

Personally Identifiable Information (PII) Study

In accordance with HB21-1111

Joint Technology Committee
Jan. 27, 2023

Amy Bhikha, Chief Data Officer
Michael McReynolds, Legislative Liaison

Chelsea Wyatt, Gartner Consulting Senior Managing Partner
Farhat Naweed, Gartner Consulting Senior Director
Nikhil Nayak, Gartner Consulting Associate Director
Bharat Bagaria, Gartner Consulting Expert Partner, Data & Analytics



COLORADO
Governor's Office of
Information Technology
Serving people serving Colorado

Gartner[®]



House Bill 21-1111

Legislators asked:

- Study **where personally identifiable information (PII) is stored** by state agencies throughout Colorado
- Identify entities that have **access** to PII stored by state agencies
- **Determine the costs and processes** necessary to centralize the storage and protection of PII
- Complete the study and have the advisory group present its findings and recommendations to the Joint Technology Committee

Personally identifiable information means information that may be used, along or in conjunction with any other information, to identify a specific individual, including but not limited to:

- Name
- Date of birth
- Place of birth
- Social security number
- Tax identification number
- A password or passcode
- Official government-issued driver's license or identification card number
- Information contained in an employment authorization document
- Information contained in a permanent resident card
- Vehicle registration information
- License plate number
- Photograph
- Electronically stored photograph, or digitized image
- Fingerprint
- Record of a physical feature
- Physical characteristic
- Behavioral characteristic
- Handwriting
- Government passport number
- Religion
- Health insurance identification number
- An employer, student, or military identification number;
- Financial transaction device
- School or educational institution attended;
- Source of income
- Medical information
- Biometric data
- Financial and tax records
- Home or work addresses or other contact information
- Family or emergency contact information;
- Status as a recipient of public Assistance or as a crime victim
- Race
- Ethnicity
- National origin
- Immigration or citizenship status
- Sexual orientation
- Gender
- Identity
- Physical disability
- Intellectual and developmental
- Disability



Advisory Group

Per statute, the 30-member Advisory Group was comprised of:

- Members of the Government Data Advisory Board (GDAB) - one representative per state agency - meets monthly to address statewide data strategy and interoperability
- Chaired by Amy Bhikha, Chief Data Officer
- Privacy and Subject Matter Experts:
 - CISO
 - Governor's Office
 - OIT Legislative Liaison
- Input from 40+ agency data experts

The Advisory Group met twice a month for six months

- Meetings were open to the public and members were invited to bring in expertise from their agency to offer comment and opinion



Government Data Advisory Board (GDAB)

**CIO Designee:
CDO**

**Attorney
General**

**Secretary
Of State**

**Gov
Office**

Judicial

Regulatory Affairs

- Dept of Personnel & Admin
- Dept of Regulatory Agencies
- Dept of Revenue

Education

- Dept of Education
- Dept of Higher Education
- School District Representation

Public Safety

- Dept of Corrections
- Dept of Military and Veterans Affairs
- Dept of Public Safety

Workforce & Economy

- Dept of Labor and Employment
- Dept of Local Affairs
- OEDIT

Environment & Renewable Energy

- Dept of Agriculture
- Dept of Natural Resources
- Dept of Public Health & Environment
- Dept of Transportation
- CO Energy Office

Health

- Health Care Policy & Financing
- Dept of Human Services
- Dept of Public Health & Environment
- Dept of Early Childhood
- Behavioral Health Administration



Existing Efforts with PII & Identity

GDAB Subcommittees, Completed Deliverables & Next Steps

Data Inventory	Data Sharing	Data Governance
<ul style="list-style-type: none">• Data Inventory Scope & Requirements• Requirements for Tool Selection• Tool Selection	<ul style="list-style-type: none">• Standard Inter-Government Data Sharing and Data Agreement & CJIS Addendum• Data Sharing Policy and Procedure• List of Data Sharing Risks and Mediations• Data Sharing Standards and Terms• Data Sharing Template for use with 3rd Parties• Guidance on Template use, including criteria for Sharing Agreement management	<ul style="list-style-type: none">• PII Definition• PII Protocol• Definition of Data Lifecycle and Accompanying Policy and Procedure• Data Retention Policy• Data Reconciliation Process• Data Roles• Data Governance Maturity Survey

- Innovative work - New for the state (both in terms of deliverables and collaboration)
- GDAB: Creation of community at data leadership level across agencies
- GDAB: Building a culture of collaboration between agencies (including OIT)
- Data Inventory: Data Stewards, usage of data, overlaps and authoritative sources within agencies
- Unique opportunity to work with Gartner, an industry leader, and get feedback on our efforts
- Alignment with other state CDOs (CDO Network)

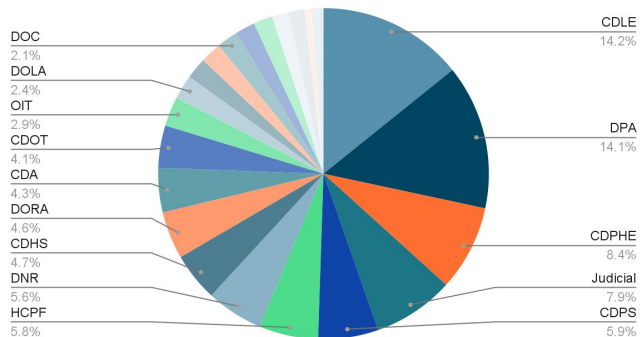


Where is PII Stored?

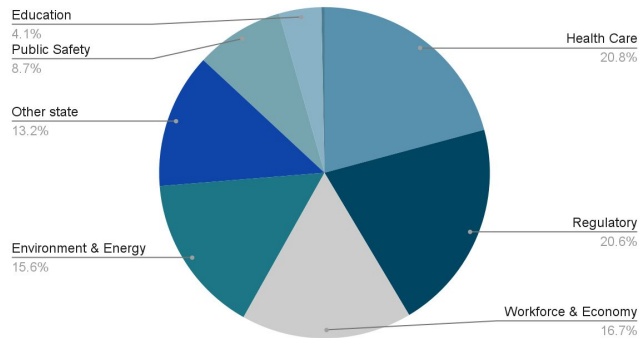
To determine where PII is stored, we leveraged the work of the GDAB and the legislatively mandated Data Inventory.

- Using technology workbooks as the starting point, 1,822 data sets were inventoried.
- 987 of these data sets were determined to contain PII. Each agency has at least some data sets with PII. The health domain has the most collected PII.
- The effort was completed manually by over 500 stewards.

PII Datasets by Agency



PII Datasets by Data Domain





Access to PII

To determine who has access to PII, we were able to use data classifications from the Data Inventory. The classifications fall under two categories, with each maintaining subclassifications of PII data.



Release of PII is Prevented (22%)

- PII assets have been identified, yet their disclosure is prevented. The prevention of disclosure follows subclassifications listed below:
 - **Internal use only** - Internal use is data that is accessible by internal staff within the agency or division.
 - **Protected, Never Released** - This data may not be released to the public or another agency.

*Note: Approximately 4% of identified assets have yet to be classified



Release of PII is Permitted (74%)

- PII assets have been identified and their disclosure is allowed. The allowance of disclosure follows sub-classifications listed below:
 - **Publicly Accessible** - The public is able to get the data without interfacing with an employee.
 - **Publicly Requestable** - The public is able to request the data from an employee or register in a system to receive the data. This could be a CORA request that does not require redaction.
 - **Sensitive** - Sensitive data is confidential information that must be kept safe and out of reach from all outsiders unless they have permission to access it. Access to sensitive data should be limited through sufficient data security and information security practices designed to prevent data leaks and data breaches.
 - **Protected, releasable with restrictions** - The public may request the data but some of the data needs to be redacted, removed or de-identified before it may be released.



Statewide Centralization

Gartner's assessment indicates complete centralization of PII or other critical data assets is unfeasible at this time for the following reasons:

- **Scale:** It would require major modification or redevelopment of hundreds of applications across the 29 Colorado agencies.
- **Cost Prohibitive:** This type of change would cost more than \$1 billion due to the modification of all the other applications required.
- **Unsupported by Vendors:** Commercial off-the-shelf vendors would not be able to support this type of uplift due to the overall scale of the data elements.
- **Impact to Operations:** PII centralization would not allow for operational business to function at most of, if not all, the Colorado agencies.
- **Data Breach:** Creation of a centralized PII solution would not resolve security concerns. The goal would be to understand where PII exists to secure and protect, not centralize. Failure in applications that would point to this data could propagate across the State.



Agency Feedback

The Advisory Group concurs with Gartner's findings, noting additional concerns:

- **Legal & Regulatory:** Differing legal mandates, rules, etc., across agencies
- **Consent:** Privacy impacts of centralization
- **Security:** Concern in scope for potential compromise of centralized location
- **Data Ownership:** Clarity on ownership and data sharing agreements
- **Procurement:** Agencies are reliant on 3rd party software, with existing contracts and established programming that would require modification
- **Logistical:** Non-consolidated agencies (e.g., Education, Judicial, SOS)
- **Priority:** Data strategy vs. agency priorities
- **Data Governance Maturity:** Differing data governance models and levels of maturity
- **Data Retention:** Differing policies by agency and usage
- **Cost:** Reiterating cost of restructure of current systems is prohibitive, including both technical and agency resources

PII Needs Analysis Report – Current State



COLORADO
Governor's Office of
Information Technology



Summary of Key Discovery Findings/Implications

1

**Need for a
Statewide
Governance Model**

2

**Inconsistent Data
Definitions**

3

**Need for a
Statewide Data
Model**

4

**Need for Agency
Specific Data
Requirements**

5

**Limited Singular
Resident View**

6

**Agency and Data
Specific
Security/Privacy
Considerations**



Summary of Key Discovery Findings/Implications

Findings

Implications

1

Need for a Statewide Governance Model

Some agencies maintain a data governance model and structure while others have elements of what a governance model may contain. Discrepancies among the agencies for how common data elements are housed, leveraged and shared can be rectified by a universal perspective for critical data.

A current set of varied approaches, data definitions and data models among the agencies demonstrates a need for a statewide PII data governance model.

2

Inconsistent Data Definitions

Each agency and business unit leverages its own definition of what constitutes PII/PHI/FTI or other critical data elements as it relates to their specific organizational mission and business processes. Agency specific data and systems subsequently maintain differing levels of security, data usage and architecture due to these discrepancies.

Non-universal definitions of critical data elements and how their usage is governed can lead to potential data leaks or create data quality issues for agencies that share and consume data from partner agencies.



Summary of Key Discovery Findings/Implications

Findings

Implications

3

Need for a Statewide Data Model

While there is a general belief that the mandates of the agency and program data needs are specific, there are commonalities at the data element level. Name, address, SSN and Colorado ID are potential examples that could knit Colorado agencies together.

A statewide data model is needed for consistent data sharing where necessary. This data model would enhance both reporting and data insight capabilities among agencies. It would also serve as a platform for a universal data definition standard.

4

Agency Specific Data Requirements

Each agency maintains specific timeframes for both data usage and storage. These depend on organizational mission and provided services. Various regulations and legal statutes relate to agency-specific data and can restrict how critical data is leveraged, shared and maintained within each agency.

Master Data Management strategies and solutions will require flexibility to ensure agency needs are upheld while limiting significant changes to existing processes or services provided. This will help ensure agency operations are not affected.



Summary of Key Discovery Findings/Implications

Findings

Implications

5

Singular Resident Views

Multiple agencies share PII and other data with each other, yet siloed data within agencies and across agencies limits staff's ability to create a singular Colorado resident view to best serve their needs.

Siloed data architecture restricts some agencies from providing some needed government services or identifying users who should not have these services. It also restricts identification of those eligible for programs and understanding resident needs across agencies/program areas/service types.

6

Security/Privacy Considerations

Critical data sets require specific protections to ensure the safety and validity of the information that data houses. Just as with regulatory restrictions, the type of data and requisite security requirements vary significantly among agencies. These depend on their specific use cases and regulatory requirements.

Master and Metadata Management strategies will need to incorporate data- and agency-specific security practices and needs while still providing central systems for critical data.



PII Consolidation Target State



Master Data Management (MDM) & Metadata Management



MASTER DATA MANAGEMENT (MDM)

A software application that would centralize a copy of information from each application with PII from all agencies, centralizing it at either the State level or within each agency, i.e., creating a “Golden Record” for each Coloradan that all agencies could use.



METADATA MANAGEMENT

A software application that used to enhance the usability, comprehension, utility or functionality of any other data point. This solution would track where PII data exists and who has access.

EXAMPLES

- Data Sharing and Reporting - A constituent changed their address at the DMV. A central MDM repository ensures that other agencies like those that manage voter registration and tax information would have access to information about the address change.
- Data Sharing and Reporting - A Coloradan qualifies for the Colorado Homeless contribution income tax credit and is not enrolled in any housing or rental assistance programs. The Department of Local Affairs can identify eligible programs and notify the Colorado resident.

EXAMPLES

- Data Access - All Colorado agencies using the metadata management system can see who has access to PII and where it exists across applications, they can also see where data is moving.
- Data Sharing - OIT can monitor how PII information is shared with third-party vendors and ensure the agreements with those vendors are being enforced effectively for not just applications with PII, but all applications within the system.
- Lineage - All agencies understand where the system of record exists, and can ensure changes there propagate to other systems. A public health patient’s name change in their patient portal could propagate to other systems if Colorado Health and Human Services structured it this way.



CO-OIT has 4 solution options to consider for meeting the needs of House Bill 21-1111

Option 1: Master Data Management (MDM)

- Start by implementing MDM at each of the individual departments (Golden record per department).
- As each organization's MDM matures, consider implementation of MDM at the State level (Golden record for State).

Option 2: Metadata Management

- Implement metadata management at each department.
- As the State mature's as an enterprise, consider implementing the metadata management solution at the enterprise level. This will accommodate the metadata associated with other data sets as well as PII.

Option 3: MDM & Metadata Management



- This is a combination of options 1 & 2 (Implement Master Data & Metadata Solutions in parallel).

Option 4: Entity Resolution & Metadata Management

- Use existing IDX architecture and expand based upon use-case classification or agency domain.
- Implement federated style of metadata management from option 2.



Approach

Legislative Requirement	Option 1: MDM	Option 2: Metadata Management	Option 3: MDM & Metadata Management	Option 4: Entity Resolution & Metadata Management
Identifies PII Data Locations	Yes	Yes	Yes	Yes
Identifies PII Data Access	No	Yes	Yes	Yes
Centralization of PII Data	Yes	No	Yes	Partial Yes
Cost (10 yr. Capital outlay)	\$40M - \$80M	\$10M - \$20M	\$50M - \$100M	\$35M - \$70M
 <p>Benefits</p>	<ul style="list-style-type: none"> Central source of cleansed, standardized and consolidated master data Minimal footprint and impact to existing architecture Provides the ability to define group and user-level rights Creates golden record at both department and state levels 	<ul style="list-style-type: none"> Can track the activities of data users to understand data usage, the most important data sets/records, related datasets, and the nature of those relationships. Advanced insight will include data lineage and historical information as data records evolve over time among agencies 	<ul style="list-style-type: none"> Benefits of option 1 and 2 apply here 	<ul style="list-style-type: none"> IDXR is used to create a common Citizen ID across different systems / applications. Can track the activities of data users to understand data usage, the most important data sets/records, related datasets, and the nature of those relationships.
 <p>Drawbacks</p>	<ul style="list-style-type: none"> This approach does not update the original source record for those consolidated data elements. Does not provide insights into the PII metadata e.g., data usage, data access etc. 	<ul style="list-style-type: none"> Only monitors the passive or active attributes of the datasets rather than the actual record No golden PII customer record is created 	<ul style="list-style-type: none"> This approach does not update the original source record for those consolidated PII data elements. 	<ul style="list-style-type: none"> IDXR functions like a registry MDM solution. This does not create a golden record. Unless expanded to and consolidated among all agencies, multiple instances of IDXr will be needed to a specific agency data regulatory or policy restrictions.



High Level Costs – Rough Order of Magnitude (ROM) 10-year Capital Cost*

Cost Buckets	Option 1: MDM	Option 2: Metadata Management	Option 3: MDM & Metadata Management	Option 4: Entity Resolution & Metadata Management
Software Licensing (Cost for 10 yrs.)	\$20M - \$40M	\$5M - \$10M	\$25M - \$50M	\$15M - \$30M
One Time Implementation (1-3 yrs.)	\$10M - \$20M	\$2.5M - \$5M	\$12.5M - \$25M	\$10M - \$20M
Ongoing M&O (7-9 yrs.)	\$10M - \$20M	\$2.5M - \$5M	\$12.5M - \$25M	\$10M - \$20M
Total (10-year TCO)	\$40M - \$80M	\$10M - \$20M	\$50M - \$100M	\$35M - \$70M

*Note - this cost model shows an initial estimate of the 10-year capital cost of acquiring the technology or implementing the process. This initial estimate **does not include agency or OIT staff time or backfill / hiring requirements. Technical costs only.**



Option 1: Master Data Management – PII Data

Master Data Management is a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and accountability.

Description	Colorado Impact	Implementation Style	Security Considerations
<p>1. MDM is about maintaining a "single trusted version" for critical concepts that describe what an organization does. It would enabled the Departments and CO-OIT to work together ensuring the accuracy, security, and stewardship of the PII data.</p>	<p>1. A MDM solution will allow the State to bridge across fragmented silos of PII customer/citizen data and create a trusted customer/citizen profile (golden record).</p> <p>2. A comprehensive PII master data management approach consists of processes such as data collection, accumulation, data cleansing, data comparison, consolidation, quality control and data distribution within departments and across departments to ensure consistency and control.</p>	<p>1. MDM Solutions are typically implemented within 4 distinct styles (Consolidated, Registry, Centralized, & Co-Existence)</p> <p>2. Of the four implementation styles, a consolidated approach would best suit CO-OIT's needs and addresses HB21-1111 requirements (i.e. location and access of the golden record).</p>	<p>1. A Consolidated Style would require CO-OIT unite all identified PII element into a singular infrastructure.</p> <p>2. CO-OIT in conjunction with each agency , each agency would be would be required to identify all security and regulatory restrictions for PII data across the public agency domains as it attempts to consolidate.</p> <p>3. OIT would remain responsible for the maintenance of the MDM solution and PII data contained while each agency would continue to maintain their systems which contain PII data.</p>



Option 2: Metadata Management

Metadata Management is a set of capabilities that enables continuous access and processing of metadata that support ongoing analysis of information.

Description	Colorado Impact	Implementation Style	Security Considerations
<ol style="list-style-type: none">1. Metadata Management Solutions serve to collate and communicate the inventory of data assets, communicate the business contexts of information, communicate the glossary of business terms, provide monitoring, auditing and traceability, and serve as a dynamic collaboration environment.2. Metadata Management Solutions will provide insight into data element Semantics, Location, Access, Trust, and Utilization.	<ol style="list-style-type: none">1. Metadata Solutions can provide a glimpse into the workflow of data, data consumption, and other attributes of identified datasets,	<ol style="list-style-type: none">1. MDM Solutions are typically implemented within three distinct styles (Centralized, Federated, & Distributed).2. Of the three implementation styles, a federated approach would best suit State's needs and addresses HB21-1111 requirements.	<ol style="list-style-type: none">1. While there are no critical elements contained within a metadata management solution, the behavioral data around utilization of critical data needs to be protected. Varying levels of security and regulation apply to active metadata capture by a metadata management solution.



Option 3: Master Data Management & Metadata Management

Combining Options 1 & 2 can enable Colorado to address each of the components of HB21-1111.

Description	Colorado Impact	Implementation Style	Security Considerations
<ol style="list-style-type: none">1. The State would identify and implement Master and Metadata Management Solution(s) together.	<ol style="list-style-type: none">1. The combination of Master and Metadata Solutions will provide Colorado the necessary flexibility to create a map of where all PII elements exist while also moving towards a “golden record” for specific PII elements.2. Colorado would also garner the individual benefits of each of these dedicated deployments.	<ol style="list-style-type: none">1. Colorado could implement the same styles of Master and Metadata Solutions as suggested in Options 1 & 2.2. Colorado would be required to identify the connection points between these solutions to ensure full functionality is realized.	<ol style="list-style-type: none">1. While there are no critical elements contained within a metadata management solution, the behavioral data around utilization of critical data needs to be protected. Varying levels of security and regulation apply to active metadata capture by a metadata management solution.2. MDM and Metadata management solutions will maintain modern security functionality (i.e., encryption, authentication / authorization, etc.)



Option 4: Entity Resolution & Metadata Management

Combining Option 2 with the existing IDXr solution can enable Colorado to partially address each of the components of HB21-1111.

Description	Colorado Impact	Implementation Style	Security Considerations
<ol style="list-style-type: none">1. The State would identify and implement Metadata Management solution with the existing IDXr offering.2. IDXr currently resemble the “Registry” implementation style of a Master Data Management solution and provides an index and cross reference of PII data elements across linked systems.	<ol style="list-style-type: none">1. The combination of IDXr and Metadata Solutions will provide Colorado the necessary flexibility to create a map of where all PII elements exist. IDXr does not create a “golden record” but a registry of what PII assets relate to one another would be established.2. Colorado would also garner all the individual benefits of a Metadata Management Solution.	<ol style="list-style-type: none">1. Colorado could implement the same style Metadata Solutions as suggested in Option 2.2. Colorado would be required to identify the connection points between these solutions to ensure full functionality is realized.	<ol style="list-style-type: none">1. While there are no critical elements contained within a metadata management solution or IDXr, the behavioral data around utilization of critical data needs to be protected.2. Several deployments of IDXr could be established each with varying levels or security or regulatory needs which address agency-domain specific needs.



Target State PII Initiatives

© 2021 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. or its affiliates. This presentation, including all supporting materials, is proprietary to Gartner, Inc. and/or its affiliates and is for the sole internal use of the intended recipients. Because this presentation may contain information that is confidential, proprietary or otherwise legally protected, it may not be further copied, distributed or publicly displayed without the express written permission of Gartner, Inc. or its affiliates.

Gartner®



Gartner: State PII Best Practices



Change Management

- Create knowledge that data is a shared, strategic asset
- Share stories around other states and worldwide efforts for shared data structures



Data Literacy Education

- Clarity and consistency on definition of PII
- What state data should be tagged as PII and agency implications
- Clarity on how to protect PII



Data Governance

- Require data governance from accountable entities within each agency
- Ensure that the data governance structure matches the technology solution



Data Usability

- Validate that data is reusable
- Ensure that data is interoperable



Data Set Maintenance

- Verify that each data set provides new insights and that data sets are not repeated, especially within agencies
- Ensure scheduled maintenance exists, or at least uses guidelines from the State on data retention and disposal



Data Source Identification

- Ensure that data sources are identified for PII, and integration maps are created, especially for external use of PII
- Provide agencies with data dictionary examples/ask to create one for each data set



Security

- Ensure data is secured and protected
- Define security and risk governance deciding what is acceptable risk and how to enable risk control
- Map and Monitor all data access privileges provided to application users, developers, etc



Colorado's Current Best Practice Initiatives



Change Management

- Ongoing efforts and resourcing to define Data as a Strategic Asset
- Alignment with GDAB, CDO Network, strategic partners



Data Literacy Education

- Independent agency efforts.
- R05 DI funding would create resource for centralized literacy program.



Data Governance

- GDAB Data Gov Subcmte clarifying data roles, creating data maturity matrix and generating supporting templates.
- R05 DI funding would create focused resource for statewide data gov program.



Data Usability

- Interoperability efforts supported by maturity of data governance, and adherence to integrated and security workflows (OIT Data Standard TD DAT 001)
- Data Sharing Agreement process and inventory, supported by R05 funding.



Data Set Maintenance

- Formalization of Data Stewards and Data Inventory efforts via GDAB to provide insight.
- R05 DI funding would create funding for advanced tooling and focused resource.



Data Source Identification

- Formalization of Data Stewards and Data Inventory efforts via GDAB to to define where PII exists.
- Maturity of Data Architect role to establish integration maps, working toward statewide data model.



Security

- Align agencies with CISO/CDO efforts to ensure data is protected and security governance is established and adhered to.
- Potential to consider Data Security Assessment and corresponding framework in future.

Next Steps



COLORADO
Governor's Office of
Information Technology

Serving people serving Colorado

Gartner[®]



Gartner: Next Steps

OIT can consider several steps as it continues its journey to identify and manage PII across the agencies

- 1 Confirm legislative direction from the JTC con the four PII Data Management solution options
- 2 Develop Business Case for a PII Data Management Solution (including whether non OIT-consolidated agencies opt-in)
- 3 Complete a benefits analysis and determine viable deployment options from existing options
- 4 Identify top use cases for pilot initiatives
- 5 Develop the TCO Model into a project budget and business case and request legislative approval

Additionally, OIT can consider incorporating other initiatives that can accelerate not just PII management needs but increase capabilities across agencies with respect to data management and usage:

- Establish data governance standards and frameworks to be leveraged across agencies
- Develop a statewide data literacy program for agency personnel
- Define universal data security governance to establish acceptable levels of risk and how to enable risk control
- Please see slide 25 for additional information on PII Best Practices



Advisory Groups Suggestions

Support Current Best Practices Initiatives for PII

Potential consideration of PII Security Assessment Framework

Data Privacy - Need for potential centralized and/or agency roles

Establishment of a state level role to create and manage a statewide privacy program to ensure that PII is managed according to best practices in compliance with all applicable privacy laws. This role could created policy and provide support to agencies

Legal - Study of potential laws/rules changes



Closing Remarks

This study adds value to ongoing discussions that must continue to mature to determine the proper course.

Additional legislation **might be needed** in the future once a clear strategy is identified and agreed upon.

Specific privacy roles **would likely benefit the State** and advance PII strategies.

This is highly complex. More time, research and inter-agency discussions are necessary in order to solidify the right next steps.

Questions?



COLORADO
Governor's Office of
Information Technology

Serving people serving Colorado

Gartner[®]