decisions about how to get strategic value from data to the technical deployment and performance of databases, as well as technical and non-technical roles and responsibilities. Departments and DoIT must be able to collaborate to support high quality data that meets the County's strategic needs.

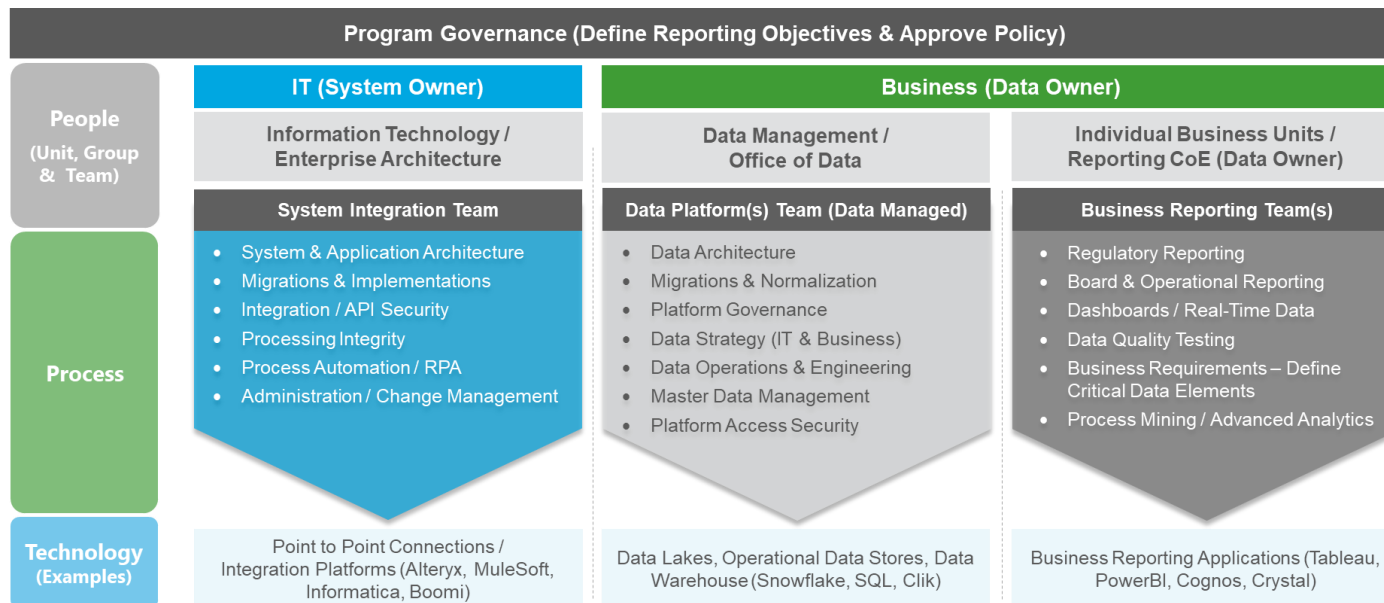The following principles are critical components of a robust data governance framework:

1. **Governance & Oversight** Governance and oversight involve establishing a structured framework for decision-making and accountability related to data assets. This includes defining policies, standards, and procedures for data management, as well as setting up a governance body to support compliance and oversee data-related initiatives.

2. **Data Management Strategy** A data management strategy outlines the vision, goals, and objectives for managing data as a valuable asset. It includes defining the processes and technologies needed to collect, store, manage, and use data effectively, aligning these efforts with the organization's strategic priorities.

3. **Data Owners and Stewards** Data owners and stewards are responsible for the management and oversight of specific data sets. Data owners have the authority to make decisions about data, while data stewards validate that data is properly managed, maintained, and used according to established policies and standards.

4. **Data Labeling and Classification** Data labeling and classification involve categorizing data based on its sensitivity, importance, and usage. This helps in managing access controls, validating data security, and complying with regulatory requirements. Proper classification enables efficient data management and retrieval.

5. **Reference and Master Data Management** Reference and master data management validate consistency and accuracy of critical data entities across the organization. Reference data includes standard values used in data systems, while master data refers to key business entities such as customers, products, and employees.

6. **Data Architecture and Lineage** Data architecture and lineage define the structure and flow of data within the organization. Data architecture provides a blueprint for data management systems, while data lineage tracks the data's origin, movements, and transformations across systems, supporting traceability and accountability.

BACKGROUND (CONTINUED)

Overview (Continued)

7. **Quality and Issues Management** Quality and issues management involve establishing processes to validate data accuracy, completeness, and reliability. This includes implementing data quality controls, monitoring data quality metrics, and addressing data-related issues promptly to maintain high standards of data integrity.

8. **Data Catalog Management** Data catalog management involves creating and maintaining an inventory of data assets within the organization. A data catalog helps users discover, understand, and access data, promoting transparency and efficient data usage.

9. **Data Retention and Destruction** Data retention and destruction policies validate that data is stored for appropriate periods and securely disposed of when no longer needed. This helps in managing storage costs, validating compliance with legal requirements, and protecting sensitive information from unauthorized access.

10. **Asset Management** Asset management involves tracking and managing data assets throughout their lifecycle. This includes maintaining an inventory of data resources, monitoring their usage, and validating they are adequately protected and maintained.

11. **Interface Management** Interface management validates seamless integration and interaction between different data systems and applications. This includes designing and managing data interfaces, validating data consistency, and facilitating data exchange across systems.

12. **Report Management** Report management involves the creation, distribution, and maintenance of data reports. This validates that stakeholders have access to accurate and timely information for decision-making, while also maintaining report consistency and compliance with reporting standards.

These principles provide a comprehensive framework for managing data as a strategic asset, validating that it is effectively governed, protected, and leveraged to support the County's objectives. The following figure (Figure 1) illustrates how these activities can span both technical and non-technical roles within the organization.

Program Governance (Define Reporting Objectives & Approve Policy)		
IT (System Owner)	**Business (Data Owner)**	
Information Technology / Enterprise Architecture	Data Management / Office of Data	Individual Business Units / Reporting CoE (Data Owner)
System Integration Team	**Data Platform(s) Team (Data Managed)**	**Business Reporting Team(s)**
• System & Application Architecture • Migrations & Implementations • Integration / API Security • Processing Integrity • Process Automation / RPA • Administration / Change Management	• Data Architecture • Migrations & Normalization • Platform Governance • Data Strategy (IT & Business) • Data Operations & Engineering • Master Data Management • Platform Access Security	• Regulatory Reporting • Board & Operational Reporting • Dashboards / Real-Time Data • Data Quality Testing • Business Requirements – Define Critical Data Elements • Process Mining / Advanced Analytics
Point to Point Connections / Integration Platforms (Alteryx, MuleSoft, Informatica, Boomi)	Data Lakes, Operational Data Stores, Data Warehouse (Snowflake, SQL, Clik)	Business Reporting Applications (Tableau, PowerBI, Cognos, Crystal)

People (Unit, Group & Team) — Process — Technology (Examples)

BACKGROUND (CONTINUED)

Maturity Assessment

We assessed the maturity of data governance at the County to be "**Initial**" based on the evaluation of the 12 principles as they apply to the in-scope departments and County as a whole. These maturity ratings are essential for identifying strengths and weaknesses within the County's data governance processes and controls, guiding targeted improvements, and facilitating alignment with best practices. The ratings provide a clear benchmark for understanding the current state and identifying areas for improvement. Please refer to the following page for detailed maturity rating definitions.

Summary of Data Governance Maturity		
Data Governance Principles		**Control Maturity (1 to 5)**
DG.1	Governance & Oversight	**1 - Initial**
DG.2	Data Management Strategy	**1 - Initial**
DG.3	Data Owners and Stewards	**1 - Initial**
DG.4	Data Labeling and Classification	**2 - Managed**
DG.5	Reference and Master Data Management	**1 - Initial**
DG.6	Data Architecture and Lineage	**1 - Initial**
DG.7	Quality and Issues Management	**1 – Initial**
DG.8	Data Catalog Management	**1 – Initial**
DG.9	Data Retention and Destruction	**2 - Managed**
DG.10	Asset Management	**1 – Initial**
DG.11	Interface Management	**2 – Managed**
DG.12	Report Management	**2 - Managed**

BACKGROUND (CONTINUED)

Maturity Assessment (Continued)

The following table represents the guiding principles for rating each data governance principle.

Initial (1)	Managed (2)	Defined (3)	Repeatable (4)	Optimizing (5)
Processes are ad hoc and disorganized. There is evidence that the organization has recognized that issues exist; however, there are no standardized processes. The overall approach to management is disorganized.	Processes have developed to the point where different people undertaking the same task follow similar procedures. There is no formal training or communication of standard procedures, and responsibility is left to the individual. There is a high degree of reliance on individuals' knowledge, and errors are likely to result.	Processes are documented and communicated. Procedures have been standardized, documented, and communicated through formal training. The procedures are not sophisticated but are the formalization of existing practices.	Processes are monitored and measured. Compliance with procedures can be monitored and measured, and action can be taken where processes appear not to be working effectively. Processes are constantly improved and provide good practices. Automation and tools are used in a limited or fragmented way.	Best practices are followed and automated. Processes have been refined to a level of best practice based on the results of continuous improvement and benchmarking with other organizations and industry best practices. IT is used in an integrated way to automate the workflow, providing tools to improve quality and effectiveness.

OBJECTIVES AND APPROACH

Objectives

The purpose of this engagement was to assess the design, control structure, and overall maturity of the County's data governance framework across 12 primary data governance principles. This assessment aimed to identify strengths and weaknesses within the current data governance practices, guiding targeted improvements to enhance data quality and compliance. The scope focused on understanding the current design and control structures without testing the effectiveness of these controls, validating alignment with best practices and regulatory requirements.

Approach

Our audit approach consisted of the following phases:

Understanding and Documentation of the Process
This phase consisted primarily of inquiry and walkthroughs to obtain an understanding of the current operating policies and procedures, monitoring functions, and control structures as they relate to the processes within our scope. The following was performed as part of this phase:
- Obtained and reviewed any documented policies and procedures related to the function, as well as relevant state and County regulations, reporting, and other relevant information.
- Conducted interviews and walkthroughs with key personnel to obtain a detailed understanding of operating policies and procedures, roles, and responsibilities within the DoIT.
- Gained an understanding of procedures related to the processes within scope.
- Developed a work plan for evaluating the design of procedures and controls based on information obtained through interviews, walkthroughs, and preliminary review of documentation.

Evaluation of the Design of Controls
The purpose of this phase was to evaluate the design of key processes and controls based on our understanding of the processes obtained during the first phase. Our testing procedures included, but were not limited to:
- Determined the adequacy of policies and standard operating procedures related to data governance.
- Evaluated the design and control structure, including adherence to policies and procedures regarding key data governance principles, including:
  o Governance & Oversight
  o Data Management Strategy
  o Data Owners and Stewards
  o Data Labeling and Classification
  o Reference and Master Data Management
  o Data Architecture and Lineage
  o Quality and Issues Management
  o Data Catalog Management
  o Data Retention and Destruction
  o Asset Management
  o Interface Management
  o Report Management
- Based on our analysis and testing results, we developed recommendations for process and control modification/addition/deletion for any identified design gaps or non-compliance issues.

Reporting
At the conclusion of this internal audit, we summarized our findings into this report. We have reviewed the results with the appropriate Management personnel and have incorporated Management responses into this report.

OBSERVATIONS MATRIX

Observation	1. Enterprise-Wide Data Governance Strategy
High	Effective data governance requires a formal, enterprise-wide strategy to set standards for data management, assign accountability, and establish controls over data integrity, availability, and quality. According to the Data Management Association (DAMA), a data governance framework should cover key elements such as data management strategy, governance and oversight, data quality standards, data classification and security, ownership and stewardship, data architecture, and data cataloging.
	The County has not designed or implemented a centralized data governance strategy. Currently, data management practices vary significantly across departments, with each department independently responsible for data oversight, management, and security. While the Department of Information Technology (DoIT) has some data management processes in place, no unified approach exists to enforce consistent data governance standards, coordinate efforts across departments, or align data practices with County-wide strategic objectives.
	Without an enterprise-wide data governance framework, the County faces several risks, including data inaccuracies, inconsistent data management practices, inefficiencies in data access, and potential regulatory non-compliance. This lack of coordination hinders the County's ability to leverage data as a strategic asset, make informed decisions, protect data and validate quality standards across all departments. The County's data assets may not fully support operational needs or meet regulatory requirements, ultimately impacting service delivery and public trust.
Recommendation	We recommend the County establish an enterprise-wide data governance program that incorporates the following critical elements starting with:

- **Data Governance Strategy**—Develop a formal strategy that aligns with the County's strategic objectives and outlines the vision, goals, and objectives for data governance. This strategy should guide data management practices across departments and serve as a foundation for consistent and effective data use.
- **Centralized Policies and Standards**—Implement County-wide policies and standards covering data quality, classification, labeling, retention, destruction, and security. Validate that these policies are documented and accessible to all departments to establish clear, standardized guidelines for data handling.
- **Governance and Oversight**—Establish a data governance committee with representatives from key departments to oversee data governance initiatives, monitor compliance with policies, and promote accountability.
- **Roles and Responsibilities**—Assign data owners and data stewards for each critical data asset to support accountability for data quality, security, and compliance. Define the responsibilities of these roles to include data lifecycle management, quality monitoring, and adherence to governance standards.
- **Data Quality Management**—Define data quality requirements such as accuracy, completeness, and timeliness. Implement a monitoring program to regularly assess data quality across departments, with a mechanism to log, track, and address data quality issues.
- **Data Classification and Security**—Develop data classification protocols to categorize data based on sensitivity, regulatory requirements, and usage. Establish security measures aligned with classifications to control access and protect sensitive information.

To further strengthen the program, the following elements should be implemented as secondary priorities:

- **Data Architecture and Lineage**—Document the County's data architecture to define data flows, dependencies, and relationships among data sources. This should include data lineage documentation to support data traceability and accountability.
- **Data Catalog and Inventory**—Establish a data catalog that inventories all critical data assets, providing metadata, data definitions, and access requirements. This catalog will promote data discovery and enable departments to leverage shared data resources effectively.

OBSERVATIONS MATRIX (CONTINUED)

Observation	1. Enterprise-Wide Data Governance Strategy (Continued)
Recommendation	• **Training and Awareness**—Provide training to all personnel involved in data governance to confirm they understand policies, roles, and responsibilities. Ongoing training and awareness programs will reinforce the importance of data governance and promote compliance across departments. • **Performance Metrics and Reporting**—Define performance metrics to measure the effectiveness of the data governance program, such as data quality scores, compliance rates, and issue resolution timelines. Regularly report on these metrics to the governance committee and senior leadership to monitor progress and make data-informed adjustments as needed.
Management Action Plan	**Response:** Management concurs. Implementation of a countywide data governance strategy is a substantial undertaking that will require significant resources, both in terms of human capital and budgetary support. The ability to harness, store, and interpret data to transform information into a valuable County asset for better decision-making support, cross-departmental collaboration, risk reduction, enhanced performance management, improved scalability and efficiency, and cost reduction thru a countywide data governance strategy is a multi-year endeavor given current resource constraints and competing priorities across the organization. At this time, management can commit to establishing a Data Governance Steering Committee to develop a roadmap that incorporates elements identified in the appendix of this report that could ultimately lead to a data governance strategy. This roadmap will also include the identification of the resource needs in advance of the FY 2026 and/or FY 2027 budget processes and will complement the BOCS' and Executive Leadership's desire to move the County in the direction of collecting and measuring outcomes with more robust Key Performance Indicators (KPIs) through a comprehensive, centralized Performance Management function. **Responsible Party:** Office of Executive Management, Office of Management & Budget and DoIT **Estimated Completion Date:** FY2027

APPENDIX A: SUPPLEMENTAL RESULTS

The following results are ordered by significance to the organization and relative impact of data governance processes and controls.

1. Centralized Asset Listing for Data-Processing Applications

Management does not maintain a centralized asset listing to identify applications processing, storing, or housing data at the County or department level. We noted that an infrastructure modernization project is in progress to review and document applications and interfaces currently utilized by the various departments within the application database, but a complete inventory of applications and interfaces does not exist.

The lack of a centralized asset listing may lead to challenges in managing and securing applications and interfaces, potentially resulting in data security risks and inefficiencies in application management. The high risk associated with this observation is due to the critical nature of maintaining an accurate and comprehensive inventory for effective data governance and cybersecurity.

Recommendation The County should establish an application inventory. This inventory should document all elements deemed necessary by management, and the following items should be considered for inclusion:

- Asset name (specific names and versions versus generic)
- Asset owner (department and specific owner/team; consider alignment with vendor relationship owner)
- Asset type (in-house or outsourced)
- Business purpose (brief description of the purpose/use of the application)
- Data classification (does the system have GLBA/customer information?)
- Data volume (does the system have more than 500 records?)
- Sunset (less than two years, two to five years, or over five years
- Operational criticality (leverage the BIA for criticality and impact ratings)

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

2. Data Governance Policies

The County and sampled departments lack a comprehensive data governance policy that defines responsibilities for managing and overseeing the data governance program. This absence of a formal policy compromises accountability for data integrity, accuracy, and completeness within the environment. While some aspects, such as data storage/backup, data classification, and data retention/destruction, are covered by existing guidelines, several critical principles of data governance are not currently addressed by policy or guidelines. These principles include:

• Data ownership and stewardship • Reference and master data management • Data migration • Data architecture and lineage	• Data quality and issue management • Data catalog management • Data asset management • System interface management • Key report inventory management

The lack of a comprehensive data governance policy may lead to inconsistencies, inaccuracies, and inefficiencies in data management, potentially affecting decision-making and operational effectiveness.

Recommendation	We recommend the following:
	
	1. Determine Governance Level: Decide whether data governance will be centralized at the County level or decentralized at the department level.
	
	2. Establish Policies and Guidelines: Develop well-defined, clear, and documented policies and guidelines that outline the roles, responsibilities, and objectives of data governance at the chosen level. These should be supported by:
	   a. Procedure Documentation: Provide detailed instructions on how the organization will adhere to policy directives.
	   b. Training Programs: Implement training for individuals with defined roles within data governance to clarify their responsibilities.
	
	3. Address Key Principles: The documentation and training should cover the following critical principles of data governance, along with any others identified as essential by management:

• Data ownership and stewardship • Reference and master data management • Data migration • Data architecture and lineage	• Data quality and issue management • Data catalog management • Data asset management • System interface management • Key report inventory management

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

3. Formal Data Governance Strategy

Management has not formally established a data governance strategy, charter, or framework that outlines plans for maintaining and improving data quality, data integrity, critical success factors, and overall data governance goals at the County or department level. Although the DoIT has components of a roadmap to address critical data governance components related to PCI and other compliance items, an overall strategic approach has not been established. Successful data governance initiatives require collaboration from the business and IT at an enterprise level and cannot be driven solely by individual units or intradepartmental initiatives.

The absence of a formal data governance strategy may result in inconsistent data management practices, reduced data quality, and potential challenges in meeting compliance requirements. This could ultimately affect the County's ability to make informed decisions and achieve its business objectives.

Recommendation	The County should establish a working group between the various departments and include key functions such as information security, compliance, critical department leadership, and project teams to establish a data governance charter/strategy. A data governance strategy is a comprehensive plan that outlines how the organization plans to achieve business objectives related to data. Management should establish the goals and initiate the management and utilization of data as a strategic asset. Consider the below elements when establishing a data management strategy:
	• Business alignment: Align the data strategy with the organization's overall business strategy and goals. Validate that data initiatives and investments are closely tied to the organization's strategic priorities and support its core business processes.
	• Governance and accountability: Define the governance structure for data management, including roles, responsibilities, and decision-making processes. Identify data owners, data stewards, and data governance committees responsible for overseeing data-related activities and validating compliance with policies and regulations.
	• Data life cycle management: Outline the stages of the data life cycle and define processes and best practices for data acquisition, data entry, data storage, data integration, data quality management, data archiving, and data disposal. Specify how data will be managed from creation to retirement.
	• Data architecture: Describe the target data architecture and data integration strategies that will support the organization's data needs. Identify data sources, data repositories, data integration tools, data standards, and data models required to enable effective data management and data integration across the organization.
	• Implementation road map: Develop a detailed implementation road map that outlines the sequencing, timelines, and dependencies of the data initiatives. Prioritize the key projects and initiatives, allocate resources, and establish milestones and metrics to measure progress and success.

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

4. Formally Established Data Owners and Stewards

Data owners and stewards have not been formally established at the county or department level. Management has not implemented a data governance policy or guideline defining responsibilities for managing and overseeing data assets, which is essential for supporting accountability for the integrity, accuracy, and completeness of data within the environment at the County or department level. While some historical data points exist within geospatial data maintained by the Department of Information Technology (DoIT), management has not identified data owners and stewards for critical data elements at the County or department level.

The lack of formally established data owners and stewards may lead to unclear accountability for data management, potentially resulting in inconsistencies, inaccuracies, and inefficiencies in data handling and decision-making processes.

Recommendation	The County should establish data owners and stewards who play a crucial role in managing and maintaining data assets. The data governance program should outline the responsibilities of owners and stewards, including:

- Documenting data definitions
- Managing data quality
- Reviewing data access and security
- Complying with data privacy and regulatory requirements

Once data owners and stewards are established, management should:

- Provide Training: Offer training to validate that individuals understand their responsibilities and the importance of their roles in data governance.
- Maintain Inventories: Develop and maintain inventories of critical data elements, data sources, and key reports, along with associated data owners and stewards. These inventories should detail data sources, dependencies, and lineage to promote accountability and transparency in data management.

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

5.	Reference and Master Data Management

Management has not documented data classifications and data labeling requirements at the County or department level. While in practice, Management has established data classifications within specific systems at each of the departments based on applicable regulations (e.g., PCI, HIPAA, etc.), Management does not require and has not implemented a mechanism by which to track data classifications for data elements, including the systematic enforcement of data classifications for unstructured data. Therefore, data is not currently labeled with the appropriate data classifications for unstructured data. DoIT has purchased Varonis as a tool with the capability to automatically scan and apply data classifications for unstructured data. Varonis has the ability to apply predefined rules to scanned data to identify and label data based on classifications (specifically, this tool is well adapted at identify certain common non-public personal information (NPPI) such as social security numbers, etc.); however, this tool has not yet been configured to actively enforce data labeling within the environment.

The lack of documented data classifications and labeling requirements at the County or department level may lead to inconsistencies in data handling, potential non-compliance with regulations, and challenges in managing data security and privacy.

Recommendation	We recommend the following:

- Define the requirements of data owners and stewards to classify data based on existing guidelines.
- Develop training to educate the business and data owners on the classification and labeling of the data for which they are responsible. This training should make reference to existing policies/guidelines and other documentation to validate that end users are aware of their responsibilities and the resources available to them.
- Once data has been properly classified, implement controls to appropriately control data based on its classification.
- Implement processes for monitoring and validating the quality and consistency of data labeling and classification practices.
- Consider adopting a rule-based tool to automatically apply data labels to data to validate that data is labeled properly and to lessen the burden on end users to apply data labels.

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

6. Data Architecture and Lineage Documentation

Data architecture and lineage documentation have not been established at the county or department level. Management has not formally established a data architecture design process to validate that data flow documents for key systems are up to date and accurately represent the enterprise data environment at the County or department level.

The absence of formal data architecture and lineage documentation may lead to challenges in understanding data flow, data dependencies, and the overall data environment, potentially affecting data management and decision-making processes.

Recommendation	We recommend the following:
	• Define and Document Data Ownership: Clearly define and document data ownership for critical data elements. Once established, outline the requirements for data owners regarding the documentation of data flow and architecture for these elements.
	• Develop Data Flow Diagrams: Create data flow diagrams to provide a visual representation of the organization's data architecture, including the structure, components, and relationships. Key components to consider in these diagrams include:
	○ Data Sources: Identify databases, applications, external systems, data feeds, application programming interfaces (APIs), and file systems.
	○ Data Integration and ETL Processes: Show the flow of data between various components, including transformation and data mappings.
	○ Data Models and Data Structures: Include data models and data structures that define the organization and relationships of data, such as data schemas or other modeling techniques.
	○ Data Analytics and Reporting: Depict components and technologies used for data analytics, reporting, and business intelligence, including data visualization tools, reporting platforms, and analytical models.
	○ Data Archiving and Purging: If relevant, illustrate the mechanisms and processes used for data archiving and purging, including how data is retained for historical purposes or compliance requirements and how it is eventually removed from the system.

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

7. Formal Data Quality Requirements

Data quality requirements have not been established at the County or department level. Management has not formally defined data quality requirements at the County or department level, including uniqueness constraints to prevent duplication and missing data in datasets. While each system has a regular process to address data quality issues, these processes vary based on the ownership of the system (State vs. County) and other factors. A centralized inventory of data quality issues or incidents is not maintained due to the decentralized nature of this process. Currently, data quality issues are tracked through email or various other means.

The absence of formal data quality requirements and a centralized tracking system may lead to inconsistencies in data quality management, potential duplication or missing data, and challenges in addressing and resolving data quality issues effectively.

Recommendation	The County should document data quality standards that include data elements such as accuracy, completeness, consistency, integrity, and validity. A program should be established to outline the guidelines for achieving and maintaining the desired data quality standards. This program should include:

- Documentation of Data Processes: Clearly document data collection, data entry, data validation, data cleansing, and data ownership requirements.
- Establishment of a Central Repository: Create a central repository for tracking data quality issues and incidents.
- Conduct Training Sessions: Once the program is established, management should conduct training sessions to educate stakeholders in the data quality process about their responsibilities and the importance of data quality.
- Quality Metrics and Monitoring: After the program has been established, management should set up quality metrics and monitoring processes to aim for adherence to the standards.

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

8. Inventory for Critical / Key Reports

Management does not maintain an inventory of critical / key reports in use at the County or at the department level. An inventory of critical reports should include data owners, source systems, data definitions, consumers, access requirements, frequency, and other details identified by business owners to aid end users in understanding reports and support processes in the criticality of report accuracy and completeness. While critical reports used in financial reporting and other subsets of report populations are documented in various locations, a complete population of critical/key reports does not exist.

The absence of a comprehensive inventory of critical/key reports may lead to challenges in managing and understanding report usage, potentially affecting decision-making processes and the reliability of reports.

Recommendation

The County should establish an inventory of key reports for each department and consider centralizing this inventory for the entire County. Once established:

- Data owners should be identified for all key reports.
- Requirements around key reports should be communicated to the identified owners.
- Once the data governance program has been established, additional requirements around key reports should be identified based on the nature of the reports.

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

9. Data Cataloging Procedures

Data cataloging procedures have not been established at the county or department level. Management has not established a defined data management program or data catalog at the County or department level. A data catalog is a collection of metadata, combined with data management and search tools, which helps analysts and other data users find the data they need, serves as an inventory of available data, and provides information to evaluate the fitness of data for intended uses.

The absence of data cataloging procedures may lead to challenges in data discovery, data inventory management, and evaluating the suitability of data for specific purposes. However, the risk associated with this observation is considered low.

Recommendation	The County should evaluate the need for a data catalog to meet the various departments' needs. If a data catalog is deemed necessary, we recommend that Management:
	• Include Appropriate Metadata: Make sure that data sets within the data catalog include appropriate metadata, such as data source, quality, lineage, and ownership.
	• Utilize Data Owners: Once established, consider utilizing data owners to aid in the documentation of metadata for existing data sources and consider discovery activities to identify additional data sources for ingestion into the catalog.
	• Focus on Key Data Elements: To aid in implementing the data catalog, consider focusing first on data utilized in key reports and other critical data elements. Work to find a subset of business partners to build collaborative relationships and identify road maps for implementation and the most valuable components of the data catalog for the business.
	• Integrate into Data Governance Strategy: These efforts should then be integrated into the overall data governance strategy or roadmap to demonstrate the value of this process to the business.

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

10. Formal Data Retention and Destruction Program

Data retention and destruction requirements have not been established at the County or department level. While the Library of Virginia record retention schedules guide data retention and Laserfiche is used to retain records, a comprehensive data retention and destruction program covering all aspects of data lifecycle management has not been implemented.

The lack of a formal data retention and destruction program may lead to challenges in managing data's lifecycle and validating compliance with legal and regulatory requirements.

Recommendation	The County should document data retention and destruction practices in place to comply with requirements established by the Library of Virginia. This documentation should also address how data that falls outside the purview of these requirements is to be handled. By doing so, the County can aim to validate a consistent and compliant approach to managing the data lifecycle.

APPENDIX A: SUPPLEMENTAL RESULTS (CONTINUED)

11. Master Data Management Program

Management has not established a master data management program to organize the availability of accurate and current data values at the County or department level. A master data management program establishes controls over master data values and reference data values that enable consistent use across systems of the most accurate and timely data with regard to essential business entities. Activities around master data management include establishing a system of record for master data, reconciling and consolidating data across sources to provide a master record or source of truth value, addressing mismatched values, and enforcing master data values.

The lack of a master data management program may lead to inconsistencies in data usage and challenges in maintaining data accuracy and currency. However, since the County's systems are intentionally not referenced against each other, the risk associated with this observation is considered low.

Recommendation	The County should evaluate the need for a master data management program to meet the County's requirements. If a master data management program is deemed necessary, we recommend that management:

- Establish a Master Data Management Program: Organize the availability of accurate and current data values for critical data elements. This program should set requirements for data to promote consistent documentation of these elements and provide resources for data users.
- Leverage a Data Catalog Solution: Consider using a data catalog solution to establish uniform naming conventions with required fields. This will help management organize data so end users can access it appropriately to support reporting and business needs.
- Tagging and Labeling Source Data: Implement tagging and labeling of source data to aid in the availability and accuracy of the master data being managed.
- Address Data Migration Requirements: Enhance existing documentation or establish new policies and procedures to address data migration requirements, including transforming, archiving, and removing data from environments.

RSM US LLP
1861 International Drive
Suite 400
McLean, VA 22102
321.751.600

www.rsmus.com

For more information, visit rsmus.com/aboutus for more information regarding RSM US LLP and RSM International.

RSM