pennsylvania
OFFICE OF ADMINISTRATION
INFORMATION TECHNOLOGY

Information Technology Policy

Enterprise Data and Information Management Policy

ITP Number ITP-INF000	**Effective Date** July 11, 2016
Category Information	**Supersedes** None
Contact RA-ITCentral@pa.gov	**Scheduled Review** July 2017

1. Purpose

The purpose of this IT Policy is to provide direction for effectively managing data and information lifecycles.

2. Scope

This Information Technology Policy (ITP) applies to all departments, boards, commissions and councils under the Governor's jurisdiction. Agencies not under the Governor's jurisdiction are strongly encouraged to follow this ITP.

3. Definitions

Data: A value or set of values representing a specific concept or concepts. Data become "information" when analyzed and possibly combined with other data in order to extract meaning, and to provide context.

Dataset: An organized collection of data

Data Exchange: Data from a source system that is restructured for the target system for the purpose of accurately representing the source data. Data exchanges rely on implementing data languages such as Extensible Markup Language (XML) and JavaScript Object Notation (JSON).

Data Migration: Utilizing a design for data extraction and data loading for the purpose of permanently relocating data from one system/application to another system/application.

Information: Any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms.

Information Lifecycle: The stages through which information passes, typically characterized as creation or collection, processing, dissemination, use, storage, and disposition.

Master Data Management Plan: An agency-specific document developed by the agency's data/information governance body that documents the governance operating model, data processes (collection, reporting, release), data roadmaps, data acquisition/integration methodologies, and other relevant procedures.

Mosaic Effect: An event in which datasets that pose no disclosure threat by themselves can create a security risk or produce PII when combined with other datasets.

Personally identifiable information (PII): Any information about an individual maintained by an agency, including 1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and 2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.

4. Policy

Data and information are valuable and strategic assets to the Commonwealth of Pennsylvania, its business partners, and the public. In order to ensure that the commonwealth agencies are taking full advantage of their information resources, it is imperative for agencies to manage data and information as assets throughout its lifecycle. Utilizing best practices for data and information lifecycles will increase operational efficiencies, reduce costs, improve services, support mission needs, and safeguard sensitive information.

Agencies are to read and comply with the following sections of 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's lifecycle:

4.1. Data and Information Lifecycle Best Practices

Agencies must consider, at each stage of the data and information lifecycle, the effects of decisions and actions on other stages of the lifecycle. Agencies should 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 Develop a data governance structure

I. Data governance must be aligned with the design and development tasks within the organization's system development lifecycle. 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.

II. Agencies should 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 adoption of these policies into the agency's business and technical environments.

III. The master data management plan should address the following components:

i. 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.

ii. Metadata management. The drive for cross-functional data sharing and exchange exposed the inherent inconsistencies associated with data systems designed, developed, and implemented separately within functional silos. The modern digital enterprise information management environment must enable business-oriented metadata management, including tools and methods for:

1. Business term glossaries to capture frequently-used business terms and their authoritative definition(s)
2. 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 standard, specifications, and formats developed within different communities (e.g., financial, health, geospatial, law enforcement)
3. Data element definitions that reflect the connection to business terms and provide context-relevant definitions for use within business applications
4. Data lineage that shows the relationships between data element concepts and their representation across different models and applications
5. Integration with data governance policies to support validation, compliance, and control

iii. 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 should establish policies, standards, and procedures for file organization and formats to include/address the following elements:

1. File Version Control
2. Directory Structure/File Naming Conventions
3. File Naming Conventions for Specific Disciplines
4. File Structure
5. File change tracking and logs
6. Use Same Structure for Backups
7. Hardware and hardware obsolescence
8. Short-term and long-term storage file formats
9. Structure for Backups & Recover

4.1.2 Develop and maintain data inventory and classification procedures

I. Maintaining 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 ultimate goal of including all agency datasets, to the extent

practicable. Refer to ITP-SEC019 *Policy and Procedures for Protecting Commonwealth Electronic Data* and OPD-SEC019A *Data Categorization and Inventory Operating Template* for data inventory guidance.

4.1.3 Ensure data integrity and integration

I. Identify strategies for ensuring the accuracy and quality of data as well as preventing, detecting, and correcting errors and misuses of data. Establishing data quality standards to ensure that the data are accurate, relevant, timely, and complete for the purposes they are intended to be used with an appropriate balance between privacy and security. Periodic quality audits should be integrated into all cycles of data managements (e.g., collection, reporting, and release).

II. 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 should be established:

i. 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.

ii. Data transformation, exchange, and delivery

- 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, etc.).

iii. Data Retention Policy

- Agencies have a responsibility to maintain data and make that data available for preservation by the commonwealth both as a matter of records retention polices, legislative mandates, and program integrity.

iv. 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.

v. IT Policy Guidance

- Refer to ITP-APP022 *Financial Applications Policy* for additional guidance on financial applications that are a part of the commonwealth's enterprise resource planning (ERP) system.

4.1.4 Data Quality Management

I. Institute best practices for data quality management to improve the precision of identifying data flaws and errors as well as simplify the analysis and remediation of root causes of data flaws. In addition, 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:

i. Data parsing and standardization: Scanning data values with the intent of transforming non-standard representations into standard formats.

ii. Data correction and cleansing: Applying data quality rules to correct recognized data errors as a way of cleansing the data and eliminating inconsistencies.

iii. Data quality rules management: Centrally manage data quality requirements and rules for validation and verification of compliance with data expectations.

iv. Data quality measurement and reporting: Provide a framework for invoking services to validate data against data rules and report anomalies and data flaws.

v. Standardized data integration validation: Continual validation of existing data integration processes and embedded verification of newly-developed data integration processes.

vi. 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.

vii. Incident management: Standardized approaches to data quality incident management (reporting, analysis/evaluation, prioritization, remediation, tracking).

4.1.5 Data Storage and Retrieval

I. Agencies should identify the useful life of their data and establish policies, procedures, and governance frameworks with the mechanisms 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:

i. Useful Life of Data and Information
ii. Criticality and sensitivity of data and information
iii. Records Management and Retention Schedules
iv. Hardware and hardware obsolescence
v. Backup Medium, Systems, and Schedules
vi. Backup integrity/validation checks
vii. Continuity of Government and Disaster Recovery
viii. Security controls and safeguards
ix. Storage Facilities (e.g., Geographical Locations, Cloud, or On-Premise)
x. Contractual Terms & Conditions for third party providers
xi. Appropriate environmental conditions will increase the life-span of media

4.1.6 Implement Effective Internal Controls

I. Agencies must adhere to the internal control framework outlined in Management Directive 325.12 *Standards for Internal Controls in Commonwealth Agencies* in order to develop effective data management processes. This MD establishes policy, responsibilities, and procedures for implementing effective internal control systems within Commonwealth agencies. With this MD, the Commonwealth is adopting the U.S. Government Accountability Office's, Standards for Internal Control in the Federal Government (often referred to as Green Book). Green Book now incorporates updated principles which include control activities over Information Technology to support the achievement of objectives. Specifically, Principle #11 of the Green Book notes that "Management should design the entity's information system and related control activities to achieve objectives and respond to risks."

II. The following is a link to the Federal Internal Control Standards - http://www.gao.gov/assets/670/665712.pdf.

4.2 Data Migration Procedures and Best Practices

Agencies must utilize OPD-INF000A *Migration Audit Checklist Template* to facilitate all data migration initiatives that follow under this policy's definition of Data Migration (see Definitions section). Agencies may add additional criteria to OPD-INF000A in order to satisfy business requirements, mandates, audits, or other legal requirements. The completed OPD-INFO000A must be retained utilizing proper records management procedures to satisfy any audit requests. NOTE: The *Migration Audit Checklist Template* 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 should be maintained for a minimum of three (3) years for audit purposes.

Any deletion of original source data after the minimum five business day retention period is to 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

Agencies should 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. Establish mechanisms to monitor system logs and triggers 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 following elements:

i. Pre-migration Environment Analysis and Impact Assessment

- Evaluate previous data migration initiatives to capture lessons learned, re-use of project artifacts
- Sponsors and key stakeholder involvement (e.g., business and IT communities, and others)
- Evaluate the scope and complexity of the migration initiative
- Access the data types and volume to be migrated
- Analyze the migration approach, options, and activities to be performed
- Scope the data being migrated (remove non-value added, non-required, or historical data)
- Examine the viability, risks, level of effort, resources, timeframe, constraints, security, dependencies, and impacts
- Migration initiative involvement with third-party entities and cross-organizational support
- Skills, competencies, and subject matter expertise needs and availability
- Software, hardware, licensing, tools, systems requirements, capabilities, and availability to support migration relative to both the source and target environments
- Legacy environment decommission requirements, strategy, and approach
- Affiliated contractual, policies, procurements, SLAs, and/or legal considerations
- Cost and funding availability associated with the data migration

ii. Migration Design Blueprint

- Create a high level source to target mapping to identify objects and relationships that will be linked and /or translated during the migration
- Create conceptual/logical/physical models to define the structures, attributes, configurations, and specifications of the source and target environments
- Design a detailed data migration mapping of source to target
- Data backup schedules (i.e., source and target environments)
- Data restoration and recovery procedures

- Design for any interfaces that are required to extract, clean/scrub, and translate the data from your legacy systems or to load the data into the target systems
- Define an executable legacy systems decommission roadmap
- Design the test strategy and plan that outlines the following:
    - What will be tested and by whom
    - Sequence, interdependencies, cycle, and duration of tests to be performed
    - Test environment requirements, configurations, and preparations
    - Test pass/fail criteria
    - Test procedures, pre and post conditions for test scripts and environments
    - Test schedule
    - Test and validation checklists
    - Test results reviews and approvals
    - Retests
    - Define a data quality management plan to include the following elements:
        - Outlining the criteria to ensure and validate the integrity, accuracy, and security of the data before and after data migration
        - Defines processes and procedures for data migration records exception and error handling and resolution procedures
    - Outlines processes and procedures for identifying, categorizing, prioritizing, documenting, and tracking defects, root cause, and corresponding resolution status
    - Establish on-going reviews of predefined task, inspection, testing, and validation checklists to guide, facilitate awareness, and determining data migration status, conditions, risks, issues, and outcomes.
    - Define business and systems operational readiness assessments with final acceptance criteria

iii. <u>Data Migration Planning, Execution, and Validation</u>

Create a formal data migration plan with the following elements:

- Data migration playbook with check lists defining gateposts for "Go" and "No-Go" decision criteria
- Ownership, roles, responsibilities, and resources to perform the data migration tasks
- Integrated quality controls and assurance as defined in the quality plan
- Building, configuration, and testing of the source and target environments and associated tools to support data migration initiative
- Building, testing, and validating the effectiveness and accuracy of mechanisms and utilities used for data extractions, cleaning, conversions, translations, and loading between source and target systems
- Conducting pilots or test runs as required to refine/validate data migration design approach, mappings, testing approach, and estimated durations to complete data migration
- Evaluate and refinement of the data migration test plan
- Backup of original source data for recovery, restoration, and business continuity
- Retention schedule and storage of original source data and systems for validation, recovery, restoration, and business continuity
- Data owner approvals and authorizations for deletion/purging of original source data
- Source and target data sets comparisons (e.g., total number of data elements, size, record counts, dollar amounts, etc.)

- Legacy environment decommissioning roadmap execution
- Executing and monitoring each of the data migration tasks in alignment with predefined playbook
- Communications Plan
- Risk and issues management plan
- Project schedule with key milestones and deliverables
- Conduct and assess business and systems operational readiness assessments in order to make recommendations and determinations for final approval and sign-off
- Post execution monitoring, impact assessments, and reporting for latent data anomalies and/or defects
- Data migration project closeout and lessons learned

4.3 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 personally identifiable information (PII). Agencies should regularly review their data security management plans and processes and initiate change management processes as needed to remain updated to the latest threats and ensuring compliance with security policies and legislation.

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 reference and consider incorporating the National Institute of Standards and Technology (NIST) Federal Information Processing Standard Publication (FIPS-PUB) 199 *Standards for Security Categorization of 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 is 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. Outline guidance about the appropriate managerial and user data activities for handling records throughout all stages of the data lifecycle, 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 should:

- Collect or create only data or information necessary for the proper performance of agency functions and business requirements and which has practical utility;
- 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;
- 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;

- 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
- Take into account other publicly available information when determining whether particular information should be considered PII.

5. Responsibilities

Agencies are recommended to follow the best practices outlined in this policy and are required to:

- Adhere to Management Directive 325.12 *Standards for Internal Controls in Commonwealth Agencies* (Section 4.1.7)
- 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. (Section 4.1.2)
- Follow and complete the OPD-INF000A *Migration Audit Checklist Template* and maintain the complete checklist following the appropriate records management procedures for all data migration initiatives (Section 4.2)
- Retain data migration source data for a minimum 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 should be maintained for a minimum of three (3) years for audit purposes. (Section 4.2)

6. Related ITPs/Other References

- OPD-INF000A – *Migration Audit Checklist Template*
- ITP-APP022 - *Financial Applications Policy*
- ITP-SEC019 - *Policy and Procedures for Protecting Commonwealth Electronic Data*
- OPD-SEC019A – *Data Categorization and Inventory Operating Template*
- ITP-SEC020 – *Encryption Standards for Data at Rest*
- ITP-SEC024 - *IT Security Incident Reporting Policy*
- ITP-SEC025 - *Proper Use and Disclosure of Personally Identifiable Information (PII)*
- ITP-SEC031 – *Encryption Standards for Data in Transit*
- Management Directive 325.12 – *Standards for Internal Controls in Commonwealth Agencies*
- Executive Order 2016-07 – *Open Data, Data Development, and Data Governance*
- Federal Internal Control Standards - http://www.gao.gov/assets/670/665712.pdf
- National Institute of Standards and Technology (NIST) Federal Information Processing Standard Publication (FIPS-PUB) 199 - *Standards for Security Categorization of Federal Information and Information Systems*
- National Strategy for Information Sharing and Safeguarding
- NIST guidance on Security and Privacy Controls for Federal Information Systems and Organizations
- National/Homeland Security and Privacy/Confidentiality Checklist and Guidance

7. Authority

Executive Order 2016-06, *Enterprise Information Technology Governance*

8. Publication Version Control

It is the user's responsibility to ensure they have the latest version of this publication, which appears on https://itcentral.pa.gov for Commonwealth personnel and on the Office of Administration public portal: http://www.oa.pa.gov/Policies/Pages/default.aspx. Questions regarding this publication are to be directed to RA-ITCentral@pa.gov.

9. Exemption

In the event an agency chooses to seek an exemption, for reasons such as the need to comply with requirements for a federally mandated system, a request for waiver may be submitted via the Commonwealth of PA Procurement and Architectural Review (COPPAR) process. Requests are to be entered into the COPPAR Tool located at http://coppar.oa.pa.gov/. Agency CIO approval is required.

This chart contains a history of this publication's revisions:

Version	Date	Purpose of Revision
Original	07/11/2016	Base Document