**Data Classification Guide**

Identifying the goals, processes, and benefits of data classification



Office of the Chief Information Security Officer, State of Texas

Texas Department of Information Resources

300 W. 15th Street, Suite 1300

Austin, Texas 78701

Version 1.1 | July 6, 2018

##

## Acknowledgments

Appreciation is offered to the following individuals and their organizations for their cooperation and support.

Lona Chastain Texas Workforce Commission

Kent Dyer Texas Workforce Commission

Sarah Jacobson Texas State Library and Archives Commission

Sean Miller Railroad Commission of Texas

David Morris Texas State Soil and Water Conservation Board

Jim Nolan Texas Comptroller of Public Accounts

Nancy Pleasant Texas Comptroller of Public Accounts

Michael Reagor Texas State Library and Archives Commission

Laura Russell Texas Parks and Wildlife Department

Ruth Soucy Texas Comptroller of Public Accounts

|  |
| --- |
| Contents[Acknowledgments 2](#_Toc518646267)[Executive Summary 1](#_Toc518646268)[Background 1](#_Toc518646269)[Benefits of Classifying Data 2](#_Toc518646270)[Proposed Solution 3](#_Toc518646271)[Sample Security Controls 4](#_Toc518646272)[Implementation Guidance 4](#_Toc518646273)[Proposed Data Classification Taxonomy 4](#_Toc518646274)[Version History 5](#_Toc518646275) |

## Executive Summary

Data classification is the process of categorizing data into various types, forms, sensitivity level, or any other grouping of similar characteristics. When a piece of information (e.g., a document, memo, or customer record) is created, the owner assigns a standard classification level which defines the prescribed handling requirements for that piece of information, among other things. Such categories dictate the controls necessary to best protect the confidentiality, integrity, and availability of the data.

Data classification makes securing data much more efficient, because it instantly identifies and communicates the minimum level of protection required for any piece of data as well as the audience that may view it. For example, a document that is classified as "confidential" is easily understood to require additional protections and controls.

The Office of the Chief Information Security Officer at the Texas Department of Information Resources (DIR) worked with a taskforce of agency stakeholders to develop a model data classification taxonomy for state agencies and institutes of higher education. The classification scheme is detailed separately from this guidance document. This document is meant to present the background, underlying assumptions, and logic behind the decisions the taskforce made in arriving at this model.

## Background

Texas Administrative Code (TAC) Chapter 202 requires all agencies and institutions of higher education to classify their data.[[1]](#footnote-1) However, TAC 202 does not explicitly define classification levels beyond the “confidential” category.[[2]](#footnote-2) The lack of standardization in data classification schemes across the state creates challenges such as inefficiency in communications, discrepancies in controls applied between agencies, and in rare cases, a neglect to implement data classification policies and procedures entirely. To address these challenges, the Office of the Chief Information Security Officer (OCISO) worked with representatives from multiple state agencies to develop a baseline data classification scheme that can be adopted and modified to meet the varying needs of agencies and institutions of higher education.

Based on the experience of these representatives and their understanding of security standards and best practices, the OCISO proposes a simple classification scheme for all agencies to consider. The representatives based their classification scheme on current Texas law, both 1 TAC 202 and the Public Information Act, as well as the relevant federal standards (FIPS 199, NIST SP 800-59 and 800-60).

|  |
| --- |
| The labels used in this data classification scheme are in no way meant to subvert, contradict, supplant, or conflict with the Texas Public Information Act. In all cases, the public release of agency data is governed by the Texas Public Information Act and Chapter 552, Texas Government Code. The data classification scheme presented in this guide is intended to be a means to identify and address the safeguards, precautions, and handling requirements necessary to prevent accidental data disclosure.  |

## Benefits of Classifying Data

Data classification is the basis for identifying an initial baseline set of security controls for information and information systems, which creates numerous benefits for the organization.

Effectively classifying data makes security decisions more efficient for employees, data owners, and IT staff, because it instantly identifies and communicates the level of protection required for any piece of data and who can access it. Establishing a common statewide vernacular can further amplify this efficiency through clear and non-ambiguous communication.

Appropriate data classification can also enable a more efficient use of IT capital. Specifically, data that has been categorized at a level requiring more protection can provide an objective justification for certain capital expenditures to help protect that data.

An organization can design its systems architecture with varying information sensitivity levels in mind if there is an awareness of the location, type, and handling requirements of the data. This may assist in achieving economies of scale with security services and protection through shared network and security zones. For example, an information system containing information protected by state privacy laws may be stored with other information systems containing similar sensitive information which are regulated by a third-party agreement.

Agency contingency and disaster recovery planning personnel can use the outputs of the data classification process to ensure that the infrastructure is sufficiently protected and that recovery efforts focus on high impact systems.

Finally, artifacts of a data classification process can also serve as inputs to Business Impact Analysis (BIA) reviews, Information Sharing and System Interconnection Agreements, and audit trails.

## Proposed Solution

The proposed data classification scheme outlines four classification labels.

* *Public* – Information that is freely and without reservation made available to the public.
* *Sensitive* – Information that could be subject to release under an open records requests but should be controlled to protect third parties.
* *Confidential* – Information that typically is excepted from the Public Information Act.
* *Regulated* – Information that is controlled by a federal regulation or other third-party agreement.

### Public

The *Public* information label is used for information such as published reports, press releases, and information published to the agency’s public website. Such information requires no authentication and is freely distributable by all agency personnel.

### Sensitive

Moving the *Sensitive* label, much of the information is still subject to public release under an open records request, but the information should be vetted and verified before release. These types of data include items such as employee records and gross salary information. While these records and information are considered “public” under the Texas Public Information Act, they should still be afforded a higher level of protection to ensure confidential data (e.g., net salary information) is not comingled. Many agencies will choose to release this type of information only through select employees who are familiar with the state and federal rules regarding disclosure.

### Confidential

The *Confidential* label is used to identify information that is typically excepted from public disclosure, whether specified in law or through a decision by the Open Records division of the Texas Office of the Attorney General. Confidential data include information such as attorney-client communications, protected draft communications, and computer vulnerability reports.

### Regulated

The fourth label, *Regulated*, may or may not be applicable to an agency, based on its mandate, customers, and business operations. Regulated focuses on the types of data typically regulated by federal statute or third-party agreements. Agencies that maintain protected health, federal tax, payment card, or certain personal information will have specific requirements placed on that data by a non-Texas regulation. Therefore, regulated data has specific handling requirements that are unique to their regulations and do not apply to all agencies.

Often in data classification projects, the adage “the perfect is the enemy of the good” can impede implementation. The data classification scheme presented is not perfect for every agency or every occasion. The workgroup, however, feels it should be considered a good starting point to begin a data classification and handling program within an organization. As an agencies data classification practices mature, the classification scheme may change or evolve over time.

## Sample Security Controls

Included in the data classification template are *Sample Security Controls*. These controls are meant to be a reference and starting point for agencies to build from. As an example, some agencies may fully adopt the *Roles and Responsibilities* section, others may decide (through documented risk acceptance) that they do not need to implement encryption to protect fixed media. The data classification template serves as a starting point to help identify areas that should be a part of a mature data classification and handling program.

## Implementation Guidance

As mentioned several times, the template provided should be considered a starting point for internal agency deliberation. Not all agencies will be prepared to implement all the handling requirements or have the discipline and resources to classify all data at an elemental level. The following are a few tips for implementing a data classification program within an organization.

* Assess the readiness of the organization to accept data classification as a standard process. If the organization has a mature culture of security, the data classification scheme and handling requirements can be more detailed. For agencies just starting the data classification process, a simpler scheme with fewer handling requirements can help in gaining traction.
* As part of the readiness assessment, key influencers and executive staff must be involved early in the process. Individuals who are part of the planning and development of the strategy are more likely to support it during implementation.
* Build data classification into the agency’s System Development Lifecycle. An initial data classification should occur during development, guiding the security controls that must be implemented.
* For agencies that are unsure they can classify their data directly, try classifying networks instead of the data. All networks should be classified at the highest level of data it contains, so if a network contains sensitive data, then it should be classified as a sensitive network. The network classification then will mandate the type of security controls that the network must possess. As the data classification program matures, the agency can get more granular with the elements it classifies.

## Proposed Data Classification Taxonomy

The accompanying spreadsheet is a template presenting the four data classification labels discussed in this guide. It provides sample security controls for each classification level. Refer to the template for the proposed taxonomy and use it as a starting point to identify areas that should be a part of the agency’s data classification system.

## Version History

Current tools are available on the Texas Cybersecurity Framework website.

| Release Date  | Description |
| --- | --- |
| 25-Mar-2014 | Version 1.0 of the Guide and Template released. |
| 06-July-2018 | Version 1.1 of the Guide and Template released. |

1. 1 TAC 202.24(b)(1): State agencies are responsible for defining all information classification categories except the Confidential Information category, which is defined in Subchapter A of this chapter, and establishing the appropriate controls for each. [↑](#footnote-ref-1)
2. 1 TAC 202.1(5): Confidential Information – Information that must be protected from unauthorized disclosure or public release based on state or federal law (e.g., the Texas Public Information Act, and other constitutional, statutory, judicial, and legal agreement requirements). [↑](#footnote-ref-2)