Commonwealth of Virginia

Enterprise Architecture Standard (EA-225)

Enterprise Information Architecture (EIA) Requirements

vita.virginia.gov

June 10, 2025

Revision History

Enterprise Information Architecture Requirements: Version History		
Revision	**Date**	**Description**
V1	10/2023	Initial version
V2	12/10/2024	ODGA updates to add governance benefits, additional code sections, best practices, and reporting requirements
V3	6/10/2025	CSRM Data Classifications integrated

Review Process
This requirements document was posted on VITA's Online Review and Comment Application (ORCA). All agencies, stakeholders, and the public were encouraged to provide their comments through ORCA. All comments were evaluated, and Individual commenters were notified of action(s) taken.

Requirements and Agency Exceptions
The requirements within this document are mandatory. Agencies deviating from these requirements must request an exception for each desired deviation by completing and submitting an exception request as described within the *Commonwealth Enterprise Architecture Policy*. The approved *Enterprise Architecture Exception* will be sent to the requester via Archer, and must be received prior to developing, procuring, or deploying such technology. Failure to follow the requirements specified in this document is considered non-compliance.

Glossary
As appropriate, terms and definitions used in this document are in the COV ITRM IT Glossary. The COV ITRM IT Glossary is available on the ITRM Policies, Standards, and Guidelines web page at the VITA website: COV ITRM IT Glossary

Contents

Vision and Strategy
Vision
The Enterprise Information Architecture Standard shall build a strong information foundation by integrating Data Governance, Data Stewardship, and Metadata Management. Data Governance establishes responsibility and integrity. Data Stewardship ensures accuracy through ownership. Metadata Management provides context for data. Together, they enable new data insights and enhance decision-making for all COV agencies. Compliance with this standard will be assessed through regular audits, adherence to documented data quality metrics, and periodic reviews of data governance practices to ensure that agencies are following the established guidelines and maintaining high standards of data management.
Strategy
Objective 1: Establish Data Governance and Data Stewardship Appoint Data Stewards to oversee data accuracy and completeness, ensuring data assets are properly managed and maintained throughout their lifecycle. **Objective 2: Data Management Foundation, Quality and Practices** Improve data management with the use of modernized discovery to enable fundamental data quality improvement practices. **Objective 3: Facilitate Data Sharing Frameworks** Utilize tools and create frameworks for seamless data sharing, enabling transparent collaboration and informed decision-making across COV agencies. **Objective 4: Drive Innovation and Security through Unified Data Management** Exercise full potential of COV data by driving transformative changes in service delivery and fostering data innovation and security.

Purpose

The purpose of this standard is to guide the purchase, design, implementation, and on-going operation of Commonwealth data management. Strong data governance and management benefits the Commonwealth by providing:

1. Improved Decision-Making and Resource Allocation

- **Access to Reliable Data**: A strong data governance program ensures data is accurate, complete, and accessible, enabling informed policy decisions.
- **Targeted Investments**: Better data helps identify areas of greatest need, ensuring state resources are deployed where they are most effective.

2. Foundation for Innovation and Economic Growth

- **Data as a Strategic Asset:** A robust governance program lays the groundwork for leveraging data in advanced analytics, AI, and predictive modeling.
- **Economic Competitiveness:** A data-driven government attracts businesses and fosters partnerships that rely on reliable and accessible public data.

3. Alignment with Federal Standards and Funding Opportunities

- **Federal Requirement:** Several federal programs and initiatives require robust data management and reporting standards to ensure accountability, compliance, and effective implementation. These programs span various sectors, including health, education, infrastructure, and more.

Failing to meet these data management and reporting requirements can lead to penalties, reduced funding, or compliance issues.
- **Funding:** Strong data governance ensures the state can not only meet these standards but also use the data effectively to maximize the benefits of federal programs.

4. Increased Public Trust and Transparency
- **Accountability:** Consistent, well-documented data practices foster transparency in government operations and decision-making.
- **Data Privacy and Security**: A governance framework ensures compliance with data protection laws, safeguarding residents' personal information and reinforcing trust.

5. Enhanced Service Delivery
- **Personalized Services:** High-quality data allows for tailored programs and services that address the specific needs of Virginia's residents.
- **Faster Response Times:** Streamlined data processes improve the state's ability to respond to emergencies and emerging issues.

6. Operational Efficiency and Cost Savings
- **Reduced Redundancies:** Proper governance minimizes duplicate efforts across agencies by standardizing data practices and encouraging data sharing.
- **Risk Mitigation**: Proactively managing data reduces the likelihood of costly errors, security breaches, and legal liabilities.

Scope

This standard is applicable to all Commonwealth agencies (hereinafter collectively referred to as "agencies") that are responsible for the management, development, purchase and use of information technology resources in the Commonwealth of Virginia and applies to all data assets.

This standard does not apply to research projects, research initiatives, or instructional programs at public institutions of higher education.

In addition to the requirements below, all COV IT technology solutions comply with the standards found on VITA's Policies, Standards & Guidelines.

This document is a dynamic, foundational framework and will continue to evolve as COV data management practices mature. This framework provides a set of criteria against which agencies can assess compliance with COV data management standards and with the overall maturity of agency data management practices.

Authority

Code of Virginia, §2.2-2007	Powers of the CIO
Code of Virginia, §2.2-2007.1	Additional duties of the CIO relating to information technology planning and budgeting
Code of Virginia, §2.2-2009(A)	Additional duties of the CIO relating to security of government information

Code of Virginia, §2.2-2012(A)	Additional powers and duties related to the procurement of information technology
Code of Virginia, § 2.2-203.2:4	Office of Data Governance and Analytics; Chief Data Officer; creation; report
Code of Virginia § 2.2-3800	Government Data Collection and Dissemination Practices Acti

Introduction

The roles of Data Owners, Data Stewards and Data Custodians are pivotal in ensuring the effective sharing and management of critical data assets across COV agencies. These roles form the foundation of our comprehensive COV Data Governance Framework and overall EIA strategy. By promoting efficient data utilization, we facilitate better decision-making and improve service delivery. Implementing a robust governance framework and adhering to best practices help establish a culture of responsible data management and encourage collaboration among agencies. This strategic approach enables COV agencies to maximize the benefits derived from their data assets.

There are three key roles covered by this standard: Data Owner, Data Steward, and Data Custodian.

Data Owner:

- Senior executive/business leader accountable for ultimate organizational responsibility for data and has budgetary responsibility. They make high-level decisions about data usage, access, and policies including decisions regarding data classification and protection. They are accountable for data quality, accuracy, integrity and value to the organization.

Data Steward

- Subject matter expert who handles day-to-day data management including ensuring data quality and consistency within a specific domain or subject area. They implement data governance policies set by data owners including access and usage, manage data quality, as well as develop and maintain data definitions and metadata. Stewards are responsible for classifying the data, ensuring it is appropriately labeled, and defines authorized access to the data.

Data Custodian

- Technical staff, including Suppliers, responsible for protecting and preserving data assets. data storage and security. They implement required classification and access controls and security measures as specified by data owners, manage storage, backups, recovery, and technical maintenance. They are often a database administrator or data engineer.

Requirements for the role of Data Owner

EIA-099: The Data Owner shall be accountable for data quality, accuracy, integrity and value to the organization for all data sets they are designated as the owner.

EIA-100: The Data Owner shall define, establish, and/or adopt data governance frameworks that align with organizational objectives and regulatory requirements.

EIA-101: The Data Owner shall be accountable for decisions on data usage, data classification, access, and protection.

EIA-102: The Data Owner shall ensure that the following required policies are developed and current for their agency: Data Governance, Metadata Management, Data Quality, Data Security, Data Privacy, Data Retention, and Data Stewardship.

EIA-103: The Data Owner shall establish a data sharing agreement with the Commonwealth Data Trust and create and maintain inter-agency Memorandum of Understanding agreements where applicable.

EIA-104: The Data Owner shall support the sharing of data between agencies while adhering to all statutory and regulatory requirements.

EIA-105: The Data Owner shall ensure the quality, integrity, and reliability of their respective data sets by enforcing data quality standards.

EIA-106: The Data Owner shall ensure that data management practices comply with all relevant laws and regulations to mitigate risks associated with non-compliance.

EIA-107: The Data Owner shall establish or participate in a Data Governance Council that meets at least twice a year to review progress on key data initiatives, review policies, and manage the data governance program.

EIA-108: The Data Owner shall develop or contribute to an agency data strategy which is shared with the Executive Data Board annually.

EIA-109: The Data Owner shall designate Data Stewards for each data domain to perform day-to-day operational tasks associated with data assets, system documentation, and metadata.

EIA-110: The Data Owner shall collaborate with Data Stewards to maintain data quality, identify potential governance issues, and enforce governance policies.

EIA-111: The Data Owner shall determine who can access the data sets and shall review these access privileges annually to ensure they remain appropriate.

EIA-112: The Data Owner shall specify and enforce data access controls based on agency data classification to protect information resources from unauthorized access, use, alteration, or destruction.

EIA-113: The Data Owner shall be accountable for ensuring controls are in place to maintain agency data confidentiality, integrity, and availability.

EIA-114: The Data Owner shall maintain an inventory of all data assets for which they are accountable.

EIA-115: The Data Owner shall complete the role-based training required in Sec 527 – Cybersecurity Awareness Training Standard.

EIA-116: The Data Owner shall champion participation in COV's Data Governance Boards and Councils.

EIA-117: The Data Owner shall review and update their data classifications annually.

EIA-118: The Data Owner shall review the data risk log at least annually and ensure escalation and remediation of issues. A sample template is available at Agency Resources.

EIA-119: The Data Owner shall monitor data privacy and data security compliance and address issues expeditiously.

EIA-120: The Data Owner shall provide an annual report to the COV Data Governance Council to include a data stewardship scorecard, data security incidents summary, privacy compliance attestation, and data risk log including opened and closed issues for the last 12 months.

Requirements for the role of Data Steward

Data Security and Privacy

EIA-201: The Data Steward shall provide physical and procedural safeguards to protect the organization's information resources as specified by Data Owners.

EIA-202: The Data Steward shall implement monitoring techniques and procedures for detecting, reporting, and investigating data-related incidents.

EIA-203: The Data Steward shall assist Data Owners in evaluating the effectiveness of data controls and monitoring compliance with these controls.

Data Governance and Compliance

EIA-204: The Data Steward shall provide technical subject matter expertise in support of data policies, standards, and best practices to enhance data governance across the Commonwealth.

EIA-205: The Data Steward shall implement data governance policies set by data owners including access and usage.

EIA-206: The Data Steward shall support informed, data-driven decision-making through compliance with Commonwealth data policies, standards, and best practices.

EIA-207: The Data Steward shall ensure adherence to Commonwealth data standards and data sharing requirements.

Metadata

EIA-208: The Data Steward shall establish and maintain a metadata repository or leverage the COV metadata repository.

EIA-209: The Data Steward shall curate and document contextual metadata about key data assets according to standard operating procedures.

EIA-210: The Data Steward shall actively seek ways to improve metadata management processes and adopt best practices.

EIA-211: The Data Steward shall identify and register COV Critical Data Assets for metadata collection, including data types, relationships, constraints, classification, and data lineage. **COV Critical Data Asset** are any data sets considered essential to the successful fulfillment of the missions of one or more COV agencies.

EIA-212: The Data Steward shall maintain supplemental data artifacts, including documentation, data definitions, and entity-relationship diagrams.

EIA-213: The Data Steward shall ensure and verify the data classification attributes of their associated COV Critical Data Assets.

EIA-214: The Data Steward shall be responsible for the lifecycle of COV Critical Data Assets from creation to retirement, including data retention and archival as defined in the Library of Virginia retention schedules and in partnership with the agency Retention Officer.

EIA-215: The Data Steward, in conjunction with the Data Owner, shall monitor and verify the quality of system metadata for accuracy, consistency, and reliability.

EIA-216: The Data Steward shall identify assigned master data fields in their data sets by tagging them as such in metadata definitions.

EIA-217: The Data Steward shall contribute to the Commonwealth's Business Glossary.

Sharing

EIA-219: The Data Steward shall implement data sharing and analytics projects that promote data accessibility, sharing, and reuse, thereby reducing redundancy across the Commonwealth.

EIA-220: The Data Steward shall promote the collection and sharing of metadata by registering all data assets, both structured and unstructured, in the Commonwealth Data Catalog to enhance data discoverability and utility.

EIA-221: The Data Steward shall identify datasets that can be shared with the citizens of the Commonwealth, either in summarized, redacted, or detail form, and publish them to Virginia's Open Data Portal.

Data Quality

EIA-222: The Data Steward shall be responsible for ensuring data quality through data profiling and cleansing.

EIA-223: The Data Steward shall implement data quality processes and standards within their agency and ensure compliance with standards.

Training

EIA-224: The Data Steward shall complete Data Stewardship training annually.

EIA-225: The Data Steward shall partner with the agency ISO to ensure all employees and contractors complete required data privacy and data security training.

EIA-226: The Data Steward shall encourage employees and contractors to complete data literacy training.

EIA-227: The Data Steward shall participate in the COV Data Stewards Group.

EIA-228: The Data Steward shall provide communication and education to data users on the appropriate use, sharing, and protection of the Commonwealth's data assets to foster a secure data environment.

EIA-229: Data Stewards and Data Custodians shall be part of incident response teams for issues where metadata plays a crucial role in incident resolution.

EIA-230: The Data Steward shall actively participate in change management by reviewing and approving changes to system metadata using version control.

EIA-231: The Data Steward shall facilitate communication and collaboration among teams regarding their assigned data set metadata.

Reporting

EIA-232: The Data Steward shall develop and maintain a data stewardship scorecard to track key metrics and provide the scorecard to the Data Owner at least quarterly. A sample template is available at Agency Resources.

EIA-233: The Data Steward shall maintain a risk log of all data issues and ensure successful remediation or escalation of issues. A sample template is available at Agency Resources.

Collaboration

EIA-234: The Data Steward shall coordinate and resolve technical stewardship issues for standardized data to ensure a cohesive data management strategy.

EIA-235: The Data Steward shall collaborate with data owners and system owners to ensure data needs are met and acts as a liaison between data owners and users.

Requirements for the role of Data Custodian

EIA-300: The Data Custodian shall implement and manage data storage solutions to ensure data is stored securely and is readily accessible as needed.

EIA-301: The Data Custodian shall establish and maintain data backup and recovery procedures to prevent data loss and ensure data can be restored in case of a system failure.

EIA-302: The Data Custodian shall monitor and manage the performance of applications and corresponding data storage systems to ensure they operate efficiently and with appropriate safeguards.

EIA-303: The Data Custodian shall implement security controls to protect data from unauthorized access, use, alteration, or destruction, following the guidelines provided by Data Owners and Data Stewards.

EIA-304: The Data Custodian shall ensure that data resides on storage solutions and configurations that comply with the Data Availability Standards.

EIA-305: The Data Custodian shall assist with incident response, work to resolve data-related incidents, and provide support for data recovery and restoration.

EIA-306: The Data Custodian shall maintain detailed records of data storage locations, backup schedules, and recovery procedures.

EIA-307: The Data Custodian shall collaborate with Data Owners and Data Stewards to ensure data is classified correctly and stored according to its sensitivity and importance.

EIA-308: The Data Custodian shall ensure that data storage solutions are scalable to handle increasing data volumes.

EIA-309: The Data Custodian shall conduct regular reviews and updates of data storage and protection practices to incorporate new technologies and best practices.

EIA-310: The Data Custodian shall implement and manage data encryption solutions to protect data at rest and in transit, ensuring compliance with security policies and regulations.

EIA-311: The Data Custodian shall complete the role-based training required in Sec 527 – Cybersecurity Awareness Training Standard.

Metadata Requirements

EIA-400: The COV Metadata Repository has been established within the Office of Data Governance and Analytics to aggregate metadata definitions created by Data Owners and Data Stewards.

EIA-401: The Office of Data Governance and Analytics shall ensure the repository is accessible to authorized users.

EIA-402: Agencies that have their own metadata repository shall export that data and share it with the Office of Data Governance and Analytics.

EIA-403: The COV Metadata Repository shall capture and store comprehensive metadata for all critical COV data assets, including data quality, format, lineage, structure, and access controls.

EIA-404: The COV Metadata Repository shall utilize a Data Taxonomy (Appendix III: Data Taxonomy for Metadata Collection) to provide a consistent and standardized approach to representing metadata across agencies.

EIA-405: Data Stewards shall document the following metadata elements at a minimum for every dataset they manage:

- Dataset Name and Description
- Classification levels (e.g., public, internal, confidential, highly confidential)
- Data Owners and Data Stewards
- Data Source/Origin
- Frequency of updates
- Retention periods
- Quality level metrics

EIA-406: The Office of Data Governance and Analytics shall implement processes for updating and maintaining metadata.

Data Quality Management Requirements

EIA-500: Data quality management processes shall be incorporated to monitor, measure, and improve the quality of data exchanged between systems, ensuring that data remains accurate, complete, and consistent across integrated systems.

EIA-501: The Data Steward shall establish a data quality baseline based on the following properties: accuracy, completeness, consistency, relevancy, validity, timeliness, and uniformity.

EIA-502: Agencies shall use Data Quality Tiers to categorize their data quality.

EIA-503: The data discovery and exploration process shall identify and report on any data quality issues, aiming to address these issues to ensure data quality is maintained over time.

EIA-504: Data Stewards shall set specific data quality metrics and targets for COV Critical Data Assets (e.g., achieving 95% accuracy or meeting Tier 1 or Tier 2 data quality standards for customer contact information).

EIA-505: Data Stewards shall maintain and improve, as necessary, the quality of associated COV Critical Data Assets.

EIA-506: Agencies shall follow standard processes for assessing, maintaining, and reporting high-quality data (see properties above) that is fit for intended uses in operations, decision-making, and planning.

EIA-507: Data lineage tracking tools shall be used to record the origin of a data set, all transformations applied, and the final destination to enable transparency and traceability for regulatory compliance and data quality assurance.

EIA-508: Agencies shall evaluate data assets against data quality tiers (Appendix II) and record them in the metadata repository.

Data Protection

EIA-601: The Data Custodian must work with the ISO and Data Stewards to implement the principle of least privilege, granting data access only to those who need it for their job functions.

EIA-602: The Data Custodian must work with the ISO to use strong authentication and authorization mechanisms to control access to data.

EIA-603: The Data Custodian must work with the ISO and Data Stewards to encrypt confidential and regulatory data at rest and in transit.

EIA-604: The Data Custodian must use approved encryption algorithms and key management practices as described in Sec 530.

EIA-605: The Data Custodian must work with the ISO and application developers to ensure no sensitive data is stored in development or test environments.

EIA-606: File level encryption must be established for all cloud and on-premises databases with sensitive data to provide encryption at rest.

EIA-607: Any tables that contain PII must be encrypted and columns with Protected PII obfuscated through encryption or masking.

EIA-608: All encryption keys must be stored in secure storage such as Azure Key Vaults or a Hardware Security Module (HSM).

EIA-609: Network traffic between agencies must use end to end encryption and traffic must be separated from other Agencies. Network traffic that does not need to be securely transmitted may use the public internet.

EIA-610: Agency must adhere to all security controls in Sec 530 to protect data including, but not limited to, network firewalls, intrusion detection, endpoint security etc.

EIA-611: All new database accounts, permission changes, and group memberships must be logged and require management or Data Steward approval.

EIA-612: Database logging must be enabled to include user database login sessions, and the logs must be kept for at least 90 days or as required by Sec 530.

Data Governance Framework

EIA-700: The Office of Data Governance and Analytics shall establish an Executive Data Board, Data Governance Council, and Data Stewards Group as defined in § 2.2-203.2:4. Office of Data Governance and Analytics; Chief Data Officer; creation; report.

EIA-701: The Executive Data Board, Data Governance Council, and Data Stewards Group shall adhere to the requirements of their respective charters including membership, quorum, and meeting frequency.

EIA-702: Agency Heads shall establish the appropriate agency data governance framework, including an agency data governance council, as required by their agency size defined below.

	Small Agency (<100 FTE's)	Medium Agency (100-1000 FTE's)	Large Agency (1000+ FTE's)
Data Owner	Usually, the agency or department head. Resourced part-time alongside other responsibilities.	May be multiple data owners assigned by functional area May include a dedicated percentage of their time for data-related duties.	High-ranking officials supported by a dedicated governance budget and resources. May include a dedicated percentage of their time for data-related duties
Data Steward	Limited to key operational areas (e.g., finance, HR). Typically, an additional responsibility for existing staff.	Dedicated or semi-dedicated roles in high-priority areas. Cross-functional collaboration among stewards.	Full-time roles in all major business units or programs. Specialized stewards for data quality, metadata management, and privacy.
Governance Committee Structure	5-7 members: Data Owners and Stewards, IT lead, program managers.	8-12 members: Department managers, CIO, legal/compliance officer, operational representatives.	12-20 members: Executives, data officers, domain stewards. May have subcommittees for specific governance domains.
Governance Council Meetings	Informal, ad hoc	Formal, quarterly	Formal, monthly
ODGA Involvement	Advisory and Coordinative • Provide guidance on implementing foundational data governance policies. • Serve as a shared resource for expertise, tools, and training. • Facilitate compliance with statewide data standards and regulations. • Act as a liaison between the agency	Strategic and Operational Partner • Support the establishment of formal data governance roles (e.g., owners, stewards). • Assist with developing tailored governance frameworks, including data catalogs and access controls. • Provide shared services (e.g., data analytics platforms)	Oversight and Enabler • Focus on policy alignment, ensuring the agency's robust governance practices adhere to statewide standards. • Offer advanced analytics and governance tools for scalability. • Collaborate on enterprise-level initiatives (e.g., cross-agency data integration). • Monitor agency performance in data

	Small Agency (<100 FTE's)	Medium Agency (100-1000 FTE's)	Large Agency (1000+ FTE's)
	and central governance bodies.	to supplement agency resources. • Drive alignment with statewide initiatives and facilitate inter-agency data sharing.	governance and provide resources for continuous improvement.
Resources	Shared Resources Leverage existing staff across multiple roles (e.g., IT, legal, program areas). Use external consultants as needed.	Hybrid Resources (where budget allows) Dedicated data governance roles for critical areas. Shared positions for less intensive roles.	Dedicated Resources (where budget allows) Full-time dedicated teams for data governance, including technical and strategic roles.

EIA-703: Agency heads shall ensure employees are provided with training on data literacy and data governance principles, policies, and procedures.

EIA-704: Agency heads shall ensure that Data Owners, Data Stewards, and Data Custodians have the necessary skills and knowledge to perform their roles effectively.

Data Classification

Purpose
The purpose is to establish a framework for identifying information based on the characteristics of the data. Identifying the type of data ensures that information is handled in a manner that maintains the target risk posture of the organization and complies with legal, regulatory, and business requirements. Using this framework will help users identify what data, documents and other information needs to be restricted.
Data classification and sensitivity classification are two separate steps. Data classification will identify the data type in use and sensitivity is defined based on the impact to confidentiality, integrity and availability. Keeping the classification items separate provides the ability to identify datasets that when combined may become sensitive.

Scope
This standard applies to all employees, contractors, consultants, and third-party partners who create, manage, store, transmit, or access organizational data. The standard governs data in all formats (digital, paper, verbal, etc.).

Classification Governance & Roles

All classification decisions and controls fall under existing EIA Governance. Key actors:

Data Owner: Defines classification label, approves sensitivity, and authorizes access.

Data Steward: Ensures data is labeled, inventories are maintained, and policy changes are communicated.

Data Custodian: Implements technical and procedural controls required by classification.

Classification Criteria

Each dataset must be evaluated against the following criteria:

Confidentiality: Impact of unauthorized disclosure (e.g., legal penalties, reputational harm).

Legal/Regulatory: Whether subject to FTI, HIPAA, PCI-DSS, FERPA, etc.

Business Impact: Financial, operational, or reputational loss if data is lost or exposed.

Access Needs: The determination among employees and third parties that legitimately require access to data.

Classification Labels

Public
Definition: Freely available to the general public; no restrictions on sharing.
Risk: Negligible risk to individuals or organizations if accessed or disclosed
Controls: None, available through unrestricted channels such as government websites, public databases, or open data portals.
Information Handling: Freely shared with the general public and without violating laws or regulations related to data privacy or security.

Internal Use Only
Definition: For State personnel and approved affiliates.
Risk: Reputation or efficiency loss if leaked.
Controls: Authentication, usage monitoring, documented sharing approvals.
Information Handling: Access to this data must follow data handling and storage policies. Data sharing authorization must be documented prior to providing the data to an identified party. Technology and mediums used to share data must be authorized prior to providing the data to an identified party.

Personal
Definition: Information generated or stored by personnel that is not directly related to the Commonwealth's operations.
NOTE: This data will be accessible and disclosable by the Commonwealth.
Risk: Disclosure of data included with this classification should not impact the risk posture of the Commonwealth in any way. Any data that does impact the risk posture of the Commonwealth cannot be classified as Personal.
Controls: Organizations retain the right to access/disclose; should not be shared with anyone in the organization using Commonwealth systems or data storage.
Information Handling: Personal data should not be shared with anyone in the organization using Commonwealth systems or data storage.

Confidential
Definition: Data that includes information that should only be accessible to authorized individuals.
Risk:

1. Unauthorized access or exposure could lead to significant consequences for the organization, its employees, citizens, or customers.
2. Unauthorized disclosure could have moderate to severe adverse effects on the organization.
3. Exposure could result in financial penalties, damage to reputation, violation of regulatory terms, violation of the law or potentially result in legal action.

Controls: Organization must maintain documentation, and access logs for roles or users accessing this data.

Information Handling:

1. Access to this data must follow data handling and storage policies.
2. Data sharing authorization must be documented prior to providing the data to an identified party.
3. Technology and mediums used to share data must be authorized prior to providing the data to an identified party.
4. Organizations must document authorized locations to store this data.

AI Training Data

Definition: Any structured or unstructured data identified as permitted for training, validating, or testing artificial intelligence models. This data may include, but is not limited to, text, images, video, sensor data, or any other data type used to develop and refine learning algorithms.

Risk:

1. Unintentional data leakage or exposure of data generated from the trained model.
2. Creation of bias or inaccuracy from the model due to data that is not suitable for training or learning.
3. The resulting model may end up with security vulnerabilities.

Controls: Separate access accounts per model/application

Information Handling: Prior to usage the agency must document a policy supporting the data is eligible for usage in AI systems and training. Data owner must document the approval of usage of the data in the AI system, including understanding once the data is used they may not be able to remove it from the trained system.

Regulatory Labels

Applied in addition to primary classification labels when datasets fall within specific legal or regulatory regimes.

Personally Identifiable Information (PII)

Personal information means the first name or first initial and last name in combination with and linked to any one or more of the following data elements that relate to a resident of the Commonwealth, when the data elements are neither encrypted nor redacted:

1. Social security number
2. Driver’s license number or state identification card number issued in lieu of a driver’s license number
3. Financial account number, or credit card or debit card number, in combination with any required security code, access code, or password that would permit access to a resident’s financial accounts.

Personal Medical Information (PMI)

Definition: Medical information means the first name or first initial and last name in combination with and linked to any one or more of the following data elements that relate to a resident of the Commonwealth:

1. Any information regarding an individual's medical or mental health history, mental or physical condition, or medical treatment or diagnosis by a health care professional
2. An individual's health insurance policy number or subscriber identification number, any unique identifier used by a health insurer to identify the individual, or any information in an individual's application and claims history, including any appeals records.

Protected Health Information (PHI)
Definition: Individually identifiable health information transmitted by electronic media, maintained in electronic media or transmitted or maintained in any other form or medium by a Covered Component (HIPAA applies to Covered Components, which are health care providers, health plans, and clearinghouses that engage in certain types of transactions electronically). PHI is considered individually identifiable if it contains one or more of the following identifiers:

- Any information regarding an individual's medical or mental health history, mental or physical condition, or medical treatment or diagnosis by a health care professional.
- An individual's health insurance policy number or subscriber identification number, any unique identifier used by a health insurer to identify the individual, or any information in an individual's application and claims history, including any appeals records.

Family Educational Right and Privacy Information (FERPA)
Definition: FERPA is a federal law that affords parents the right to have access to their children's education records, the right to seek to have the records amended, and the right to have some control over the disclosure of personally identifiable information from the education records. When a student turns 18 years old, or enters a postsecondary institution at any age, the rights under FERPA transfer from the parents to the student ("eligible student"). The FERPA statute is found at 20 U.S.C. § 1232g and the FERPA regulations are found at 34 CFR Part 99.PCI

Federal Tax Information (FTI)
Definition: FTI is defined as any return, return information or taxpayer return information that is entrusted to the commonwealth by the Internal Revenue Services. Federal law provides that all returns and return information are confidential. No current or former employee of the IRS, state or federal agency may access or disclose returns or return information unless specifically authorized under provisions of the Code. A return means any tax or information return, estimated tax declaration or refund claim (including amendments, supplements, supporting schedules, attachments or lists) required by or permitted under the Code and filed with the IRS by, on behalf of, or with respect to any person.

Social Security Administration Information (SSA)
Definition: Data subject to the social security administration data exchange agreement.

Payment Card Information (PCI)
Definition: Information printed or stored on a physical card. PCI covers all cardholder data, including Primary account number (PAN), Cardholder's name, Card expiration date, and Security code.

Critical Infrastructure (CI)
Definition: Any information system that controls, processes, transmits, receives or stores electronic information in any form including data, voice, or video that is either vital to the functioning of critical infrastructure of the commonwealth or the United States or so vital that the incapacity or destruction of such systems would have a debilitating impact on commonwealth or national security, economic security, or public health and safety.

Law Enforcement (LE) Data
Definition: Information that is essential to the law enforcement mission. This could be any criminal investigative data that may jeopardize an ongoing investigation or prosecution; jeopardize the safety of an individual; or cause a suspect to flee or evade detection.

Supervisory Control and Data Acquisition (SCADA) Data
Definition: Data in systems and networks that monitor, manage, and control automation, production and distribution in industrial environments or equipment used to support physical environments.

Intellectual Property (IP) Data
Definition: Data such as inventions, artistic works, designs, symbols, names, and images used in commerce, which are protected by law allowing the owner exclusive rights to the data.

Attorney-Client Privileged Data
Definition: Data that is subject to attorney-client privilege is considered confidential. This includes any written advice of legal counsel to state, regional or local public bodies or the officers or employees of such public bodies, and any other records protected by the attorney-client privilege. In addition, any data that may be considered part of any attorney work product is also confidential. This includes legal memoranda and other work product compiled specifically for use in litigation or for use in an active administrative investigation.

Sensitivity Labels

Once data is classified with the classification label the data should be evaluated for sensitivity. Data should be classified as sensitive relative to confidentiality, integrity and/or availability. All data identified as sensitive should also apply the appropriate Sensitive Data Label.

Sensitive – Confidential
Definition: Any data that must be protected from access by one or more authorized parties

Sensitive – Integrity
Definition: Any data that must be protected from unauthorized modification

Sensitive – Availability
Definition: Any data that has a target recovery time objective of 8 hours or less or a recovery point objective of 4 hours or less

Review and Updates:
Data classification should be reviewed annually and updated based on changes in legal, regulatory, or business needs. Security audits must include verification of compliance for sensitive systems hosting data that has been classified.

Definitions and Terminology

As appropriate, terms and definitions used in this document can be found in the COV ITRM IT Glossary.

Accessibility: A measure of how easily data can be accessed, retrieved, and used when needed.

Accuracy: A measure of how well data represents reality or a given standard.

Artificial Intelligence (AI): Refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions.

COV Critical Data Asset: Any data set considered essential to the successful fulfillment of the missions of one or more COV agencies. These critical data assets are identified by the designated Data Owner in each agency, in collaboration with the Data Steward and may be structured (e.g. databases) or unstructured (e.g. Excel, PDF's)

Completeness: Indicates whether all necessary data is present.

Consistency: Ensures that data does not have discrepancies when compared across different data sources or systems.

Data Asset: Any collection of data recognized as valuable for decision-making, operations, and strategic planning within an organization. This includes structured data, such as databases and spreadsheets, and unstructured data, such as documents and multimedia files. Data assets must be managed and maintained throughout their lifecycle to ensure accuracy, completeness, and usability.

Data Classification: The process of organizing data into relevant groups ("classes") based on shared characteristics, such as sensitivity, risk, and applicable compliance regulations.

Data Cleansing: The process of identifying and correcting (or removing) corrupt or inaccurate data from a dataset.

Data Catalog: A centralized repository or platform that organizes, categorizes, and indexes metadata and information about data assets. It serves as an inventory, providing insights into the structure, lineage, quality, and usage of data, facilitating efficient data management, discovery, and governance for COV.

Data Custodian: An individual or organization in possession of data on behalf of Data Owners, responsible for protecting the data from unauthorized access, alteration, destruction, or use, and administering general controls such as backup and recovery.

Data Discovery: The process of identifying and understanding all data assets, including relationships and dependencies, to enable informed decisions about data management and governance.

Data Domain: The logical grouping of data that shares a common theme or purpose. A way to categorize and organize data into meaningful units, making it easier to manage, understand, and analyze. Examples include patient, provider, claims, student, staff, crime, officer, vehicle, air quality etc.

Data Integrity: The assurance of the accuracy, completeness, and consistency of data throughout its lifecycle, from creation to consumption.

Data Lineage: The flow of data through its lifecycle, from creation to consumption, enabling traceability and governance.

Data Owner: An individual responsible for defining, managing, and controlling data use while ensuring compliance with standards within an agency. The Agency Head or designee assigns the Agency Data Owner(s) for functional/subject areas under their jurisdictional control or authority and provides adequate resources to develop and maintain these areas in support of the Commonwealth's Data Management Program. Data Owners are typically senior business leaders accountable for data and usually have budgetary responsibility.

Data Steward: An individual designated by an agency to represent interagency data needs, ensuring that proposed standards meet those needs. Agency Data Stewards work on behalf of their Agency Data Owners and should understand the agency's data, research data usage, have the authority to obtain agreement from Data Owners, and address data issues for the agency. Data Stewards are typically in the functional areas, not in Information Technology (IT).

Data Quality: The measure of how well data meets criteria for accuracy, completeness, consistency, validity, uniqueness, timeliness, and fitness for purpose.

Granularity: The level of detail or precision of the data.

Integrity: Refers to the overall correctness and coherence of data, ensuring no contradictions within related data sets and maintaining relationships between data elements.

Master Data: Represents "data about business entities that provide context for business transactions." Common categories of master data include parties (individuals, organizations, roles), products, financial structures (such as ledgers and cost centers), and locations.

Master Data Domain: Categories of master data that define and group related foundational information essential for agency operations, such as customers, products, suppliers, and assets.

Reliability: Refers to the trustworthiness and dependability of the data.

Reference Data: Data used to classify or categorize other data. Typically, reference data is static or changes slowly over time, including items like COV-specific units such as ethnicity, agencies and departments, budget codes, disease or procedure codes, school districts. External sources that reference standards should be used when possible.

Relevance: A measure of how pertinent the data is to a specific purpose or context.

Traceability: The ability to trace data back to its origins, which is critical for auditing and verifying data sources.

Timeliness: A measure of how current or up-to-date data is.

Uniqueness: Ensures that there are no unnecessary duplicate records in the database.

Validity: Refers to data that conforms to a specific format, range, or pattern.

Appendix I: RACI

Key

- **R** = Responsible (does the work)
- **A** = Accountable (final authority and approval)
- **C** = Consulted (provides input)
- **I** = Informed (kept updated)

Activity	Data Owner	Data Steward	Data Custodian	System Administrator
Data Classification & Sensitivity Designation	A	R	I	I
Business Data Definitions	A	R	C	I
Data Quality Standards	A	R	C	I
Data Access Policy Development	A/R	C	C	I
Approving Data Access Requests	A/R	R	I	I
Implementing Access Controls on Data	I	I	A/R	R
Data Quality Monitoring	R	A/R	C	I
Data Quality Issue Resolution	R	A/R	C	I
Master Data Management	A	R	C	I
Metadata Management	C	A/R	C	I
Database Backups & Recovery	A	I	C	R
System Performance & Availability	I	I	R	A/R
Data Retention & Disposal Schedules	A	R	C	I
Executing Data Retention & Disposal	I	C	A	R
Privacy Impact Assessments	A	R	C	I
Data Breach Response (Business Impact)	A	R	C	I
Data Breach Response (Technical)	C	C	C	A/R
Regulatory Compliance (Business)	A	R	I	I
Data Integration & ETL Processes	C	C	R	A
Vendor Data Sharing Agreements	A	R	I	I

Notes

- Agencies should adapt this matrix to their specific organizational structure and needs
- Some smaller agencies may combine roles (e.g., Data Custodian and System Administrator)
- For enterprise-wide data, the Data Owner role may reside at the agency head or deputy level
- Regular review and updates to this RACI matrix help maintain clarity as programs mature

Appendix II: Data Quality Tiers

Tier 1 (Highest Quality)
Definition: Data that meets the highest standards of accuracy, consistency, and completeness.

Characteristics:

- 99-100% accuracy and integrity.

- Fully consistent with enterprise-wide standards and definitions.
- Complete with negligible missing or null values.
- Current and timely.

Tier 2
Definition: Data that is largely accurate but may have minor inconsistencies or occasional missing values.

Characteristics:
- 95-98% accuracy and integrity.
- Mostly consistent with minor deviations from enterprise-wide standards.
- Few missing or null values.
- Generally, up to date.

Tier 3
Definition: Data that is adequate for many use cases but may have notable defects and inconsistencies.

Characteristics:
- 85-94% accuracy and integrity.
- Moderate inconsistencies observed.
- Some missing or null values that might impact certain analyses.
- Slight delay in data freshness or timeliness.

Tier 4 (Lowest Quality)
Definition: Data that requires significant cleansing and validation before it can be reliably used for decision-making.

Characteristics:
- Below 85% accuracy and integrity.
- High levels of inconsistency and deviation from enterprise standards.
- Significant amounts of missing or null values.
- Data might be outdated or not timely.

Appendix III: Data Taxonomy for Metadata Collection

Metadata Element	Description
Data Domain & Classification	• Broad categories or types of data (e.g., customer, product, financial, employee). • Classification levels (e.g., public, confidential, restricted).
Data Entity & Attributes	• Specific tables, objects, or datasets within a domain. • Individual fields or columns within an entity, including data types, lengths, and constraints.
Data Source & Origin	• Where the data originates (e.g., CRM system, ERP system, external vendor). • Physical or logical location (e.g., server, database, schema).
Data Relationships	• How different entities and attributes relate to one another (e.g., foreign keys, hierarchies).
Data Usage & Purpose	• The primary use cases or business processes that rely on the data. • Contextual information that clarifies the importance, role of the data, and usage restrictions.

Metadata Element	Description
Data Owners & Stewards	• Individuals or teams responsible for the data's quality, accuracy, and governance. • Points of contact for questions or issues related to the data.
Data Lifecycle & Temporality	• The stages of the data's life, from creation to archiving or deletion. • Frequency of updates, retention periods, and archival rules.
Data Quality & Integrity Metrics	• Quality level metric and Tier (e.g., completeness, accuracy, consistency). • Known issues, anomalies, or constraints. (Data Quality Tier)
Data Security & Compliance	• Security measures in place (e.g., encryption, access controls). • Compliance requirements and their implications for the data. • Data Classification (Sensitive / Non-Sensitive)
Data Lineage & Transformation	• The progress of data through systems, including any transformations, aggregations, or cleansing it undergoes.
Business Terminology & Glossary	• Commonly used business terms and their definitions. • Relationships between business terms and technical data entities or attributes.
Reference & Master Data	• Lists or sets of data values that remain relatively static and are referenced by various systems (e.g., product/tracking codes, product categories)