Federal Zero Trust Data Security Guide

PUBLISHED OCTOBER 2024

REVISED MAY 2025 (ADDITION OF CHAPTER 4: MANAGE THE DATA)

TABLE OF CONTENTS

TABLE OF CONTENTS

FIGURES

TABLES

AUTHORITY

This document was developed by the **Zero Trust (ZT)** Data Security Working Group in furtherance of its directive under the **Office of Management and Budget (OMB)** Memorandum M-22-09, *Moving the U.S. Government Towards Zero Trust Cybersecurity Principles*.[1] The Working Group is a joint committee comprised of members from the Federal **Chief Data Officer (CDO)** Council and **Chief Information Security Officer (CISO)** Council, as well as other Federal stakeholders. OMB M-22-09 charged the Working Group with: (1) developing a guide for Federal agencies that addresses how existing Federal information categorization schemes can support effective data categorization in a security context; (2) developing enterprise-specific data categories that are not addressed by existing Federal categories; (3) identifying members who will act as leads, or designate leads within their agencies, to convene a **community of practice (CoP)** that can assist agencies in tackling specific areas of focus; and (4) identifying and supporting pilots of emerging approaches and best practices among agencies. This document is intended to assist agencies on their ZT journey as they continue to implement ZT principles year-to-year and across Administrations.

Disclaimer

This document is intended to provide best practices to help agencies consider data security issues that are relevant to the implementation of ZT requirements. This document does not establish or modify any requirements in law, regulation, or policy. The best practices presented in this document are intended to be consistent with Federal policies and laws, including, but not limited to: **Executive Order (EO)** 14028 on *Improving the Nation's Cybersecurity*;[2] EO 13744 on *Making Open and Machine Readable the New Default for Government Information*;[3] Foundations for Evidence-Based Policymaking Act (Pub. L. 115–435); and OMB M-19-18, *Federal Data Strategy — A Framework for Consistency*.[4] Nothing in this document may be construed to replace or supersede any authorities, policies, guidance, standards, or other articles made mandatory and binding for Federal agencies under statute. Agencies should consult with their counsel, policy officials, and relevant stakeholders to determine the appropriate process for implementing ZT data security requirements.

This document is non-binding and is a reference for Federal agencies. While nongovernmental organizations may use this document on a voluntary basis and this document is not subject to copyright regulations, attribution would be appreciated by the Federal **Chief Information Officer (CIO)** Council, CISO Council, and CDO Council.

[1] OMB M-22-09: *Moving the U.S. Government Towards Zero Trust Cybersecurity Principles*, https://www.whitehouse.gov/wp-content/uploads/2022/01/M-22-09.pdf

[2] EO 14028 on *Improving the Nation's Cybersecurity*, https://www.federalregister.gov/documents/2021/05/17/2021-10460/improving-the-nations-cybersecurity

[3] EO 13744 on *Making Open and Machine Readable the New Default for Government Information*, https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

[4] OMB M-19-18: *Federal Data Strategy – A Framework for Consistency*, https://www.whitehouse.gov/wp-content/uploads/2019/06/M-19-18.pdf

This document focuses on the Data Pillar of the **Cybersecurity and Infrastructure Security Agency's (CISA) ZT Maturity Model (ZTMM)**[5] and is not intended to cover every aspect of ZT. Certain commercial entities, equipment, or materials may be referenced in this document to adequately describe an experimental procedure, theoretical application, or concept. Such reference is not intended to imply recommendation or endorsement by the Working Group, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

ACKNOWLEDGEMENTS

This document was developed by CDOs, CISOs, **Senior Agency Officials for Privacy (SAOPs)**, and representatives from over 30 Federal departments, agencies, bureaus, and entities, in consultation with other stakeholders.

[5] CISA ZTMM, https://www.cisa.gov/sites/default/files/2023-04/zero_trust_maturity_model_v2_508.pdf

CHAPTER 1: INTRODUCTION

1.1: DATA MANAGEMENT IS CRITICAL TO MAKING ZERO TRUST A REALITY

Our vision is to establish a resilient and secure digital environment where users, assets, and access to resources are continuously validated and verified.

The cyber risk landscape is continuously evolving — and our adversaries are evolving along with it. The United States is facing unprecedented threats as malicious actors advance their tactics and unlock new ways to attack our systems, including using emerging technologies, such as **artificial intelligence (AI)**, to launch increasingly sophisticated cyber campaigns.

To counter these threats, agencies are making Federal systems more defensible by employing ZT principles — which means trust is never implicitly granted and must be continually validated.

ZT moves away from the traditional approach of protecting the network perimeter — a "castle and moat" model as seen in Figure 1 — to instead assume that a network may be compromised at any time, anywhere, and by anyone. Through the ZT lens, we focus on **securing the data itself, rather than the perimeter protecting it.** This concept is known as "ZT data security."

FIGURE 1: Traditional Security vs. Zero Trust

Traditional Security

Traditional network perimeter-based security, with its assumption of implicit trust inside the perimeter, has failed to protect enterprise assets. This belief that everything is safe and can be trusted once inside actually paves the way for attackers to cause chaos through unimpeded lateral movement.

Zero Trust

Zero trust assumes that all networks — enterprise-owned or not — are untrusted and that an attacker is present in the environment. It denies default access to data and workload, continually authenticates and authorizes each access request, and monitors and analyzes the risks to the assets.

ZT data security is not a tool that you can buy off the shelf and achieve overnight — it is a multi-year strategy that requires adaptability, investment, and collaboration. This guide aims to help agencies on their journey to ZT data security by being practical and operational, yet flexible enough for practitioners to adjust to meet their agency's unique needs and mission requirements. The intended audience for this guide includes practitioners charged with securing data, such as system owners, data management practitioners and stewards, system administrators, and cybersecurity engineers.

This guide breaks the ZT data security journey into three chapters:

1. **Define the Data:** This chapter helps readers find and identify the totality of their data landscape, learn how to accurately categorize and handle data, and define the sensitivity and criticality of data.
2. **Secure the Data:** This chapter focuses on implementing the appropriate security monitoring and controls, such as encryption, for data and incorporating risk management and **identity, credential, and access management (ICAM)**. It also explains considerations for privacy and compliance.
3. **Manage the Data:** This chapter explores how ZT data security principles can be embedded in **data lifecycle management (DLM)** to ensure robust protection at every stage. It discusses how readers can equip their team with the necessary skills and adapt their approach to address emerging technologies.

Implementing a holistic security approach like ZT is challenging in any environment, especially where components may operate independently. However, we must remember that cybersecurity is a team sport and — with demonstrated senior leader buy-in and stakeholder collaboration — we can overcome the challenge. Looking ahead, we must build and sustain a cooperative relationship between data management and cybersecurity teams across Government. By embracing innovation, cultivating partnerships, and acting swiftly, we can protect our data from adversaries.

1.2: CONNECTING THE DOTS BETWEEN ZERO TRUST AND DATA

Data security is at the heart of Zero Trust.

Protecting sensitive data assets is at the heart of the ZT model. As such, data management practices must be **secure by design**, where the security of data is treated as a core requirement from the beginning of and throughout its lifecycle — security cannot be a mere aspiration or afterthought. The ZT model forces security upstream in DLM, starting at the logical design stage of the data as depicted in Figure 2.

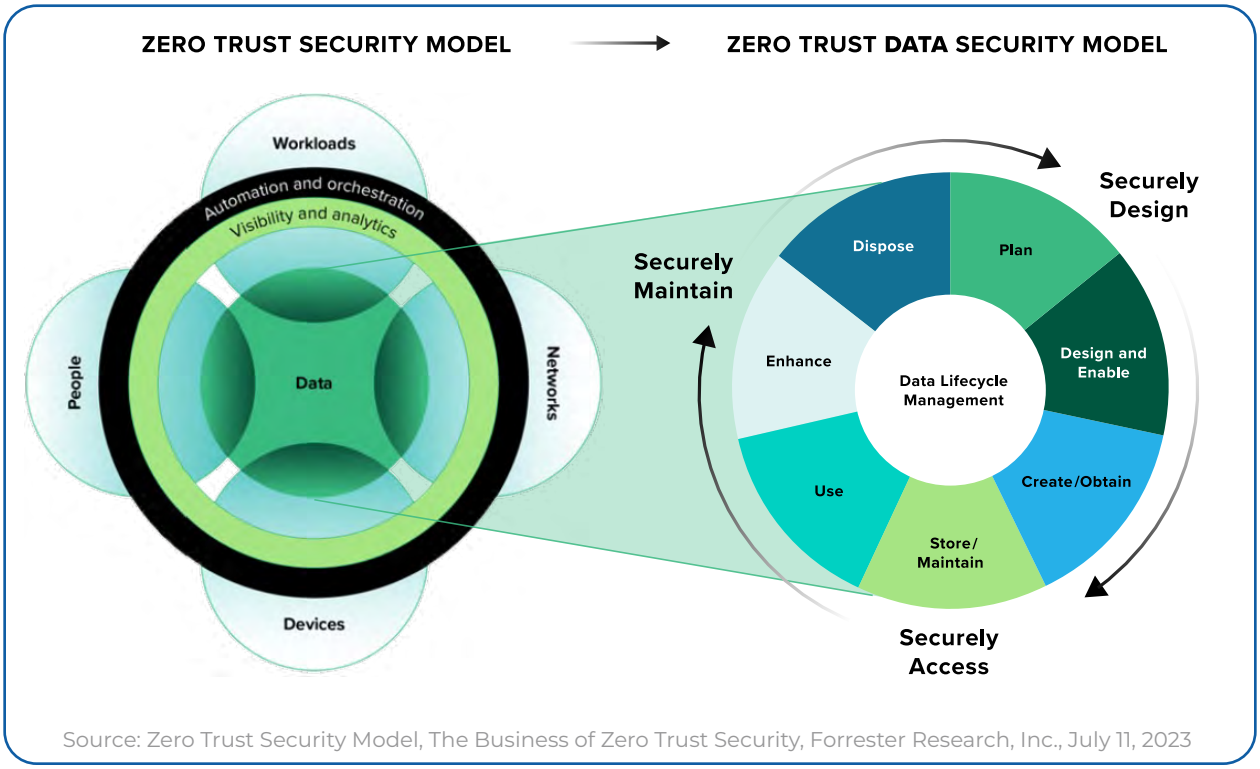
Establishing security as a core design criterion from the outset can also help facilitate seamless transition of data into other environments. This is particularly useful as agencies are increasingly moving their data to cloud-based, shared environments.

Data as the New Perimeter

"[We must] … extend the protect surface by defining data down to the cellular level as the new perimeter and adopting dynamic data tagging, labeling, and encryption technology that empowers data stewards and consumers and assures access from anywhere, anytime."

— Department of the Air Force *Zero Trust Strategy*, DAF Zero Trust Strategy v1.0 - SAF/C, https://www.safcn.af.mil/Portals/64/Documents/Strategy/DAF%20Zero%20Trust%20Strategy%20v1.0%20(002).pdf

FIGURE 2: Secure Data Lifecycle Management Is at the Heart of the Zero Trust Model



Source: Zero Trust Security Model, The Business of Zero Trust Security, Forrester Research, Inc., July 11, 2023

1.3: ZERO TRUST DATA SECURITY PRINCIPLES

Practitioners can reference the principles in Table 1 to inform decision-making and ensure alignment with their agency's strategy, mission, and values.

TABLE 1: Zero Trust Data Security Principles

Principle	Why?
Adopt a data-centric view	Data is everywhere and exists in different formats with varying levels of sensitivity and value. Protecting critical data requires visibility, analytics, and automation across the entire digital ecosystem.
Implement standardized least privilege and strictly enforce access control	The bedrock principles of ZT are that all entities are untrusted, least privilege access is enforced, and comprehensive security monitoring is implemented.
Promote data resiliency and integrity	The value of data is maximized when it is available, accessible, and trustworthy. The Federal government relies on quality data to conduct business and deliver services to the public.
Integrate data and security literacy	Data and security practitioners must understand each other's nomenclature to effectively safeguard their agencies' data and enable appropriate use.
The impact of data security must be measurable and actionable	Meaningful analytics that produce actionable insights can help to prevent breaches and reduce the impact of breaches when they do occur.
Data security is risk-informed throughout the data lifecycle	Each stage of the data lifecycle has specific security requirements. Security controls must address risks **to** the data, **from** the data, and **in** the data.
Balance priorities — make the most with what you have	ZT principles will shape existing practices, processes, and perimeters. As practitioners understand these changes, they can assess whether their current cyber infrastructure meets these evolved needs.

DID YOU KNOW?

The DAMA Guide to the **Data Management Body of Knowledge (DAMA-DMBOK2)** is a well-known reference guide for data management professionals, and has been used by practitioners at Federal agencies.

The DAMA-DMBOK2 defines the goals of data security as:

- Enable appropriate, and prevent inappropriate, access to enterprise data assets.
- Understand and comply with all relevant regulations and policies for privacy, protection, and confidentiality.
- Ensure that the privacy and confidentiality needs of all stakeholders are enforced and audited.

1.4: APPLYING ZERO TRUST DATA SECURITY GUIDANCE

The journey to ZT data security will be complex and extensive, so practitioners should develop a ZT data security roadmap to help manage the task effectively. Practitioners should ensure the roadmap aligns with their agency's ZT implementation plan, as required by EO 14028[6] and OMB M-22-09.[7]

Practitioners should start by assessing their agency's current maturity level and identifying the desired future maturity level with a clear timeframe. Utilizing an existing maturity model, such as CISA's ZTMM,[8] provides a way to evaluate the agency's current state, identify the target state goals and metrics to measure progress toward those goals, and refine the roadmap as needed.

If a practitioner's agency doesn't have a published or fully mature ZT implementation plan, they may draw inspiration from other agencies' published plans. While this guide can be used as a starting point, practitioners are encouraged to leverage other existing roadmaps and resources. Remember — don't reinvent the wheel or create silos!

[6] EO 14028 on *Improving the Nation's Cybersecurity*, https://www.federalregister.gov/documents/2021/05/17/2021-10460/improving-the-nations-cybersecurity

[7] OMB M-22-09: *Moving the U.S. Government Toward Zero Trust Cybersecurity Principles*, https://www.whitehouse.gov/wp-content/uploads/2022/01/M-22-09.pdf

[8] CISA ZTMM, https://www.cisa.gov/sites/default/files/2023-04/zero_trust_maturity_model_v2_508.pdf

To inform the drafting of their roadmap, practitioners should:

1. Assess the current maturity level of data management and security of their organization.
2. Consider the unique needs of their organization, including current and planned changes to cybersecurity posture, mission(s), and existing or anticipated organizational technology, strategic, business, and data management roadmaps and changes.
3. Understand the existing landscape of business initiatives, technology transformations, and security projects.
4. Document where existing data management and data security capabilities can be reused and where new capabilities are needed.
5. Identify and prioritize related initiatives such as **controlled unclassified information (CUI)** and privacy.
6. Set goals for future maturation and identify a timeframe to achieve those goals.

These steps may have already been performed in the agency's ZT implementation plan. If so, practitioners should verify the findings are up to date before proceeding.

1.4.1: UNDERSTAND THE STARTING POINT — FOCUS ON SMALL WINS AND BUILD OUT

A ZT data security roadmap should address the stages outlined in Table 2. Although Table 2 includes *suggested* actions and leads for each stage, practitioners should determine whether these are applicable to their organization.

TABLE 2: Stages of a Zero Trust Data Security Roadmap

Stage	What	Who
1. Discover, inventory, categorize, label, and map the data flows *See* Chapter 2 *to learn more*	• Establish the foundation and infrastructure for the ZT data security roadmap. • Find all the data across the agency. • Map the critical data flows between existing systems, processes, and data stores to discover the trust relationships and logical mission trust zones.	**Lead:** The agency's CDO or designated data steward(s) should lead this stage to develop an understanding of the current data landscape, data stewardship, categorization, and labeling in metadata. **Coordinate with:** Data governance body, SAOP, records retention officials, and other stakeholders early in the process to identify the steps in the roadmap
2. Conduct risk analysis *See* Chapter 3 *to learn more*	• Conduct a risk analysis to articulate data integrity issues that can cause financial, reputational, third-party, and supply chain losses. • Prioritize focus areas based on risk, impact, and achievability.	**Lead:** The agency's ZT or cybersecurity practitioners typically lead this and subsequent stages. **Coordinate with:** Data management and privacy practitioners and agency experts involved in ongoing risk analysis and threat modeling

Stage	What	Who
3. Align with the ZT architecture	• Review the agency's ZT implementation plan. • Set timelines and quick feedback mechanisms.	**Lead:** The agency's ZT or cybersecurity practitioners typically lead this and subsequent stages. **Coordinate with:** The team responsible for implementing ZT, such as the agency's ZT Program Office, and data management practitioners.
4. Design the controls and monitoring *See* Chapter 3 *to learn more*	• Identify tools to close data security gaps. • Implement data-centric security controls, monitoring, logging, and alerting. • Refactor the architectural, **Federal Information Security Modernization Act (FISMA)**, financial boundaries, and data standardization and centralization.	**Lead:** The agency's ZT or cybersecurity practitioners typically lead this and subsequent stages. **Coordinate with:** Data management and privacy practitioners and agency experts involved in ongoing risk analysis and threat modeling.
5. Embrace automation and orchestration	• Identify opportunities to integrate and orchestrate automation to manage the operational data.	**Lead:** The agency's CDO or designated data steward(s) should lead this stage. **Coordinate with:** Security and privacy practitioners and agency experts involved in automation (e.g., **Chief AI Officer [CAIO]**)

CHAPTER 2: DEFINE THE DATA

While agencies have been required to inventory their data assets for some time, a comprehensive ZT approach to data management requires going beyond and — in some ways — *redefining* what agencies may be accustomed to thinking of as "data." This chapter provides guidance on data management and governance practices essential to ZT data security. It leverages a rich, established body of work published by the CDO Council rooted in existing Federal statutes and **National Institute of Standards and Technology (NIST)** guidelines.

Agencies must first identify and locate their sensitive data assets, inventory those assets, establish data categories using norms aligned to their missions, and label those assets accordingly. This chapter focuses on how to:

Agency data management and privacy practitioners should lead the operationalization of the guidelines outlined in this chapter, working closely with cybersecurity practitioners to conduct or update their data inventory.

1. Identify and locate data assets to understand the entirety of an agency's sensitive data landscape.
2. Create a data asset catalog and inventory that aligns with the agency's mission and needs as the single source of truth for agency data assets.
3. Categorize the data assets so the appropriate security controls can be applied.
4. Label (i.e., mark or tag) the data sources and associated metadata to support automated protection.

2.1: LEGAL AND REGULATORY REQUIREMENTS FOR DATA INVENTORY AND CATEGORIZATION

Multiple laws, regulations, and Federal guidance already require the maintenance of an inventory of data assets. As such, agencies may be able to build on existing inventories to jumpstart their ZT data inventory, categorization, and labeling initiative. The list below provides examples of existing laws, regulations, and guidance that may require inventory-related efforts. Note that this list only includes examples in effect at the time of this guide's publication, and **it's the agency's responsibility to stay up-to-date with the latest legal requirements.** Examples include, but are not limited to:

- The Foundations for Evidence-Based Policy Making Act of 2018 ("Evidence Act," Pub. L. 115-435)[9]

[9] The Foundations for Evidence-Based Policy Making Act of 2018, https://www.congress.gov/bill/115th-congress/house-bill/4174

- OMB M-21-27, *Evidence-Based Policymaking: Learning Agendas and Annual Evaluation Plans*[10]
- OMB M-19-18, *Federal Data Strategy — A Framework for Consistency*[11]
- OMB M-23-04, *Establishment of Standard Application Process Requirements on Recognized Statistical Agencies and Units*[12]
- The Paperwork Reduction Act of 1980 (Pub. L. 96–511)[13]
- The Privacy Act of 1974 (Pub. L. 93-579)[14]
- The E-Government Act of 2002 (Pub. L. 107-347)[15]
- FISMA 2014 (Pub. L. 113-283)[16]
- The **Health Insurance Portability and Accountability Act (HIPAA)** of 1996 (Pub. L. 104-191)[17]
- "Desirable Characteristics of Data Repositories for Federally Funded Research,"[18] published by the National Science and Technology Council
- CISA **Binding Operational Directive (BOD)** 18-02: Securing **High Value Assets (HVA)**[19]
- CISA ZTMM Version 2.0[20]
- "A Framework for Data Quality,"[21] published by the **Federal Committee on Statistical Methodology (FCSM)**

[10] OMB M-21-27: *Evidence-Based Policymaking: Learning Agendas and Annual Evaluation Plans*, https://www.whitehouse.gov/wp-content/uploads/2021/06/M-21-27.pdf

[11] OMB M-19-18: *Federal Data Strategy — A Framework for Consistency*, https://www.whitehouse.gov/wp-content/uploads/2019/06/M-19-18.pdf

[12] OMB M-23-04: *Establishment of Standard Application Process Requirements on Recognized Statistical Agencies and Units*, https://www.whitehouse.gov/wp-content/uploads/2022/12/M-23-04.pdf

[13] The Paperwork Reduction Act of 1980, https://www.congress.gov/96/statute/STATUTE-94/STATUTE-94-Pg2812.pdf

[14] The Privacy Act of 1974, https://www.congress.gov/93/statute/STATUTE-88/STATUTE-88-Pg1895.pdf

[15] The E-Government Act of 2002, https://www.congress.gov/bill/107th-congress/house-bill/2458

[16] FISMA 2014, https://www.congress.gov/bill/113th-congress/senate-bill/2521

[17] HIPAA 1996, https://www.congress.gov/bill/104th-congress/house-bill/3103/text

[18] Desirable Characteristics of Data Repositories for Federally Funded Research, https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf

[19] CISA BOD 18-02: Securing High Value Assets, https://www.cisa.gov/news-events/directives/bod-18-02-securing-high-value-assets

[20] CISA ZTMM, https://www.cisa.gov/sites/default/files/2023-04/zero_trust_maturity_model_v2_508.pdf

[21] A Framework for Data Quality, https://nces.ed.gov/FCSM/pdf/FCSM.20.04_A_Framework_for_Data_Quality.pdf

2.2: DEFINING THE DATA ASSETS

An agency's data landscape may seem daunting and uncharted due to the sheer volume and diversity of types, formats, locations, and use cases. It can encompass **personally identifiable information (PII)**[22] stored in databases and files on endpoints, digital conversations saved in cloud environments, structured and unstructured data in data lakes and file stores, and more.

Since data is so vast and ubiquitous, agencies need a shared understanding of what data assets are in order to identify and protect them.

The Evidence Act defines data as "recorded information, regardless of form or the media on which the data is recorded" and a data asset as "a collection of data elements or data sets that may be grouped together."[23] The CISA ZTMM defines data as "all structured and unstructured files and fragments that reside or have resided in Federal systems, devices, networks, applications, databases, infrastructure, and backups (including on-premises and virtual environments) as well as the associated metadata."[24] Generally, Federal CDOs have taken the view that a data inventory is not just a "collection of data" or "recorded information," but an **intentional, meaningful categorization of data that serves a mission purpose and brings value to the agency.**

2.3: COMMON STARTING POINT TO IDENTIFY SENSITIVE DATA ASSETS AND BEGIN THE DATA CATALOG

As mentioned earlier, in many cases, agencies won't need to start their inventory from scratch — they may be able to build upon existing lists of data assets from other efforts highlighted above. Agencies must aim to organize these assets around a meaningful taxonomy based on the agency's mission and business areas. This will serve as the foundation for the data catalog and ensure that the data is organized in a way that is most useful for the agency.

While each agency will develop its unique taxonomy of data assets, a good starting point is to identify the mission-based information types. We recommend referencing NIST **Special Publication (SP)** 800-600 Rev. 2, *Guide for Mapping Types of Information and Systems to Security Categories*.[25] Data practitioners can also collaborate with the agency enterprise architects to use a Business Reference Model of Federal Enterprise Architecture,[26] if applicable.

---

[22] Per OMB Circular A-130, PII is defined as "information that can be used to distinguish or trace an individual's identity, either alone or when combined with other information that is linked or linkable to a specific individual." See https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/circulars/A130/a130revised.pdf.

[23] The Foundations for Evidence-Based Policy Making Act of 2018, https://www.congress.gov/bill/115th-congress/house-bill/4174

[24] CISA ZTMM, https://www.cisa.gov/sites/default/files/2023-04/zero_trust_maturity_model_v2_508.pdf

[25] Guide for Mapping Types of Information and Systems to Security Categories, https://doi.org/10.6028/NIST.SP.800-60r2.iwd

[26] Business Reference Model of Federal Enterprise Architecture, https://obamawhitehouse.archives.gov/omb/e-gov/FEA

2.4: CREATING THE DATA INVENTORY

The Evidence Act requires agencies to "develop and maintain a comprehensive data inventory that accounts for all data assets created by, collected by, under the control or direction of, or maintained by the agency," to the extent practicable.[27] The CDO Council has been leading the charge in identifying best practices for data inventories.

In April 2022, the CDO Council released the Enterprise Data Inventories Report,[28] which states:

> *"...a data inventory is foundational to any formal data management program. Data inventories enable stakeholders to efficiently find, access, and use data assets. Inventories are also indispensable to managers who need to evaluate the extent to which the organization's data helps meet mission goals or can be shared with other agencies to meet their missions."*

The CISA ZTMM echoes that a comprehensive data inventory is a factor in improving the ZT maturity of Federal agencies. Increasing the automation of the inventory can further boost an agency's ZT maturity from "traditional" to "optimal." Therefore, many agencies that have started their ZT journey may consider prioritizing the data inventory as an immediate action. Agencies can approach the inventory through manual or automated creation as outlined in Table 3 below.

TABLE 3: Approaches to Creating the Data Inventory

Manual Creation	Automated Creation
Agencies may find that starting with HVAs and FISMA High Impact systems makes the most sense. This is because agencies are likely to have already established these assets and systems in accordance with Federal laws and policies.[29] Agencies should capture relevant data asset taxonomy and metadata in commonly available tools (e.g., Excel). It is recommended that agencies use standard tools as a starting point until they acquire modern data catalog software.	Some agencies have invested in **machine learning data catalog (MLDC)** tools, which utilize advanced algorithms and techniques to automate capabilities such as data discovery, metadata extraction, and other data management activities. Many agencies may have existing technologies and data security tools within their **information technology (IT)** environment with sensitive data discovery and classification capabilities. Agencies can use the insights generated from these tools' data discovery capabilities to help develop a data inventory.

For more details about the steps for creating a data inventory, see *Appendix A: Data Inventory* in the companion document.

[27] The Foundations for Evidence-Based Policy Making Act of 2018, https://www.congress.gov/bill/115th-congress/house-bill/4174

[28] CDO Council Data Inventory Report, https://resources.data.gov/assets/documents/CDOC_Data_Inventory_Report_Final.pdf

[29] See OMB M-19-03: *Strengthening the Cybersecurity of Federal Agencies by enhancing the High Value Asset Program*, https://www.whitehouse.gov/wp-content/uploads/2018/12/M-19-03.pdf and FISMA 2014

2.5: STANDARDS FOR THE DATA INVENTORY

Since standards for data catalogs and inventories are evolving, agencies should be vigilant about development in this space. For example, an updated data catalog standard for the United States, DCAT-US v3.0 Schema,[30] is currently being developed. It builds on the DCAT standard, DCAT-US v1.1,[31] which Data.gov has used for over a decade. This standard, in addition to the **DATA Act Information Model Schema (DAIMS)**,[32] can help practitioners set internal data catalog standards for their agency. Additionally, NIEMOpen[33] — while not a standard for data inventories — can be used to help establish common vocabulary for easier exchange of information.

Security practitioners are encouraged to collaborate with their data management counterparts to gain a foundational understanding of the data inventory and related standards. Ultimately, security practitioners should learn to apply the data management discipline to manage security-relevant, non-business data, such as logs.

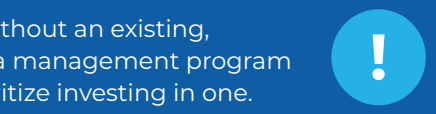
Agencies can find additional resources at https://resources.data.gov/standards/catalog/, a repository of data standards.

2.6: ROLES AND RESPONSIBILITIES

Before diving into the steps of data categorization, it is helpful to understand the roles and responsibilities of various parties involved in the process.

In agencies with mature data management programs, the data governance team typically develops the data catalog and data inventory. These agencies have established roles, such as business data steward and technical data steward, to create and maintain their data inventory, business glossary, and metadata library. Data stewards are great resources for identifying business risks. They should collaborate with their agency's security team to do risk-based categorization and labeling of their data assets. Data stewards should also collaborate with the agency's legal, records management, and privacy officials to establish the data inventory.

Agencies without an existing, mature data management program should prioritize investing in one.



While the implementation of CUI standards is underway in many agencies, a new role of information steward is also emerging. An information steward is similar to a data steward except that they steward unstructured data assets. They are the business experts likely to be best positioned to understand the risks of unstructured data assets and assign appropriate categories.

[30] DCAT-US v3.0 Schema, https://github.com/DOI-DO/dcat-us

[31] DCAT-US v1.1, https://resources.data.gov/resources/dcat-us/

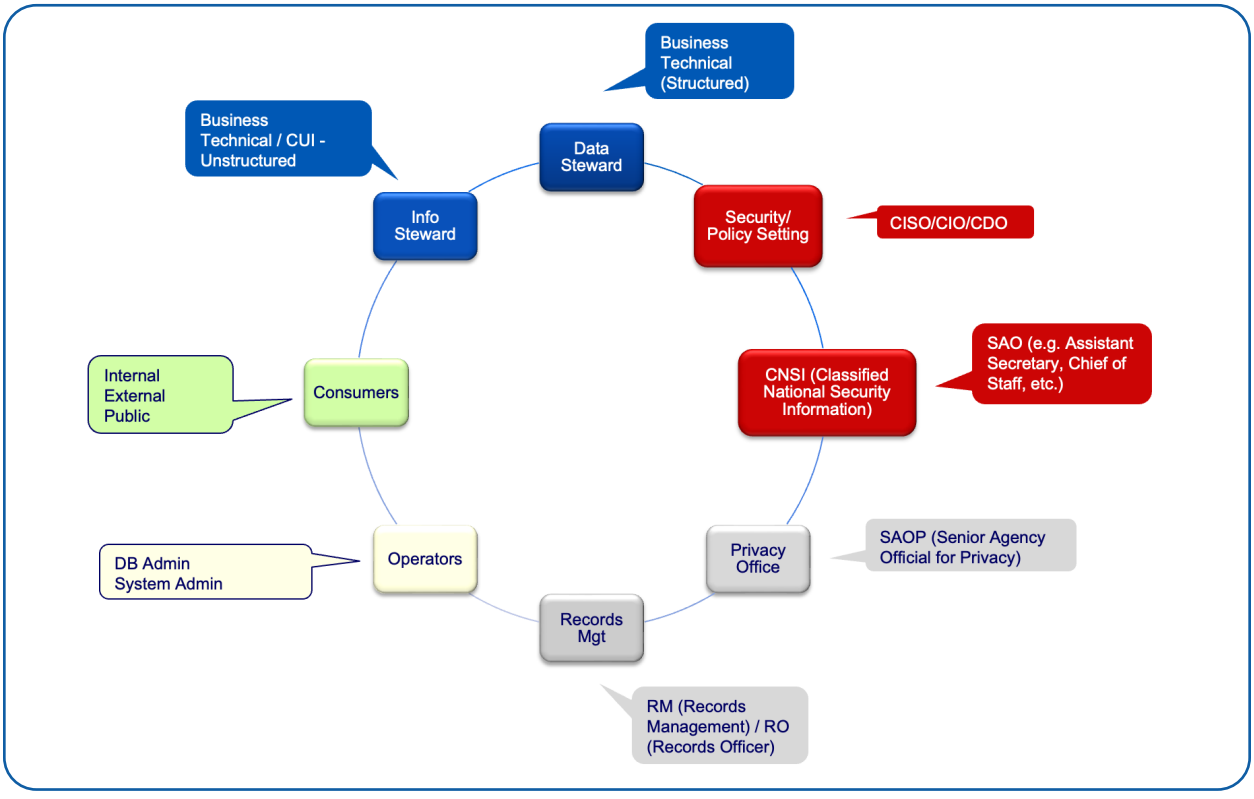
[32] DATA Act Information Model Schema (DAIMS), https://resources.data.gov/standards/catalog/daims/

[33] NIEMOpen, https://www.niem.gov/

Additionally, there are other key players in ZT data security as outlined in Figure 3. The following roles, bodies, or their designee(s) should likely be consulted in creating the data inventory and categorization:

- Agency Data Governance Body or Board
- Agency Records Officer or Senior Agency Official for Records Management
- **Agency Statistical Official (ASO)**
- CAIO
- CDO
- CIO
- CISO
- SAOP
- Senior Agency Official for CUI
- **Senior Agency Official for Geospatial Information (SAOGI)**

FIGURE 3: Zero Trust Data Security Players



Refer to *Appendix B: Data Stewardship* in the companion document for additional details about the role of the agency stewards.

2.7: CATEGORIZATION OF DATA ASSETS BASED ON IMPACT

The loss or misuse of data assets have different impacts on the mission of the Federal government. Understanding the impact of data's loss or misuse is the first step to understanding how data should be categorized and secured.

Agencies can refer to several well-known frameworks for categorizing data assets based on impact:

- **Federal Information Processing Standards (FIPS) Publication 199, *Standards for Security Categorization of Federal Information and Information Systems:***[34] Agencies should already be familiar with applying the standards set by FIPS 199 for semi-structured data assets, such as databases or XML-based data assets with underlying structure. Agencies can adapt this framework to apply to broader datasets, focusing on categorization of all information types housed throughout the system.
- **CUI:** Established by EO 13556 on *Controlled Unclassified Information*,[35] the CUI program is a great starting point for categorizing and labeling unstructured data — including emails, documents, spreadsheets, and text files — that are typically stored in a file system repository. Some agencies have found applying a common CUI labeling schema to structured or semi-structured data to be a streamlined approach.
- **NIST SP 800-60 Vol. 2 Rev. 1, *Guide for Mapping Types of Information and Information Systems to Security Categories: Appendices:***[36] This publication contains security categorization recommendations and rationale for mission-based, management, and support information types. It includes provisional impact level recommendations for various administrative, management, and service information.

[34] FIPS 199, Standards for Security Categorization of Federal Information and Information Systems, https://nvlpubs.nist.gov/nistpubs/fips/nist.fips.199.pdf

[35] EO 13556 on *Controlled Unclassified Information*, https://obamawhitehouse.archives.gov/the-press-office/2010/11/04/executive-order-13556-controlled-unclassified-information

[36] NIST SP 800-60, *Guide for Mapping Types of Information and Information Systems to Security Categories: Appendices*, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-60v2r1.pdf

2.8: DATA CATEGORIES AND LABELS

Agencies must develop or adopt methodologies to label their data assets. We recommend using the CUI sensitivity types as a starting point for labeling sensitive data, whether structured, semi-structured, or unstructured. Labeling should be done at the document level only and not at the data attribute level. Agencies can also use records management file plans that identify and categorize data according to the **National Archives and Records Administration (NARA)** Federal Records Schedule. For example, most Federal contracts are managed using **General Records Schedules (GRS)** 1.[37]

Label deployment: Labels should not be deployed to the whole agency all at once. The best practice is to deploy labels in an incremental fashion when a business need is identified.

- **Initial labels:** Labels should strive to address the largest percentage of data categories within an agency (e.g., privacy, financial).
- **Label scoping:** Some labels may only apply to specific agencies (via Active Directory Groups). Some organizations dealing with highly sensitive data may warrant special protection labels.
- **Methods:** Technologies enable manual and automated labeling. There can be a need for both methods, as well as challenges associated with each approach.
    - **Manual labeling** can be useful in instances where automated categorization and labeling is difficult to achieve accuracy due to the type of content or data. Manual labeling is challenging to scale, can have adverse impacts to employee workflows, and is prone to human error and manipulation.
    - **Automated labeling** capabilities are improving across tools today, though agencies must test this functionality to validate the accuracy of the categorization and labels. Automated labeling is increasingly necessary due to the growing volumes of data created today and in the future.

[37] General Records Schedules, https://www.archives.gov/records-mgmt/grs

CHAPTER 3: SECURE THE DATA

This chapter provides guidance on managing common data security risks. For the purpose of this guide, data security risks refer to potential threats and/or vulnerabilities that could compromise the confidentiality, integrity, and availability of an agency's sensitive data. These risks are amplified by the expanding capabilities of digital networks, AI, algorithms, and computational methods that enable data to be easily collected, combined, manipulated, and shared. Risk elements are outlined in Table 4.

TABLE 4: Risk Elements

Risk Elements	Data-Centric Guidance
Threat	A threat is any circumstance or event caused via unauthorized access, destruction, disclosure, or modification of data that the information system maintains. In other words, most threats target the data. Threats can be adversarial and non-adversarial.
Vulnerability	This is a weakness in an information system, system security procedures, or internal controls. While weaknesses are typically in the security controls that protect the data, they can also be in the data itself (e.g., lack of integrity, poor quality).
Likelihood	The probability of an adverse event or security incident depends on a combination of factors, such as: 1. Likelihood the event will occur. 2. Initiation by an adversary. 3. Likelihood that the event will result in adverse impacts. From the data perspective, the likelihood of an adverse event or security incident depends on the value of the data to the organization and its adversaries. An adversary is more likely to apply resources to higher value assets and data.[39]
Impact	The impact to an organization of an adverse security incident is based on the value and sensitivity of the data to stakeholders (e.g., the organization and its customers). See Chapter 2 for more information on assessing impact.

[38] CISA's HVA Program Management Office connects the authorities of OMB and CISA to identify the most valuable systems and harmonize a government-wide approach to protect HVA system functions and the information they contain. For more information, visit https://www.cisa.gov/resources-tools/programs/high-value-asset-program-management-office.

3.1: A BRIEF ANATOMY OF DATA SECURITY RISKS

Data security risks can be unintentional or intentional and stem from sources such as:

- External actors.
- Insider threats.
- Trusted third parties.
- Human/machine errors.
- Natural disasters.

Additionally, data security risks can stem from the data itself, such as issues related to data integrity, quality, and lifecycle management.

These risks have the potential to harm the national security and economic interests of the United States, so it's imperative that practitioners understand their roles and responsibilities and are held accountable for managing information security risk.[39]

Consider the following representative examples of data security risk types outlined in Table 5.

TABLE 5: Data Security Risk Types

Risk Type	Representative Examples
Risks **TO** the data *Risks to the confidentiality, integrity, and/or availability of data from external and internal threat actors, or from an agency's own IT practices*	• **Cybersecurity threats:** Ransomware, data extortion, insider threats, data breaches, eavesdropping, person-in-the-middle attacks, or accidental misuse of data • **Data transmission risks:** Insecure communication channels for data in transit, such as lack of secure encryption, improper use or configuration of encryption modules, hardware, ciphers, etc. • **Data storage risks:** Failure to properly encrypt data at rest, such as insecure or improper usage or configuration of encryption algorithms, cross origin sharing policies, or improper encryption key management • **Data resiliency risks:** Incomplete or untested backups, lack of redundancy, or loss in the availability and integrity of data • **DLM risks:** Improper handling of data throughout its lifecycle, including its initiation, development, implementation, operation, maintenance, and disposal • **Storage failure:** Data loss or corruption due to physical or logical failures of storage systems or devices • **Incomplete erasure:** Unauthorized use of residual data that is not fully erased from storage systems or devices

[39] These resources are a good starting point to help practitioners understand their roles and responsibilities: NIST SP 800-39, *Managing Information Security Risk-Organization, Mission, and Information System View* (https://csrc.nist.gov/pubs/sp/800/39/final); OMB Circular A-123: *Management's Responsibility for Enterprise Risk Management and Internal Control* (https://obamawhitehouse.archives.gov/omb/circulars_a123_rev); OMB Circular A-130: *Managing Information as a Strategic Resource* (https://www.federalregister.gov/documents/2016/07/28/2016-17872/revision-of-omb-circular-no-a-130-managing-information-as-a-strategic-resource).

Risks **FROM** the data usage *Risks based on how an agency uses or processes data*	• **Data governance risks:** Insufficient or inadequately enforced data governance may lead to non-compliance or mismanagement of data • **Data accountability and stewardship:** Improper data management or accountability, such as a lack of clear data ownership or failing to oversee proper data usage • **Unintended data usage:** Data being used for purposes other than its intended or authorized use, such as using data to instruct AI to manipulate people or systems through deepfakes • **Data aggregation risks:** Unintended consequences from combining multiple datasets, such as privacy violations or misinterpretation of data • **Processing errors:** Mistakes made during data processing, analysis, or transformation, leading to incorrect outputs or conclusions • **Algorithmic bias:** Flawed data can lead to automated data processing systems (e.g., AI) producing biased or unfair results • **Misinterpretation of data:** Incorrect conclusions are drawn from data analysis, leading to improper decision-making • **Privacy risks:** Exposure of sensitive information, such as CUI and PII
Risks **IN** the data *Risks based on the qualities of the data itself*	• **Data entry errors:** Incorrect or inconsistent data input leading to inaccurate or incomplete information • **Outdated data:** Use of obsolete or stale data that no longer accurately reflects the current situation, which may lead to improper decision-making • **Data timeliness risks:** Delayed or slow processing times that result in decisions being made without timely, comprehensive information • **Relevance risks:** Accumulation of unnecessary data or an over-reliance on historical data may distract from key insights • **Unverified sources:** Using data from untrustworthy or unverified sources that lead to false conclusions • **Unreliable data collection methods:** Using data collected through inconsistent or faulty methodologies
Risks **FROM SHARING** data *Risks related to sharing data with agency business partners and third parties, including suppliers with incidental access*	• **Third-party management:** Inadequate security practices by third parties may lead to violations of data's confidentiality, integrity, or availability • **Data misuse:** Data being used for purposes beyond the agreed upon use, potentially violating privacy or compliance requirements • **Data leakage:** Accidental or intentional exposure of shared data during transmission, processing, or storage • **Loss of control:** Diminished ability to monitor and enforce security measures once data is shared with external entities

?

DID YOU KNOW?

Data security practitioners should consult NIST's **Cybersecurity Framework (CSF)** 2.0[40] or the NIST **Risk Management Framework (RMF)**[41] for guidance on effectively addressing specific risk scenarios. However, practitioners should remain cognizant of additional data-centric risk scenarios that may exist outside of these frameworks.

[40] NIST CSF 2.0, https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf

[41] NIST RMF, https://csrc.nist.gov/projects/risk-management/about-rmf

3.2: WHERE TO BEGIN: BRINGING SECURITY TEAMS AND DATA TEAMS TOGETHER

Data teams and security teams can mutually benefit from a collaborative relationship. Data stewards can partner with their security counterparts to build their understanding of security risks and streamline data risk assessments. An agency's security team will likely have standardized tools and processes for assessing risks and may have already quantified risks for certain agency systems.

Likewise, security teams can benefit from data stewards' understanding of the scope of the data landscape and the totality of assets that must be secured. Data stewards should readily know the business value of data assets, which is crucial for accurate identification of data security risks and implementation of appropriate controls.

You don't know what you don't know, so it's essential to work together and bridge potential knowledge gaps to ensure that the agency is implementing a robust data security infrastructure that can mitigate potential data security risks.

You may be able to leverage existing channels for partnership. For example, you can foster collaboration through your agency's formal enterprise-level risk management program, if applicable.

3.3: RISK MANAGEMENT FROM A DATA SECURITY PERSPECTIVE

The NIST RMF provides a holistic process for managing cybersecurity and privacy risks to systems and organizations. Agencies are required to implement the RMF[42] and can further utilize the RMF's methodology to identify data security risks.

When applying the RMF process for data security risk identification, the primary roles will lie with the agency's security team, such as ZT or risk management practitioners. However, the security team should coordinate across their agency to identify common controls[43] and remain flexible in their approach as the roles of data management practitioners may evolve in a ZT data security model.

!

REMEMBER

While it's essential to focus on achieving data security objectives, it's equally important to consider those who will be impacted by your actions. Ensure you communicate with impacted users across your agency and allow sufficient time for feedback to uncover challenges and explore solutions.

See *Appendix D: Roles in Data Security Risk Management* in the companion document for a representative list of roles and responsibility guidance relevant to data security risk management.

[42] Federal agencies are required to implement the NIST RMF, per OMB Circular A-130 (and other guidance from OMB) in fulfillment of FISMA requirements for managing security and privacy risks.

[43] NIST RMF Prepare Task P-5 states that common controls can be identified at different levels of an organization, such as departmental, bureau, subcomponent, individual system, or program area.

3.4: THIRD-PARTY RISKS

Agencies rely on both inter/intra-agency partnerships, as well as commercial organizations, to leverage expertise and innovation to effectively execute priorities and missions. These relationships can result in globally distributed and interconnected network ecosystems spanning **information, communications, and operational technology (ICT/OT**). Third-party relationships are often long-term, trust-based, contractual, and integrated with an agency's daily workflows.

> This guide refers to all external entities as "third party" regardless of agency or commercial status, or relationship within the supply chain.

From a data-centric perspective, sharing data with third parties — whether intentionally or via incidental access — introduces additional risks to the agency with the potential to harm stakeholders, operations, reputation, and resilience. Therefore, agencies must consider the risk and appropriate controls throughout the lifecycle of the relationship with the third party. Learn about example use cases of third-party risk in Table 6.

TABLE 6: Example Use Cases of Third-Party Risk

Agency allows non-government third parties to view and use their data. In this use case, agencies retain full control over the data while allowing third parties to view and use the data.	**Agency contracts non-government third parties to manage their data.** In this use case, third parties act as the custodian for the agency's data assets. The agency may lose some or all control over the data after sharing it.
Agency sends data to non-government third parties for processing. In this use case, the agency loses control of its data after sharing it and relies on the third party to maintain appropriate controls and protections.	**Agency shares data with a variety of third parties, such as: a) other Federal agencies; b) U.S. state, local, tribal, and territorial governments (SLTT); c) foreign governments; etc.** In this use case, the agency loses control of its data after sharing it and relies on third parties to maintain appropriate controls and protections.

Common risks arising from third-party relationships:

- **Non-compliance with data privacy regulations:** Third parties may fail to comply with data privacy regulations, which may subject the agency to a fine or legal dispute.
- **Data residency and sovereignty:** Third parties may store or process data in locations that do not align with the agency's data residency or sovereignty requirements. This can raise concerns about legal jurisdiction, compliance, and control over the data.
- **Business continuity and disaster recovery:** If the third party doesn't have robust business continuity and disaster recovery plans, it can pose a risk to the agency's data availability and recovery in the event of a disruption (e.g., natural disaster, ransomware attack).

3.5: DATA SECURITY THROUGH THE PRIVACY LENS

In today's rapidly evolving digital landscape, **the preservation of individual privacy relies on the protection of sensitive data**. The Federal government's commitment to safeguarding this data is not only a legal obligation as outlined in Chapter 2, but a fundamental ethical responsibility, as exemplified in the **Fair Information Practice Principles (FIPPs)**.[44]

Privacy is a cornerstone of a comprehensive data security framework, ensuring that the collection, processing, storage, and dissemination of PII aligns with stringent privacy regulations and ethical considerations. This section will articulate the concerted efforts to navigate the intricate landscape of regulatory requirements, industry standards, and best practices pertaining to data privacy, and aims to guide agencies in the following actions:

- **Review the vast landscape of privacy frameworks requirements.**
    - Ensure compliance with pertinent laws, regulations, and standards governing data privacy.
- **Identify and mitigate privacy risks.**
    - Conduct thorough assessments to identify potential privacy risks and assess their impact.
    - Implement robust mitigation strategies to safeguard sensitive data.
- **Promote "privacy by design."**
    - Integrate privacy considerations into the fabric of your processes, systems, and initiatives.
    - Adhere to the principle of "privacy by design" to embed privacy measures from inception.
- **Collaborate with key stakeholders.**
    - Engage stakeholders, both internal and external, to foster a culture of privacy awareness, education, and collaboration.
    - Recognize that privacy is a collective responsibility.

[44] The FIPPs are a collection of widely accepted principles that agencies use when evaluating information systems, processes, programs, and activities that affect individual privacy. While the FIPPs are not requirements, they are principles that should be applied by each agency according to the agency's particular mission and privacy program requirements. For more information, visit https://www.fpc.gov/resources/fipps/.

3.5.1: INTEGRATING PRIVACY STANDARDS

Privacy standards play a crucial role in shaping data security practices within Federal agencies. They provide a set of guidelines that organizations must adhere to in order to meet legal requirements, ensure data privacy, and mitigate risks.

Practitioners should start by identifying and understanding the specific privacy laws, regulations, and standards that apply to their agency. Agencies may be subject to different laws and regulations based on the type of data they handle (e.g., health, financial, etc.). To ensure compliance, practitioners should know what each regulation authority requires in terms of data protection, privacy safeguards, access controls, reporting, and more.

Conducting a thorough review to understand what each privacy regulation requires can be time-intensive. It may be more efficient for data security practitioners to build strong partnerships and consult with their agency's privacy team in implementing ZT data security.

NIST offers various frameworks and publications that provide guidelines and best practices for cybersecurity, risk management, and data protection. For instance:

- The **NIST Privacy Framework**[45] is complementary to the NIST CSF and assists organizations in managing privacy risks by providing a structured approach to privacy management.
- **NIST SP 800-122, *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)***[46] provides guidance to protecting the confidentiality of PII in information systems.
- **NIST SP 800-53 Rev. 5, *Security and Privacy Controls for Information Systems and Organizations***[47] provides a catalog of security and privacy controls for Federal information systems and organizations.

Since these controls integrate with privacy, practitioners may find it useful to consult the Federal Privacy Council's security and privacy collaboration index and work with their agency's privacy team when deciding which controls to choose.[48]

[45] NIST Privacy Framework, https://www.nist.gov/privacy-framework

[46] NIST SP 800-122, *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*, https://csrc.nist.gov/pubs/sp/800/122/final

[47] NIST SP 800-53 Rev. 5, *Security and Privacy Controls for Information Systems and Organizations*, https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final

[48] Collaboration Index for Security and Privacy Controls, https://www.fpc.gov/assets/pdf/Collaboration%20Index%20for%20Security%20and%20Privacy%20Controls%20FINAL.pdf

3.5.2: PRIVACY PRESERVATION TECHNIQUES AND TOOLS

The following techniques and tools aim to protect sensitive information, prevent inadvertent disclosure of PII, preserve privacy, minimize the risk of data breaches, and align with the core tenants of ZT.

REMEMBER

It's important to consider appropriate scale — many third-party breaches occur as a result of an organization taking on more data than it can reasonably protect, which has further effects on the third party and the agency in data breach response. Anonymization and pseudonymization are generally used to reduce re-identification risk (i.e., the risk that PII will be revealed through the mosaic effect when combining datasets). Practitioners should note that anonymization does not alleviate all privacy concerns, and it's possible to over-collect anonymized information.

1. **Data minimization and anonymization:** Limit the collection, storage, disclosure, and processing of personal data to only what is necessary for the intended purpose. Anonymize or pseudonymize data whenever possible.
    - **Tools/techniques:** Anonymization tools, tokenization, data masking, k-anonymity, and differential privacy
2. **Encryption:** Use robust encryption methods to protect data at rest, in transit, and in use to prevent unauthorized access even if the data is compromised.
    - **Tools/techniques:** Strong encryption algorithms (e.g., **Advanced Encryption Standard [AES]**, **Rivest-Shamir-Adleman [RSA]**), encryption key management systems, secure communication protocols (e.g., **Transport Layer Security [TLS]**, **Secure Sockets Layer [SSL]**), and secure communications solutions
3. **Data masking and redaction:** Conceal specific portions of sensitive information to prevent unauthorized disclosure while maintaining usability for authorized users.
    - **Tools/techniques:** Redaction tools, dynamic data masking, and techniques to mask or hide sensitive information in documents or databases
4. **Privacy by design and default:** Integrate privacy considerations into the design and architecture of systems and processes by default, ensuring that privacy is a fundamental component of any new initiative.
    - **Tools/techniques:** Design systems with controls in place to protect the data before the system enters into operation, incorporate privacy-enhancing technologies, and adopt privacy-preserving architecture

5. **Privacy-enhancing technologies (PETs):** Leverage technologies explicitly designed to enhance privacy by safeguarding data and minimizing the risks associated with its processing and transmission.
    - **Tools/techniques:** Secure multiparty computation (SMPC), homomorphic encryption, and private information retrieval (PIR)
6. **Compliance automation:** Automate compliance checks and controls to ensure ongoing adherence to privacy regulations and standards.
    - **Tools/techniques:** Compliance and privacy management software, automated risk assessment tools, and policy enforcement mechanisms
7. **Collaboration and standardization:** Collaborate with industry and agency peers and standardization bodies to adopt common privacy-preserving techniques and tools that adhere to recognized standards.
    - **Tools/techniques:** Participation in industry forums, adherence to recognized standards (e.g., those from NIST, International Organization for Standardization [ISO]), and adoption of common best practices

Other approaches, such as computational isolation and confidential computing, can further support preservation of privacy, as outlined in Table 7.

TABLE 7: Computational Isolation and Confidential Computing

Computational Isolation (Clean Data Rooms)	Confidential Computing
Clean data rooms, also known as clean rooms or secure data rooms, can play a crucial role in ensuring security especially within the context of sensitive data handling in Federal agencies. These controlled environments — which can be physical or virtual — provide a secure space for managing highly sensitive information while minimizing the risk of unauthorized access or data breaches. Benefits include: • Secure data handling. • Confidential collaboration. • Regulatory compliance. • Data protection. • Risk mitigation. • Temporary data access. • Comprehensive security measures.	Confidential computing helps keep data safe and private by allowing sensitive information to be processed or analyzed while still encrypted. This means the data stays protected from unauthorized access, even from the system that's processing it. There are emerging technologies that could automatically secure data during processing, ensuring confidentiality, integrity, and privacy. Benefits include: • Protection of sensitive data. • Privacy preservation. • SMPC. • Homomorphic encryption. • Isolation and trusted execution environments (TEEs). • Confidential containers and secure enclaves. • Enhanced compliance and regulatory adherence. • Securing AI models. • Secure cloud and hybrid environments.

3.5.3: PRESERVING PRIVACY WHILE ENGAGING WITH THIRD PARTIES

Practitioners must establish robust protocols, guidelines, and contractual obligations to securely engage with third parties, such as other Federal agencies, U.S. SLTT governments, and foreign governments, and ensure that sensitive data remains protected and compliant throughout the entire supply chain. This may include:

1. **Vendor risk assessments.**
    - **Due diligence:** Conduct thorough assessments to evaluate the security and privacy practices of third parties, especially regarding how they handle sensitive data. Practitioners should also consider whether the third party has the capacity to handle and protect the quantity and sensitivity of data it would receive.
    - **Include privacy criteria:** Integrate criteria related to confidential computing, clean data rooms, encryption, and other privacy preservation techniques into the third-party evaluation process.
2. **Contractual obligations and agreements.**
    - **Data protection clauses:** Include specific clauses in contracts that outline the requirements for protecting sensitive data, specifying the use of confidential computing, clean rooms, encryption standards, and other privacy-preserving techniques.
    - **Compliance adherence:** Ensure third parties commit to complying with relevant U.S. regulations and standards that are applicable to the data they handle.
3. **Security and privacy standards.**
    - **Third-party guidelines:** Provide clear guidelines and standards to third parties regarding the secure handling, processing, and storage of sensitive data.
    - **Training and awareness:** Conduct training sessions or provide resources to educate third parties on best practices for safeguarding sensitive information.
4. **Regular audits and monitoring.**
    - **Continuous evaluation:** Implement regular audits and monitoring mechanisms to ensure that third parties adhere to the agreed-upon security and privacy standards.
    - **Access controls:** Monitor third-party access to sensitive data and ensure strict controls are in place to the extent practicable.
5. **Incident response and reporting.**
    - **Reporting obligations:** Establish clear reporting procedures for any security incidents or breaches and require immediate notification from third parties in case of any data breaches.
    - **Incident response:** Plan a coordinated, timely response in case of security or privacy incidents. Refer to OMB M-17-12, *Preparing for and Responding to a Breach of Personally Identifiable Information* for specific contractor responsibilities.[49]

[49] OMB M-17-12: *Preparing for and Responding to a Breach of Personally Identifiable Information*, https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2017/m-17-12_0.pdf

6. **Legal and compliance oversight.**

   - **Legal review:** Involve legal and compliance experts to ensure that contracts and agreements with third parties adequately address data security, privacy, and compliance requirements.

3.5.4: PRIVACY IMPACT ASSESSMENT

Conducting a **privacy impact assessment (PIA)** involves a systematic evaluation of how a project, system, or initiative might affect the privacy of individuals regarding the collection, use, retention, and disclosure of their PII. It's important to note that a PIA is *both* a process and a document.

For more information about PIAs, refer to the E-Government Act of 2002[50] and OMB M-03-22, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*.[51]

3.6: DATA SECURITY CONTROLS: IDENTITY, CREDENTIAL, AND ACCESS MANAGEMENT

The philosophy of ZT extends to ICAM, which plays a pivotal role in ensuring the confidentiality, integrity, and availability of data within Federal systems. ICAM encompasses a comprehensive set of policies, procedures, and technologies governing the management of user identities, their access privileges, and the associated activities within an organization's network.

Throughout this section, we will delve into the importance of strong authentication and identity verification, the principle of authorizing access based on least privilege, and the need for continuous monitoring. By understanding and implementing these ICAM principles and controls, agencies can establish a robust and secure environment that minimizes the risk of unauthorized data access, mitigates insider threats, and enhances their overall data security posture.

ICAM practices and technology controls are a primary means of enforcing data security policies for data access and enable the principle of least privilege, both at an information system level and a granular data level.

3.6.1: ACCESS CONTROL MECHANISMS

Access control mechanisms[52] are essential to ICAM with various means of implementation, such as **Context-based Access Control (CBAC), Role-based Access Control (RBAC),** and **Attribute-based Access Control (ABAC).** Table 8 explains these access control mechanisms, their pros and cons, and provides example scenarios for choosing the appropriate control. While practitioners should implement controls that best match the data that they are trying

[50] The E-Government Act of 2002, https://www.congress.gov/bill/107th-congress/house-bill/2458

[51] OMB M-03-22: *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*, https://obamawhitehouse.archives.gov/omb/memoranda_m03-22/

[52] Secure Cloud Business Applications: Hybrid Identity Solutions, https://www.cisa.gov/sites/default/files/2023-03/csso-scuba-guidance_document-hybrid_identity_solutions_architecture-2023.03.22-final.pdf

to protect, CISA has observed that access control decisions are best made using CBAC, which considers the context in which the access request is being made. CBAC combines features of RBAC and ABAC to apply dynamic access policies using device-level signals as cues.

TABLE 8: Access Control Mechanisms

Access Control Mechanism	Description and Guidance
Role-Based Access Control (RBAC)	Defines access privileges based on job roles and responsibilities. • When using RBAC, data owners and operators can use ISACA's step-by-step guide for establishing appropriate **separation of duties (SoD)**.[53] • Some products can help identify where SoD should be applied by using "role mining" techniques. Data owners and operators should ask if their organizations' ICAM capabilities offer support for role mining if they are having difficulty understanding how to create appropriate roles.
Attribute-Based Access Control (ABAC)	Generally used for more granular and dynamic access control based on user attributes and contextual information. • When using ABAC, you can consider attributes like user location, time of access and other factors to determine access privileges so long as the attributes are available at the time the decision is made. • The **National Security Agency (NSA)** and CISA recommend "utilizing a data tagging system or solution, where data is conditionally accessed via granular ABAC policies to protect data. It is also important to separate accounts that grant access to resources from those that manage them daily."[54]
Context-Based Access Control (CBAC)	Unlike traditional access control methods, CBAC continuously evaluates factors such as user identity, location, device health, and behavior to make precise access decisions. This method enhances security by ensuring that access is granted only under secure and appropriate conditions. • Ensure that CBAC is integrated into your broader ZTA. This means verifying every access request explicitly, using least privilege principles, and assuming breach to minimize potential damage.

[53] A Step-by-Step SoD Implementation Guide, https://www.isaca.org/resources/isaca-journal/issues/2022/volume-5/a-step-by-step-sod-implementation-guide

[54] Secure Data in the Cloud, https://media.defense.gov/2024/Mar/07/2003407862/-1/-1/0/CSI-CloudTop10-Secure-Data.PDF

3.6.2: ESSENTIAL IDENTITY, CREDENTIAL, AND ACCESS MANAGEMENT PRACTICES FOR PROTECTING DATA

While there are many factors that practitioners should evaluate when determining how to properly and appropriately protect data, this guide considers the ICAM practices outlined in Table 9 to be essential.

TABLE 9: Essential Identity, Credential, and Access Management Practices for Protecting Data

ICAM Practice	Description
Identity validation and verification	Involves the processes and technologies used to confirm that an individual's claimed identity is genuine and accurate.
Principle of least privilege	Mandates users, applications, and systems be granted the minimum level of access necessary to perform their functions. By limiting access rights, it reduces the risk of unauthorized access and potential security breaches, ensuring that users can only access the information and resources essential for their legitimate purposes.
User provisioning and lifecycle management	Involves the processes and technologies used to manage user identities and their access to resources throughout their entire lifecycle within an organization. This includes onboarding new users, managing changes in user roles, and deprovisioning users when they leave the organization.
Continuous monitoring	An ongoing, systematic process that utilizes automated tools to constantly observe and analyze an organization's IT systems and networks. The primary goal is to detect security threats, performance issues, and compliance problems in real time, allowing for prompt identification and resolution of potential risks.
Authentication	The process of verifying the identity of users, devices, or systems before granting them access to resources. It ensures that only authorized entities can access sensitive information and systems, thereby protecting against unauthorized access and potential security breaches.
Access control mechanisms	The methods and processes used to regulate who or what can view, use, or access resources within a system. These mechanisms are essential for ensuring the security and integrity of data by preventing unauthorized access and managing permissions based on user roles and contexts.
Federation and single sign-on	Authentication mechanisms are designed to simplify and secure access to multiple applications and systems. Federation allows users to access resources across different domains or organizations using a single set of credentials, while **single sign-on (SSO)** enables users to log in once and gain access to multiple applications within the same domain without needing to re-authenticate.

3.7: DATA SECURITY MONITORING AND CONTROLS

This section provides guidance on enhancing data security and implementing rigorous monitoring for data — whether it's at rest or in transit — within a ZT framework. We will delve deeper into the realm of data-centric security monitoring, logging, and alerting. These are crucial components in identifying and responding to anomalous or suspicious activities that could potentially compromise data security.

3.7.1: USE DATA-CENTRIC SECURITY CONTROLS TO SECURE DATA AT EVERY LEVEL, IN EVERY LOCATION

Cybersecurity technologies from firewalls to endpoint security tools are used to help protect data, but data-centric security controls consist of a specific domain of technologies that bring the controls closer to the data itself. Common examples of data-centric security controls include encryption, rights management, and **data loss prevention (DLP)**. Data access controls, such as use of RBAC, CBAC, and ABAC are used in conjunction with data-centric security controls to ensure appropriate levels of access and data use. The controls an agency selects to use should be chosen to mitigate the data risks identified, protect the data while enabling appropriate use based on its sensitivity level, and achieve specific objectives.

3.7.2: DEFINE POLICIES AND SELECT THE APPROPRIATE CONTROLS

Data security controls are used to enforce policies. Practitioners should define their policies based on data sensitivity, identifying who can access the data, under what conditions, and what they can do with it.

Access enforcement can limit a threat actor from interacting directly with a protected resource or data, while segment enforcement limits the ability for a threat actor to create further loss to the organization after breaching access and perimeter controls. Practitioners should expect to use a combination of various data-centric security controls and data access controls to address key objectives as outlined in Table 10.

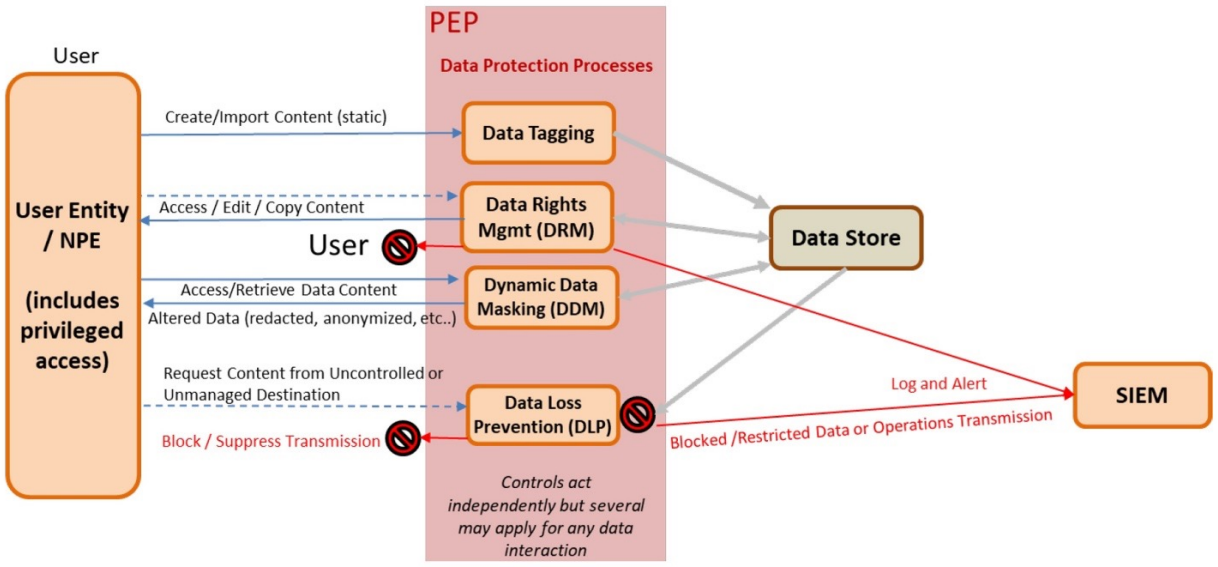
TABLE 10: Data-Centric Control Type Objectives

Control Type Objective	Example Technology Control
Reduce unauthorized access	Data access controls (e.g., RBAC, CBAC, ABAC), data access governance, rights management, data masking, data redaction, confidential computing
Reduce loss event frequency	DLP, data detection and response
Reduce loss magnitude	Encryption, tokenization, data masking, data redaction
Provide visibility into data and user activity	Database activity monitoring, DLP, data detection and response, **user and entity behavior analytics (UEBA)**
Preserve privacy	Privacy-preserving technologies, clean data rooms, confidential computing

For descriptions of example data-centric security controls, see *Appendix C: Security Monitoring and Controls* in the companion document.

Aligning technical policies with data access needs enables **policy enforcement points (PEPs)** to provide robust, data-centric protection. In implementing ZT, agencies would use a policy engine to deploy the policy and enforce the policy through a PEP. Access requests would be evaluated against the policy and adjudicated in real time. This approach ensures that sensitive data is adequately protected, supporting regulatory compliance, and reducing the risk of data breaches. See Figure 4 for an agency example of this alignment and use of data-centric security controls.

FIGURE 4: Data-Centric Security Protections



Source: Department of Defense Zero Trust Reference Architecture, https://dodcio.defense.gov/Portals/0/Documents/Library/(U)ZT_RA_v2.0(U)_Sep22.pdf

3.7.3: CONTINUOUSLY MONITOR FOR SECURITY CONTROL EFFECTIVENESS

Continuous monitoring capabilities across several key areas will help agencies stay on top of their changing data environment, ensure ongoing alignment between policies and controls, and respond quickly to suspicious activity, as outlined in Table 11.

TABLE 11: Continuous Monitoring Capabilities

Capability	How To Enable and Operationalize
Automated data inventory scanning: This ensures that the data inventory remains up to date. This capability typically includes tools and scripts to scan an organization's network, file systems, databases, and cloud repositories to discover data assets and perform initial identification.	• Data inventory scanning is often a feature within data security tools, referred to as data discovery functionality. Data repositories may also have this capability. It is also available as a standalone technology. • When new data assets are discovered or when existing assets change, use automated alerts to notify data stewards or administrators. This enables timely action to properly inventory and secure the data.
Data monitoring: This enables immediate alerts for potential security incidents. It includes real-time monitoring of data access, data activity and modifications, and user behavior. It also includes anomaly detection to identify suspicious activity.	• DLP capabilities are a common approach for data monitoring to alert on policy violations related to data movement. • Identify security analysts responsible for triaging and investigating alerts generated from monitoring tools. • With the use of monitoring technologies, ensure proper considerations and communications related to employee privacy. Develop procedures for an appropriate response to alerts, particularly as it relates to addressing insider risks.

3.7.4: LOG, AUDIT, AND ALERT TO MAINTAIN SECURITY AND ENABLE INVESTIGATIONS

Data-centric logging, auditing, and alerting are crucial for ZT, especially from the PEP. The PEP is responsible for enforcing the organization's security policies, and its logs are vital for maintaining security. Important logs from the PEP include:

1. **Access logs:** These logs record who accessed what data, when, and from where.
2. **Policy decision logs:** These logs track the decisions made by the PEP, such as allowing or denying access based on the policies.
3. **Change logs:** These logs track any changes made to the policies or system configurations.

Auditing involves regularly reviewing these logs to identify any anomalies or suspicious activities. Automated tools can help in this process by flagging activities that deviate from established norms or violate security policies. Audit logs track user actions and system changes to ensure accountability and traceability. They provide a chronological record of activities, crucial for audits and compliance checks. They may be used for evidence purposes in compliance, security, and computer forensic investigations.

Alerting is the process of notifying relevant personnel when such anomalies are detected. This allows for immediate action to mitigate potential threats.

The Role of the Security Operations Center in Monitoring and Response

Ideally, agencies will have a high-functioning **Security Operations Center (SOC)** to play a pivotal role in data security efforts. SOCs continuously monitor network traffic, user behavior, and system activities for any signs of anomalous or suspicious activity. They use advanced analytics and threat intelligence to identify potential threats that may otherwise go unnoticed. In the absence of an SOC or existence of an SOC with limited staffing, agencies may rely on a **managed security services provider (MSSP)** for this function. **Managed detection and response (MDR)** providers are an option for organizations to obtain 24/7 monitoring of their security events and augment existing SOC analysts. Alternatively, agencies can visit https://www.cisa.gov/resources-tools/services/security-operations-center-soc-optimization-advisory-service to learn more about their SOC Optimization Advisory Service.

DID YOU KNOW

Tabletop exercises are crucial for enhancing data security by simulating scenarios that test an agency's response to data breaches and cyber threats. These exercises allow teams to analyze how different security measures intersect and impact overall data protection strategies. By involving diverse stakeholders, agencies can identify vulnerabilities, assess the effectiveness of their security protocols, and develop comprehensive response plans. This proactive approach ensures that data security measures are robust, inclusive, and capable of addressing the complex challenges posed by modern cyber threats.

3.8: PRACTICAL EXAMPLE: STEPS FOR POLICY ENFORCEMENT CONTROLS

To effectively monitor data usage and detect any suspicious activities, technical practitioners should implement robust security measures. Here's a practical example to apply these principles:

Step 1: Prioritize known data types. Begin by focusing on well-defined data categories, as defined in Chapter 2, such as files containing PII like **social security numbers (SSNs)**, personnel forms (e.g., SF-50), or CUI. These data types are commonly targeted and should be secured first.

Step 2: Identify risks to mitigate. Determine the purpose for the security control and why it is necessary, alongside an understanding of how the data needs to be used or flow. For example, the agency needs to share files containing CUI with third parties, and employees should use the agency-approved, secure file-sharing solution to do this rather than email or public cloud storage services.

Step 3: Use security controls to enforce policies. Implement security controls to prevent data from being transferred from a secure environment to a less secure one. For instance, if data is being sent from an internal network to an external one (e.g., via email or to cloud storage services like Dropbox), it should trigger a security protocol. In this example, use DLP tools to automate detection and response to this policy violation.

Step 4: Adopt an iterative approach. Start with basic actions such as logging any data movement attempts. Gradually escalate the response to include user notifications, alerts to an SOC, data transfer quarantines, and — eventually — blocking unauthorized transfers.

See *Appendix C: Security Monitoring and Controls* in the companion document for an example implementation plan.

CHAPTER 4: MANAGE THE DATA

This chapter explains the importance of embedding ZT security practices throughout the data lifecycle to ensure data is securely managed across all stages and in the face of emerging technologies.

This chapter focuses on how to:

1. Develop a data management strategy with actionable roles and responsibilities.[55]
2. Prepare data for the adoption of emerging technologies, such as AI, to ensure responsible and effective implementation.
3. Establish appropriate data security controls for each lifecycle stage.
4. Provide practical approaches for data stewards and security professionals to collaborate on applying appropriate security controls.

4.1: BUSINESS VALUE OF SECURELY MANAGING DATA

Securely managing data is not just a technical necessity, but it is also a strategic business imperative. Consider these points to gain buy-in for secure lifecycle management at the outset rather than retrofitting security measures.

- **Organizational resilience and efficiency:** By implementing comprehensive data management strategies aligned with ZT principles, agencies can ensure that data is protected and readily available to authorized users, enabling a more agile and informed data use.
- **Legal compliance:** By implementing appropriate access controls and data categorization, agencies can demonstrate compliance with regulations and policies as highlighted in Section 2.2. This can mitigate the risk of data breaches and simplify audit processes.
- **Cost optimization:** By optimizing and streamlining data storage, reducing redundancies, and establishing a single source of truth, agencies can decrease their data management overhead and enable practitioners to efficiently access and leverage data assets.
- **Brand reputation management:** By demonstrating a commitment to data protection, transparency, and strategic use of information resources, agencies can enhance their reputation and credibility, ensuring appropriate access to the appropriate data at the appropriate time.

[55] See OMB M-25-05, *Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open Government Data Access and Management Guidance* for additional data management guidance.

4.2: MANAGING DATA SECURITY THROUGHOUT THE DATA LIFECYCLE

Table 12 elaborates on the lifecycle stages illustrated in Figure 2: Secure Data Lifecycle Management Is at the Heart of the Zero Trust Model, outlining suggested actions and providing a practical example for those actions. Practitioners should incorporate ZT principles while designing and implementing security measures throughout the data lifecycle.

As shown in Figure 5 below, ZT principles are applied in data lifecycle by ensuring continuous verification, least privilege access, encryption, and proactive governance at every stage of the data lifecycle, from creation to disposal.

FIGURE 5: Secure Data Lifecycle Management

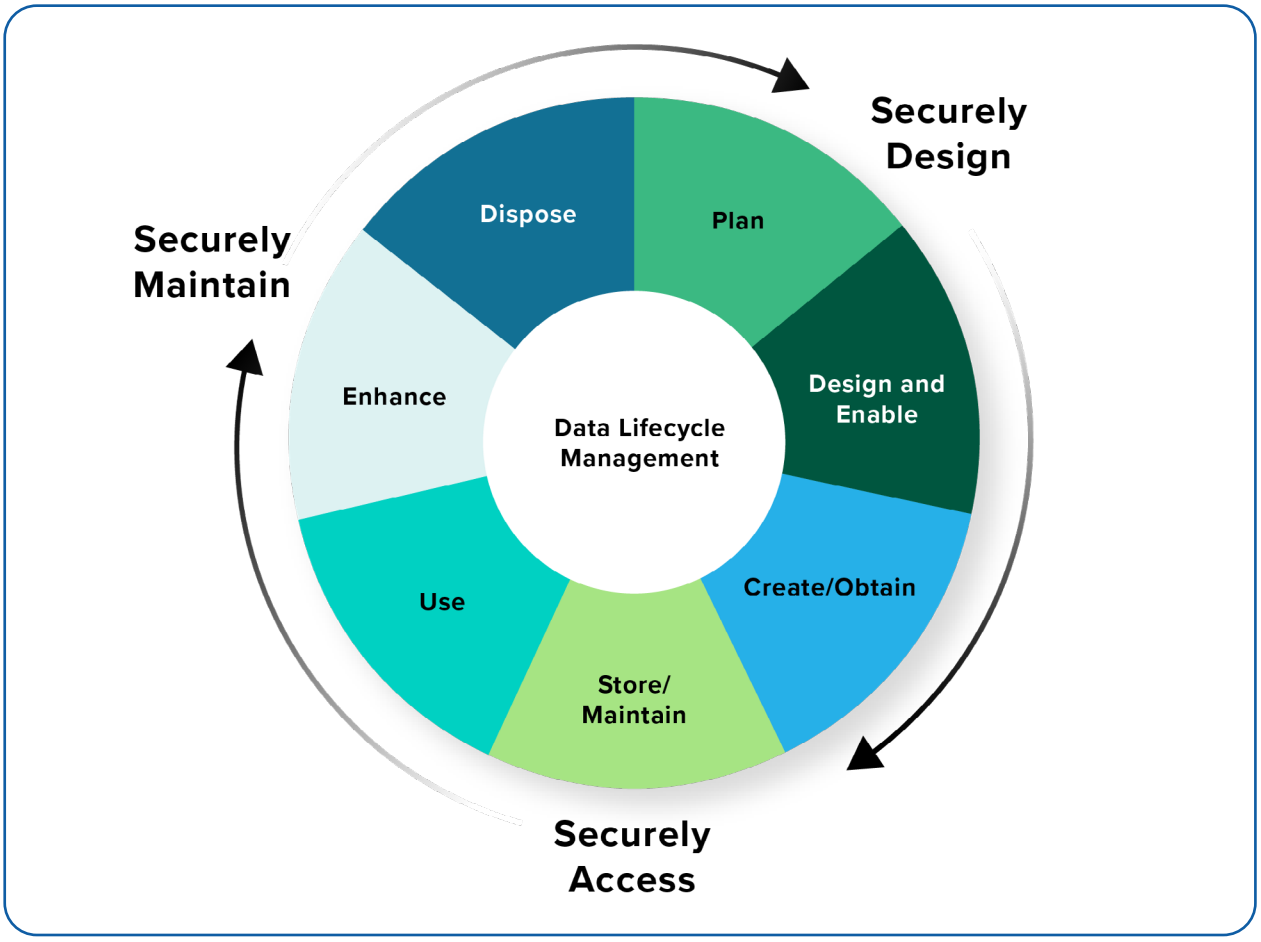


TABLE 12: Suggested Data Security Actions Throughout the Data Lifecycle With Practical Examples

	SUGGESTED ACTIONS	PRACTICAL EXAMPLE: AN AGENCY'S FINANCIAL DATA ASSET
Plan	Discover and document the design requirements of the data asset: 1. Discover and document the expected security and privacy of the data (see Section 3.5). Ask questions such as: • *Where will the data come from?* • *Will the data asset contain sensitive data or generally non-public data?* • *What are the availability and disaster recovery considerations?* • *Will this data be shared with or accessed by other agencies and/or third parties?* • *Are there any known mandates for sharing the data asset openly?* 2. Discover and document the nature of usage. Ask questions such as: • *Who will be allowed to access the data and why?* • *What are the least privilege access criteria?* • *Which location will users access the data from?* • *How will the data be used?* • *Should users be allowed to modify the data? If so, how will they modify the data?* 3. Conduct a risk analysis of the data asset (see Section 3.1). 4. Establish a preliminary determination of the data asset category (see Section 2.7).	A financial data asset, such as budget data, may contain both non-public (e.g., department-level personnel spending) and publicly shareable elements (e.g., major project spending). Disseminating public data while keeping non-public elements confidential requires careful design considerations, including categorization, risk analysis, and controls for effective separation.
Design and Enable	Design security measures based on the findings discovered above. You might take the following steps: 1. Design your metadata strategy and determine security-related metadata elements. Ask questions such as: • *How does the data flow? What is its lineage?* • *Who owns the data?* • *What are the appropriate data sharing/access rules and agreements?* • *How will you monitor who has access to the data?* • *Are quality requirements established?* • *What are the quality requirements?*	Establish tiered access and usage policies tailored to both public and non-public data elements. While aggregate spending data at the agency level may be made public, more detailed, granular data elements/metadata may need to remain confidential and require additional security controls.

Design and Enable	2. Determine storage architecture based on availability needs, including disaster recovery options. Ask questions such as: • *What are the security implications for a centralized vs. distributed storage architecture?* • *What are the built-in options for data security controls?* 3. Design security controls, including ICAM controls (see Section 3.6) and other data-centric controls (see Section 3.7).	Commonly used controls include encryption and data masking. Database platform providers often offer built-in tools to create access-controlled views, ensuring non-public data is protected from unauthorized access. Agencies might opt to maintain separate copies of public and non-public data to enhance security and simplify management, while ensuring synchronization of data elements.
Create/Obtain	Implement and test the controls. Focus on the following activities: 1. Implement the security controls and monitoring following the design decisions made in the previous stage. If the data is being sourced from outside, additional security measures, such as data validation and access controls, should be put in place. 2. Allocate resources for testing the security controls during the implementation stage of new data assets to address security flaws early.	Ensure security controls are effective for all use cases (public and private) of the agency financial data asset. Certify security controls before migrating the data asset to production. Continuously reevaluate controls as new data assets are derived from the financial data. For example, if creating training data for financial AI models by removing PII, update access and data masking controls accordingly.
Use/Enhance	Monitor the controls to prevent events such as access violation, data misuse, poisoning, and theft. Focus on: 1. Enforcing authentication and access control (see Section 3.6.1) 2. Monitoring continuously (see Section 3.7) 3. Continuously assessing the effectiveness of security controls (see Section 3.7.3).	Access to the financial data asset is strictly granted after applying policies informed by continuous, contextual, and risk-based verification. Access policies must be granular because of the co-mingling of public and non-public data elements of the data asset. Apply controls like data masking and monitor usage and activity.

Store/Maintain	Ensure secure storage and maintenance by focusing on: 1. **Managing redundancies:** Maintain a single authoritative source whenever possible to reduce the risk surface with the added benefit of reduced infrastructure cost. 2. **Securing backup and restore:** Backup across distributed environments like cloud and data centers is crucial for data resiliency and ransomware defense. Focus on: · *Regular **disaster recovery (DR)** tests to ensure recoverability within established timeframes and objectives.* · *A user-friendly centralized management and reporting capability for the backup.* 3. **Data storage controls:** Monitor security controls embedded in physical storage devices and cloud environments to protect data at rest. Examples include options for quantum-safe encryption and key management and compliance with data residency requirements for cloud workloads.	Agencies might choose to maintain separate copies of shared public data, ensuring they are synchronized in real time or near real time to reflect changes from the authoritative data source. Implement controls and monitoring during data transfer and sharing to protect against threats to data in transit. For data resiliency and to protect data at rest, implement controls and test recovery procedures.
Dispose	Ensure secure disposal of the data asset by focusing on: 1. **Data retention:** Monitor compliance with agency data retention policies. 2. **Data disposal:** Monitor compliance with agency secure data disposal policies.	Focus on the following example activities: **Data Sanitization:** Before disposing of IT equipment (e.g., laptops or servers), use certified data wiping software to overwrite all financial data, ensuring it cannot be recovered. The process is logged, and verification reports are retained for audit purposes. **Secure Erasure:** For SSDs or flash drives, secure erasure tools compliant with standards such as NIST 800-88 are used.[56] Certificates of erasure are generated and stored to demonstrate compliance. **Degaussing Media:** If financial data is stored on magnetic tapes or hard drives, use degaussing equipment to disrupt the magnetic fields, making the data unreadable. This step is documented in the asset disposal report. **Physical Destruction:** After data sanitization, the physical media (e.g., hard drives, tapes) are destroyed-typically by shredding or crushing-either on-site or at a certified facility. The destruction is witnessed, and a certificate of destruction is issued and kept for records.

[56] NIST SP 800-88 Rev. 1, *Guidelines for Media Sanitization*, https://csrc.nist.gov/pubs/sp/800/88/r1/final

Data Management Is Increasingly a Must-Have Competency for Cybersecurity Practitioners

As agencies implement ZT in diverse environments with varying infrastructures and an increasing number of devices, cybersecurity practitioners face the challenge of managing a flood of security-relevant data such as signals, telemetry, alerts, and log entries. This data is often unstandardized, making it hard to normalize and analyze.

To address these challenges, cybersecurity practitioners should collaborate with data management experts and understand the fundamentals of handling large volumes of data. At a minimum, the cybersecurity practitioners should adhere to the following best practices:

1. **Identify security data:** Know what data you have, its sources, and its flow. Treat most security data as sensitive and apply risk analysis and safeguards.
2. **Design an architecture:** Work with data management experts to improve your data management proficiency and understand security data architecture, storage, access, and analytics. Strategically use the built-in data management architecture of security analytics platforms.
3. **Implement robust security controls:** Use technical, administrative, and physical controls to protect sensitive data.
4. **Monitor and analyze security data:** Use security analytics platforms to monitor and analyze data.
5. **Adopt lifecycle practices:** Apply lifecycle practices (see Section 4.2) from creation to disposal. Develop and follow retention policies.

4.3: DATA STEWARDS ARE ENABLERS IN SECURING DATA THROUGHOUT ITS LIFECYCLE

As highlighted throughout this guide, ZT data security relies on close collaboration across stakeholders, with agency CDOs, CISOs, and CIOs facilitating the integration of data stewardship and security. This section elaborates on the role of agency data stewards in implementing ZT data security as described in Appendix B. Cybersecurity practitioners should leverage their agency's data governance body and data stewards' insights into the sensitivity and risks of specific data assets to design effective controls and monitor data access, modification, and flow within and across agencies and their partners.

Refer to *Appendix B: Data Stewardship* in the companion document for additional details about the role of the agency stewards.

Table 13 suggests roles and responsibilities for each lifecycle stage, which may vary by organization. To clarify roles, you can create a responsibility assignment matrix (also known as a Responsible, Accountable, Consulted, and Informed Chart or "RACI Chart") and **Standard Operating Procedures (SOPs)** to outline each role's contributions to managing the organization's data security.

TABLE 13: Zero Trust Data Security Roles Across the Data Lifecycle

Stage / Roles	Plan	Design and Enable	Create/ Obtain	Use (incl. Monitoring)	Enhance (incl. Monitoring)	Store/ Maintain	Dispose
Data Stewards	Lead	Co-lead	Lead	Co-lead	Lead	Lead	Co-lead
Cybersecurity Practitioner	Follow	Co-lead	Follow	Co-lead	Follow	Follow	Co-lead

Implications for Cybersecurity Teams

In a mature ZT data security environment, security teams serve as enablers for data stewardship by:

- Sharing knowledge to educate data stewards about the evolving landscape of data security technology and practices.
- Providing tools for secure data classification and monitoring.
- Automating compliance checks to reduce manual effort.
- Supporting stewards with threat intelligence to refine access controls.
- Make data stewards part of the incident response during data breaches.

This dynamic positions stewards as key stakeholders in defining security priorities while allowing security teams to focus on broader threat mitigation efforts.

4.4: DATA PROTECTION BOUNDARIES

Data protection boundaries delineate scope and limitations for how an agency can approach security controls when the agency does not have complete ownership over an environment. Understanding such boundaries can clarify relevant data risks, inform where to place and enable security controls, and identify who (or what) is responsible for those controls.

Table 14 provides a few examples of data-centric scenarios where data protection boundaries apply. It highlights approaches to risk mitigation that adhere to ZT data security principles outlined in Section 1.3, with particular emphasis on enforcing least privilege access and risk-informed security controls throughout the data lifecycle.

Note that the boundary of a data asset can be amorphous and not always aligned with the system boundary. Sometimes, data assets are an aggregation of multiple underlying data assets, such as a data fabric. Practitioners should consider this to design security controls that appropriately span the data asset's boundary.

TABLE 14: Example Scenarios of Data Protection Boundaries and How To Approach Controls

Scenario	Considerations
Protect data in a mobile and distributed work environment where employees are working from the office, home, or any other remote location	• Apply consistent verification processes for data access, regardless of user location. Limit access to sensitive data based on job roles. Ensure that only authorized personnel have access to specific information. • Implement secure remote access solutions, such as MFA and RBAC. • Continuously audit and monitor systems to track user activity. • Identify and implement specific procedures and technical solutions for **Government Furnished Equipment (GFE)** and **Privately Owned Equipment (PoE)** data access, regardless of user location. • Implement DLP tools to monitor and control the flow of sensitive data. • Use endpoint security solutions that can detect and prevent unauthorized applications or processes, such as unapproved email clients or file-sharing software.
Protect data stored and used by cloud service providers (CSPs)	• Delineate responsibilities and scope of control for the organization vs. CSP. • Assess built-in controls from the CSP versus areas requiring additional third-party controls, such as encryption, logging, or backup. • Implement **multifactor authentication (MFA)** to reduce the risk of unauthorized access. • Encrypt data both at rest and in transit to ensure that only authorized users can access it. • Use key management solutions to safeguard cryptographic keys and secrets. • Limit access to sensitive data based on job roles and ensure only authorized personnel have access. • Implement monitoring and auditing systems to track user activity and detect any suspicious behavior.

Scenario	Considerations
Protect data shared with third parties	• Determine required security controls and data storage, handling, and processing requirements based on the categorization of the data asset. • Include requirements in contracts that clearly define the terms and conditions of data sharing, including data usage, protection measures, and responsibilities of each party. Include confidentiality clauses to legally bind third parties to protect shared data. • Anonymize data where possible and anonymize or pseudonymize data before sharing to reduce the risk of exposing sensitive information. • Use data masking techniques to obfuscate sensitive data elements when sharing.

Extend the concept of data protection boundaries in your approach to include layered **supply chain risk management (SCRM)** boundaries. This will enable you to mitigate the risks associated with the products and services that are a part of your agency's supply chain, including cybersecurity products and services that you use to support the Data Pillar of ZT, as well as other pillars of ZT. Refer to NIST SP 800-161r1-upd1, ISO 28000:2022 for additional guidance.[57]

4.5: CONTINUOUS MONITORING AND RISK ANALYSIS

A key part of securely managing data is maintaining the overall health of the data's infrastructure and ecosystem to prevent unauthorized breaches and intrusions. Establishing a robust continuous monitoring and risk analysis process enables organizations to identify and respond to conditions or activities that could potentially compromise data security. This proactive approach includes regular risk analyses to identify potential risks, ensuring that vulnerabilities and anomalies are detected and addressed promptly. Table 15 describes different approaches and their objectives for continuous monitoring and risk analysis.

TABLE 15: Examples of Approaches and Objectives for Continuous Monitoring and Risk Analysis

Approach	Description	Objective
Attack surface management	Technology capability to continuously discover, inventory, and assess the context, relationships, control environments, and exposures of an entity's IT asset estate. Comprised of external attack surface management and cyberattack surface management.	Gain visibility into external- and internal-facing assets to better understand security posture. Enables assessment and validation of exposures to prioritize remediation.
Data security posture management	Technology capability to identify data risks and compliance gaps in cloud and on-premises environments based on attributes like data sensitivity, data flows, and configurations.	Gain visibility into data risks to identify remediation steps for improving data security posture.

[57] NIST SP 800-161r1-upd1, Cybersecurity Supply Chain Risk Management Practices for Systems and Organizations, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-161r1-upd1.pdf; ISO 28000:2022, Security and resilience — Security management systems — Requirements, https://www.iso.org/standard/79612.html.

Approach	Description	Objective
Database activity monitoring	Technology capability to track and identify anomalous database activity.	Gain visibility into database activities to detect and stop unauthorized access.
Data and operational resiliency testing	• Regular testing of backup, business continuity, and disaster recovery plans. • Sampling and testing of backup data to ensure usability and consistency with plans.	Ensure recovery processes and objectives can be met and identify potential gaps or impediments to meeting those plans.
Breach response testing	May include regular tabletop exercises of incident and breach response plans, simulation of actual breaches, phishing tests for security awareness, and reporting of suspicious activities.	Ensure incident and breach response plan objectives can be met and improve response. Engage in human risk management to train staff about threats and change behaviors to reduce risks.

4.6: BALANCING RISK AND REWARD: HOW TO PREPARE DATA FOR EMERGING TECHNOLOGIES

As agencies increasingly pursue adoption of AI and other emerging technologies in service to their mission, it is critical to revisit the role that data management plays in enabling and sustaining that innovation. The role of data as a strategic asset upon which AI systems are built makes the handling of that data key to successful AI adoption. Without robust data preparation, AI systems are prone to errors, biases, or otherwise unreliable outputs, which undermines agencies' ability to provide cost-effective and efficient innovation. Effective data management is crucial to ensure AI systems have access to high-quality, relevant, and accurate data that is fit-for-purpose for a given AI use case. For example, information about the source, provenance, selection, quality, and appropriateness of data — as well as the time period across which the data was collected — are all useful considerations for data preparation ahead of integration into an AI system. Techniques such as **retrieval augmented generation (RAG)** enhance the capabilities of a **large language model (LLM)** like ChatGPT by adding an information retrieval system that provides real world data via specified authoritative data sources. Here are some key aspects of this relationship:

1. **Data collection:** AI systems need vast amounts of data to train and enhance their performance, but not all data should be allowed to be used to train an AI system. Data management involves collecting, storing, organizing, labeling, exploring, and optimizing the data identified as suitable and appropriate as scoped for the intended AI use within the context of an agency's mission.

2. **Data quality:** The effectiveness of an AI system hinges on the quality, categorization, and labeling of the data it uses. Poor-quality data creates poor-quality models that propagate inaccurate results, which can introduce bias or otherwise negatively impact decision-making and business outcomes. Data management ensures that the data used by AI systems is accurate, complete, and up to date. Automated metadata curation, explainable data lineage, and comprehensive business glossaries have become even more critical with AI's advent.

3. **Data security:** As AI systems handle increasingly sensitive data, ensuring the protection of that data becomes paramount. Data security involves implementing robust security measures to protect against unauthorized access, theft, manipulation, or data loss as users interact with the AI system.

4. **Data governance:** Data governance involves setting policies and procedures for managing data within an organization, including how it is collected, stored, used, and shared with AI systems. There will be natural overlap between how an organization manages data governance and related AI governance across its systems, but it is possible that not all organizations will manage these interrelationships the same.

5. **Explainable AI (XAI):** XAI allows users to understand how an AI system arrives at its decisions or outputs. This transparency is crucial for building trust in AI systems and ensuring management of bias and other responsible use. Data management supports XAI by providing clear explanations of how data is used by AI systems and how these systems arrive at their conclusions.

6. **New demand on data:** AI has pushed organizations towards faster value creation from their data, often in real time. Data platforms have evolved rapidly, solving for interoperability by integrating previously disjointed tools and platforms into unified systems where data movement latencies are minimized.

7. **New types of data**: AI has expanded the types of data that need to be managed beyond traditional tables and spreadsheets to include images, notifications, audio, geospatial data, IoT sensor data, and vector data. This includes remote sensing, synthetic aperture radar, and visual data generated by **light detection and ranging (LIDAR)** technologies, both visible and invisible.

8. **Ethical considerations and bias mitigation**: Generative AI systems can inadvertently perpetuate or amplify biases present in training data and its labeling or categorization. Federal agencies should prioritize ethical data management practices to ensure fairness and prevent discrimination in AI-enabled decision-making processes. This involves implementing rigorous data auditing procedures, diverse data collection methods, assumption analysis and continuous monitoring, and the evaluation of AI outputs for potential biases.

9. **Continuous learning and model updating:** Generative AI models require ongoing refinement to maintain accuracy and relevance. Federal agencies should establish processes for continuous data collection, curation, and model re-training. This dynamic approach to data management ensures that AI systems evolve alongside changing mission requirements and evolving challenges.

CHAPTER 5: CONCLUSION

5.1: RECAP

This guide outlines an approach to implementing ZT data security within the Federal government. We began by establishing the vision and principles of ZT, emphasizing the need to view data as the new perimeter. We focused on how to define data using various types, classifications, and categories before expanding on how to secure data with the appropriate security controls.

5.2: BUILDING THE FUTURE OF DATA SECURITY

As we look to the future of data security, it's clear that we must embrace the following to be successful:

- **Cross-functional collaboration and effective communication:** Cybersecurity is a collective effort that goes beyond agency IT and cybersecurity teams — it requires collaboration across all roles at all levels. Effective communication between teams helps ensure alignment and promote interconnectedness, fostering a culture of security where everyone contributes to the collective effort of protecting the organization's digital assets. This also enables continuous improvement as different roles can learn from each other, share best practices, provide feedback on existing policies, and collaborate on new security initiatives.
- **A cooperative relationship between data and security teams:** Fostering a CoP between data stewards and security practitioners is paramount. A persistent, collaborative environment encourages the exchange of knowledge, promotes the development of innovative solutions, fosters the development of a culture of security, and strengthens an agency's overall security posture. CoPs play a crucial role in this process, serving as a platform for continuous learning and improvement, thereby ensuring that the principles of ZT are not statically implemented, but evolve with the changing cybersecurity landscape.
- **Continuous learning and education:** Since EO 14028 and OMB M-22-09, there has been a significant increase in ZT training available to multiple audiences. Agencies and departments should invest in training and certification programs to develop cross-functional teams with common lexicons and language. This applies to all levels of the Federal workforce, particularly senior leadership, to enable them to make informed decisions that balance the benefits and risks of data-driven initiatives. Additionally, new resources and frameworks are being developed that offer innovative ways to protect and analyze data. Practitioners should endeavor to stay informed about the latest developments. By fostering a culture of continuous learning and awareness, agencies

can equip their personnel with the necessary skills to navigate and counter evolving cyber threats.

- **Intersectional analysis and assessments:** Further consideration should be provided to the data that departments and agencies already collect, supporting greater fidelity of data and associated attributes.
- **Across-the-board buy-in:** It's crucial to gain support from policy officials and senior leaders, Congress, and other stakeholders to adequately prioritize funding and resources for ZT data security efforts. Stakeholders should engage in multi-year planning to ensure their agency can keep up with the pace of cybersecurity over time.
- **Adaptability:** Practitioners need to find the right implementation strategy that meets the unique needs of their agency. Implementation strategies should be approached from both a strategic and a tactical perspective. There is no one-size-fits-all solution or tactical level implementation strategy.

In today's interconnected world, all devices will — or already do — seamlessly communicate with each other, **expanding the threat surface of an already vast data ecosystem.** Therefore, the implementation of ZT principles is paramount for the Federal government to safeguard its data assets in an increasingly complex and contested cyber environment. **By adhering to the core tenets of ZT — never trust, always verify, and assume breach — agencies can ensure that their data is categorized and handled with the utmost precision and care.**

Agencies must ultimately embrace innovation and collaboration to mature the data security architecture of the Federal enterprise and build resilience against future threats.


federal
CDO
council
CISO
Council
CIO.GOV