Data Management Policy

Effective Date:
January 06, 2025

Category:
Information

Scheduled Review:
March 31, 2026

Supersedes:
ITP-INF000, ITP-INF001, ITP-INF003, ITP-INF004

1. Authority

Executive Order 2016-06, Enterprise Information Technology Governance

2. Purpose

This Information Technology Policy (ITP) establishes enterprise-wide policy and guidance for the management of Commonwealth data and information. The policy also provides guidance related to data migrations, including those needed to support the GAAP Audit.

3. Scope

This policy applies to all offices, departments, boards, commissions, and councils under the Governor's jurisdiction and any other entity connecting to the Commonwealth Network (hereinafter referred to as "agencies").

Third-party vendors, licensors, contractors, or suppliers shall meet the policy requirements of this policy as outlined herein.

4. Policy

For definitions found within this document, refer to the *IT Policy Glossary*.

4.1 Data and Information Management

Data and information are valuable and strategic assets to the Commonwealth of Pennsylvania, its business partners, and the public. In order to ensure that agencies are taking full advantage of their information resources, it is imperative for agencies to manage data and Information as assets throughout its life cycle.

Agencies shall read and comply with this ITP, which will provide the agencies with specific information, guidelines and best practices to help institutionalize the principles of effective

information management at each stage of the Information Life Cycle.

4.1.1 Data and Information Life Cycle

Agencies shall consider, at each stage of the data and Information Life Cycle, the effects of decisions and actions on other stages of the life cycle. Reference *Records Management Policy* for additional information regarding the life cycle.

Agencies shall design new information collection and creation efforts so that the information collected or created supports downstream interoperability between information systems, as appropriate, without the need for costly retrofitting. This includes consideration and consultation of key target audiences for the information when determining format, frequency of update, and other information management decisions. Specifically, agencies should incorporate the following components into future information collection and creation efforts:

4.1.1.1 Data Governance Structure

Data governance must be aligned with the design and development tasks within the organization's system development life cycle. This permeates the lifetime of data management from the analysis and synthesis of data consumer requirements through conceptual modeling, logical and physical design, and subsequent implementation.

Agencies shall establish a data governance structure that develops an agency Master Data Management Plan which captures the governance operating model, the data processes, roadmaps, and various data management methodologies. The governance structure responsibilities also include drafting policies, proposing policies at both agency and enterprise levels, and implementing and adopting these policies into the agency's business and technical environments.

The Master Data Management Plan should address the following components:

Data integration and sharing capabilities: Increasing reliability, performance, and access to data sources as well as adopting a standard model for Data Exchanges (infrastructure, software, methodologies) ensures successful internal and external sharing of data, which will provide value to the enterprise by potentially reducing cost and other maintenance resources.

Metadata management. The modern digital enterprise information management must enable business-oriented Metadata management, including procedures, processes, and tools for:

- Business term glossaries to capture frequently used business terms and their authoritative definition(s).
- Data standards such as naming conventions, defined reference data sets, and standards for storage and exchange. (Metadata standards and specifications must be reviewed for compliance with the common core Metadata standards, specifications, and formats developed within different communities (e.g., financial, health, geospatial, law enforcement)).
- Data element definitions that reflect the connection to business terms and provide

context-relevant definitions for use within business applications.

- Data lineage that shows the relationships between data element concepts and their representation across different models and applications.
- Integration with data governance policies to support validation, compliance, and control.
- Protection of the Metadata for a record from unauthorized deletion or alteration.
- Procedures for retention or deleting in accordance with the record's retention schedule.
- Migration and validation strategies for Metadata to ensure preservation and integrity during the transfer of Metadata and records.
- Metadata associated with a record is to be categorized in at least one of the following:
  - Identity - information identifying the record.
  - Description - information determining the nature of the record.
  - Use - information facilitating immediate and longer-term record use.
  - Event plan - information used to manage the record, such as disposition information.
  - Event history - information recording past events on the record and its Metadata
  - Relation - information describing the relationship between the record and other records.

Records Systems shall define Metadata to:

- enable the identification and retrieval of records,
- associate records with changing business rules, policies, and mandates,
- associate records with agents, and their authorizations and rights with regards to the records,
- associate records with their business activities,
- track processes carried out on records, and
- track relationships with other records. The Records System should provide the ability to filter and sort report data based on the values contained in any field or column.

Metadata shall be utilized to detail information about a record and should capture the following:

- the description of the content of the record;
- the structure of the record (form, format, and relationships between record components);
- the business context in which the record was created, including the author and date of creation;
- relationships with other records and Metadata;
- a unique identifier and other information needed to retrieve the record; and
- the business actions and events involving the record throughout its existence.

Once the record has been captured, the associated Metadata must be fixed and kept as transactional evidence.

File formats: The file format in which organizational entities keep their data is a primary factor in providing the ability to properly integrate, share, and use data sets

across organizational boundaries. To facilitate data integration and sharing, agencies shall establish policies, standards, and procedures for file organization and formats to include/address the following elements:

- File Version Control;
- Directory Structure/File Naming Conventions;
- File Naming Conventions for Specific Disciplines;
- File Structure;
- File change tracking and logs;
- Use Same Structure for Backups;
- Hardware and hardware obsolescence;
- Short-term and long-term storage file formats; and
- Structure for Backups & Recover.

4.1.1.2 Data Inventory and Classification Procedures

Agencies shall maintain a complete up-to-date inventory of all Data Exchanges between systems (with internal and external entities) and data systems used to store and process data. The data inventory should specify what data elements are collected, origin or source of the data elements, justification for their collection, explanation of the intended purposes for use, and the data elements classification by type and their sensitivity levels. Agencies must update their enterprise data inventory, accounting for datasets used in the agency's information systems. If the inventory does not already exist, it can be built out over time, with the goal of including all agency datasets, to the extent practicable. Refer to the *Data Classification Policy* guidance.

NOTE: Records can represent more than one business activity and therefore can be assigned to more than one record category.

4.1.1.3 Ensure Data Integrity and Integration

Agencies shall identify strategies for ensuring the accuracy and quality of data as well as preventing, detecting, and correcting errors and misuses of data. Agencies shall establish data quality standards to ensure that the data is accurate, relevant, timely, and complete for the purposes it is intended to be used, and maintains an appropriate balance between privacy and security.

Periodic quality audits shall be integrated into all cycles of data managements (e.g., collection, reporting, and release).

The key to providing and sharing data between systems is to determine packaging (files, transactions, data streams, etc.), content and formatting (values, formats, etc.) and package Metadata details (location, origin, availability, changes, etc.). This component also defines methods to allow data availability, such as support for internally and externally delivered content, production support and change control (errors, fixes, versions, etc.) and interface and access to allow data delivery.

To improve data integration and sharing, these fundamental capabilities shall be

established:

Data Accessibility: A vital aspect of data integration is accessibility, and the information management framework must provide connectors to that wide variety of data sources, including file-based, tree-structured data sets, relational databases, and even streamed data sources.

Data transformation, exchange, and delivery: Agencies shall establish controls, mechanisms, procedures, and integration frameworks for data sets to be accessed from their original sources, and efficiently move the data from source to target destinations in an effective and secure manor. There must be a capability to transform the data from its original format into one that is suited to the target, with a means of verifying that the data sets are appropriately packaged and delivered securely. Controls and tests should be performed that data was transformed accurately and completely (e.g. before and after snapshots of record counts/dollar amounts, control totals).

Data Retention Policy: Per MD 210.5 The Commonwealth of Pennsylvania State Records Management Program, all Commonwealth employees are to manage records under their care and control. Each agency has a records coordinator to assist in getting paper, electronic and mixed media records scheduled in their agency specific records retention and disposition schedule. Agencies must control any proposed deletion of records pursuant to existing Agency Specific and Commonwealth General Records Retention and Disposition Schedules.

Ownership and Privacy: Agencies shall ensure that due care and diligence is exercised to make sure that they have considered the implications of sharing data, in terms of copyright, intellectual property ownership, and ethical requirements such as privacy and confidentiality.

4.1.1.4 Data Quality Management

A record is considered reliable if it ensures a full and accurate representation of the transactions or activities. A record is considered to have integrity if it is complete and unaltered. A record must be able to provide adequate and proper documentation of agency business for as long as the information is needed.

Agencies shall institute best practices for data quality management to ensure reliability and integrity. In addition, data quality management practices are to be used to improve the precision of identifying data flaws and errors as well as simplify the analysis and remediation of root causes of data flaws. Leveraging data quality tools and techniques to support the ability to standardize and potentially correct data when possible, flag issues when they are identified, notify the appropriate data steward, and facilitate the communication of potential data issues to the source data providers. These objectives can be met within a formal framework for data quality management that incorporates techniques for:

- Data parsing and standardization: Scanning data values with the intent of transforming non-standard representations into standard formats.
- Data correction and cleansing: Applying data quality rules to correct recognized data

errors as a way of cleansing the data and eliminating inconsistencies.

- Data quality rules management: Centrally manage data quality requirements and rules for validation and verification of compliance with data expectations.
- Data quality measurement and reporting: Provide a framework for invoking services to validate data against data rules and report anomalies and data flaws.
- Standardized data integration validation: Continual validation of existing data integration processes and embedded verification of newly developed data integration processes.
- Data quality assessment: Source data assessments (e.g., real time, generated, or compiled) and evaluation of data issues to identify potential data quality rules using data profiling and other statistical tools.
- Incident management: Standardized approaches to data quality incident management (reporting, analysis/evaluation, prioritization, remediation, tracking).

4.1.1.5 Data Storage and Retrieval

Agencies should identify the useful life of their data and establish policies, procedures, and governance frameworks for the proper storage and retrieval of their data whether it be short term or long-term preservation of their data. Agency data owners and stewards should consult with records coordinators, business owners, legal, privacy officer, security, and other stakeholders to determine the useful life, retention schedule, access frequency, security, and appropriate storage, retrieval, backup, and recovery requirements of their data. Key elements that should be considered in this area are:

- Useful Life of Data and Information.
- Criticality and sensitivity of data and information.
- Records Management and Retention Schedules.
- Hardware and hardware obsolescence.
- Backup Medium, Systems, and Schedules.
- Backup integrity/validation checks.
- Continuity of Government and Disaster Recovery.
- Security controls and safeguards
- Storage Facilities (e.g., Geographical Locations, Cloud, or On-Premise).
- Contractual Terms & Conditions for third party providers.
- Appropriate environmental conditions to increase the lifespan of media

4.1.1.5 Internal Controls

Agencies shall adhere to the internal control framework outlined in *Management Directive 325.12 Amended, Standards for Enterprise Risk Management in Commonwealth Agencies* in order to develop effective data management processes. This directive establishes policy, responsibilities, and procedures for implementing effective internal control systems within agencies. With this directive, the Commonwealth adopted the U.S. Government Accountability Office's Standards for Internal Control in the Federal Government (often referred to as Green Book).The Federal Internal Control Standards can be found at https://www.gao.gov/greenbook/overview.

4.1.2 Data Migration Procedures and Best Practices

Agencies shall continuously review their records schedules to determine if changes to their use of technology affects the value of the records in question. Ensuring usability of records includes carrying out system upgrades of hardware and software while maintaining the functionality and integrity of the electronic records created in them. This includes ensuring that migration of records addresses non-active electronic records stored off-line.

Agencies shall utilize *Migration Audit Checklist Standard* to facilitate all data migration initiatives. Agencies may add additional criteria to *Migration Audit Checklist Standard* in order to satisfy business requirements, mandates, audits, or other legal requirements. The completed checklist must be retained, utilizing proper records management procedures to satisfy any audit requests.

NOTE: The *Migration Audit Checklist Standard* may have items that do not apply to all Data Migration initiatives. Agencies should determine appropriateness of each checklist item as it relates to their Data Migration initiative on a case-by- case basis.

Original source data must be retained for a minimum of five (5) business days after a successful Data Migration and validated in live production environments. Evidence that data validation procedures were completed and results of, and support for, those validations shall be maintained for a minimum of three (3) years for audit purposes.

Any deletion of original source data after the minimum five (5) business day retention period should be authorized and approved by the business owner of the data, which may be contingent on one or more of the following criteria:

- Business owner concurrence of a successful Data Migration that has been completely validated in live production environments.
- Compliance with business data/records retention schedule and/or policies.
- Deployment strategies to run parallel systems (original and target systems) for extended time periods or business cycles.

NOTE: When migrating records, all metadata must be preserved and accompanied with the transfer.

Agencies shall define methodologies, policies, procedures, tools, and governance frameworks to manage the Data Migrations (i.e. movement of data from an old or legacy system to a new system or information systems modernization) and replication that enable rapid bulk transfers of large data sets securely and accurately.

Agencies shall establish mechanisms to monitor system logs, trigger updates to the target systems as changes happen at the source and record periodic extractions from source systems and loading into data warehouses or backup and recovery systems. When performing Data Migrations, Agencies should have a Data Migration plan that is recommended to include, but not limited to, the elements detained in *Data Migration Planning Procedure*.

4.1.3 Security of Data and Information

Establishing a comprehensive data security management plan with checks, controls, and systems is critical to prevent, identify and mitigate security risks, which will ensure the security of sensitive data (data that carry the risk for harm from an unauthorized or inadvertent disclosure) and PII. Agencies shall regularly review their data security management plans and processes and initiate change management processes as needed to remain up to date as to the latest threats and ensure compliance with security policies and legislation.

Agencies shall regularly monitor and evaluate their records controls including monitoring and reviewing access rights and permission rules for electronic records regularly; these access rights and permission rules should be updated on a regular basis. Any actions changing the level of access, altering the record, or changing the location of the record must be documented and tracked in an audit log.

As agencies consider security-related restrictions for handling data and information, focus should be centered on information confidentiality, integrity, and availability as part of the agency's overall risk management framework. In addition to all Commonwealth IT policies and management directives, agencies should consider referencing and incorporating the National Institute of Standards and Technology (NIST) Federal Information Processing Standards Publication *(FIPS-PUB) 199 Standards for Security Categorization of Federal Information and Information Systems* and *(FIPS-BUS) 200 Minimum Security Requirements for Federal Information and Information Systems*. Agencies should also review the National Strategy for Information Sharing and Safeguarding and NIST guidance on Security and Privacy Controls for Federal Information Systems and Organizations.

Ensuring compliance with security IT policies can be accomplished by clearly specifying all activities related to handling data by data stewards as well as users; defining who can access what data, for what purpose, when, and how; and outlining guidance about the appropriate managerial and user data activities for handling records throughout all stages of the data life cycle, including acquiring, maintaining, using, and archiving or destroying both regular and secure data records with mechanisms for de-identifying PII data in order to protect individual privacy for information systems development and testing purposes.

As a security best practice, Agencies shall:

1. Ensure that once a record has been captured into a records system, all events and actions related to the record by person entities and non-person entities are documented on an on-going basis.
2. Collect or create only data or information necessary for the proper performance of agency functions and business requirements and which has practical utility;
3. Limit the collection or creation of data or information which identifies individuals to that which is legally authorized and necessary for the proper performance of agency functions and business requirements;
4. Limit the sharing of data or information that identifies individuals or contains proprietary information to that which is legally authorized, and impose appropriate

conditions on use where a continuing obligation to ensure the confidentiality of the information exists;

5. Ensure that information is protected commensurate with the risk and magnitude of the harm that would result from the loss, misuse, or unauthorized access to or modification of such information; and
6. Take into account other publicly available information when determining whether particular information should be considered PII.

4.2 Database Management Systems

New and existing application development projects that could benefit from a DBMS shall use one of the current standards as defined in *Data Management Standard*.

New application development projects utilizing a DBMS shall adhere to the operational standards as defined in *Data Management Standard*. Existing production applications are encouraged to adopt these standards as well.

Agencies shall analyze the total cost of ownership when selecting one or more of the current DBMS standards. Total cost of ownership includes, but is not limited to, the cost of licensing of DBMS software, hardware platforms, backup solutions, training of operations, database administration, and development staffs. Agencies shall evaluate the alignment of business and technical requirements against the features and capabilities of current DBMS standards when selecting a DBMS.

Note: This policy applies to Enterprise Class DBMSs only.

4.3 Data Modeling

All new application development projects shall be required to use one of the current standard data modeling software products as defined in *Data Management Standard*. Likewise, all new application development projects shall also adhere to the data modeling standards detailed in *Data Management Standard*. Existing production applications are encouraged to adopt these standards as well. Benefits of Data Modeling

Data Management Guideline presents a summary of the various levels of models that may be developed to define and support the business as well as the perspectives from which these models are used in the development process.

4.3.1 Model Types

There are three basic types of data models: conceptual, logical, and physical. A brief primer on the practice of data modeling and the function of each of the model types is presented in the supplemental document *Data Management Guideline*. Please refer to this document for further details on each model type and explanations of steps

4.3.2 Data Modeling Standards

A list of data modeling standards has been compiled by the Enterprise Data Office (EDO). These standards are presented in *Data Management Standard* document. The list has applicability across all current standard products and is required to be used for all application development

efforts.

A data modeling methodology is currently being developed to provide additional guidance to agencies and will become part of the System Development Methodology. A citizen data model is also being developed and can be found in *Data Management Guideline.*

Defining a Core Citizen Data Model and Data Elements will:

- Assist in delivery of consistent and user-friendly experiences across all digital Services.
- Assist online destinations to have a consistent look and feel to ensure a single identity for Enterprise Services.
- Ensure all citizen information is concise, in plain language, and current.
- Facilitate and govern the ongoing transition of traditional non-digital services to a digital service platform.
- Establish a Citizen-First framework that promotes the innovative spirit and skills of the Enterprise through its personnel and technologies.
- Identify and Improve programs managing data, privacy, risk, and accessibility associated with Citizen Data.

4.4 Data Warehousing

When a business requirement necessitates reporting that summarizes or combines data from multiple sources, Agencies shall consider utilizing data warehouse technology.

Prior to building or implementing a new data warehouse, agencies shall determine if a data warehouse exists that meets their needs. If a data warehouse exists that contains all the required data, agencies shall utilize it rather than implementing a new data warehouse.

Agencies shall leverage the Data Warehousing solution provided by Integrated Enterprise Systems (IES) for Enterprise Resource Planning (ERP) applications or ERP data when the IES Data Warehousing solution meets agencies business requirements.

Agencies shall use one of the current standards for Data Warehousing as defined in *Data Management Standard*. Information regarding the availability and licensing of current Data Warehousing product standards is also available in that document.

Agencies shall determine the level of mission criticality of their data warehouse. The infrastructure and operational procedures necessary to support the data warehouse shall be designed and implemented to the level of mission criticality of the data warehouse.

Agencies shall take appropriate automated and manual measures to continually improve the quality and accuracy of the information in the data warehouse working towards a goal of 95% or greater. Agencies shall maintain metrics regarding the current quality and accuracy of the information in their data warehouses. Data quality and accuracy are critical to establishing the integrity of the information and user confidence in the validity of the resulting output.

Additionally, agencies shall ensure:

- All data warehouse models follow the database standards referenced in this policy and *Data*

Management Standard.

- All data warehouse models follow the Core Citizen Data Model as defined in *Data Management Guideline* for citizen-centric common data elements described in the citizen model.
- Any custom development done for ETL adheres to existing Commonwealth standards.
- Data Warehousing solutions physically operate within Commonwealth-approved environments.
- Any data warehouse contains a subset of information from operational systems optimized for data retrieval and reporting to support performance measurement against agency or enterprise goals and objectives.
- Data warehouses utilize Commonwealth approved IT resources separate from operational and transaction-related systems, thereby mitigating potential performance issues with these systems.
- Any data warehouse has an efficient extracting or harvesting process separate from operational and transaction-related systems in order to minimize the impacts to the performance or availability of these systems.
- Standard methods such as ANSI-SQL, Open Database Connectivity (ODBC), Java Database Connectivity (JDBC), or OData (Open Data Protocol) RESTful APIs are used to access any data warehouse.
- Training requirements are established and met for each model of the data warehouse (Snowflake and Star Schema) before a user will be granted access to data.
- Access to the data warehouse and the information within the data warehouse shall be based upon each user's job requirements and access level approved by the authorized agency officer and in accordance with *Management Directive 205.34 Amended Commonwealth of Pennsylvania Information Technology Acceptable Use Policy.*
- Data federation is enabled to support horizontal and vertical exchange between agencies and centralized Commonwealth-wide data warehouses utilizing data sharing agreements and role-based access control.

Agencies are encouraged to review the best practices contained in the *Data Management Guideline* document.

5. Contact

Questions or comments may be directed via email to OA, IT Policy.

6. Exception from Policy

In the event an agency chooses to seek an exception from this policy, a request for a policy exception shall be submitted via the enterprise IT policy exception process. Refer to *IT Policy Governance Policy* for guidance.

7. Revision History

This chart contains a history of this publication's revisions. Redline documents outline the revisions and are available to Commonwealth users only during the drafting process.

Version	Date	Purpose of Revision
Original	01/06/2025	Base Document