Data Classification Toolkit

Last reviewed: July 2015

Purpose

To compile resources pertaining to data classification in higher education. Although data classification is just one component of a comprehensive program to protect data, it is an important foundation. This Toolkit consolidates resources from the EDUCAUSE web site as well as from other sources, and organizes them into five basic steps.

Introduction

Data are some of the most valuable assets any institution of higher education owns, and, as is in the case with all valuable assets, they need to be protected accordingly. What constitutes "accordingly" is mostly driven by legal, academic, financial and operational requirements and is based on the criticality and risk levels of the data. Protecting data assets while supporting academic, medical and research missions that require collaborative work and the open sharing of knowledge can be a difficult balancing act. One of the most important steps in protecting data appropriately is to determine classification levels for the data, and then to proceed with the actual classification of all of your valuable data assets.

The objective of the Data Classification Toolkit is to provide a body of information, resources, and guidance that can assist higher education officials in addressing the following questions regarding classifying data:

1. Need: Why is it necessary or mandatory to classify data?
2. Roles: Who should classify what data?
3. Methods: How should data be classified?
4. Are there any best (or common) practices available?
5. Impact: What processes are dependent or impacted by data classification?

Steps

- Step 1: Determine the need and/or requirements for data classification
- Step 2: Determine the roles involved in data classification
- Step 3: Determine your institution's classification levels
- Step 4: Determine the methodology and procedures for classifying data
- Step 5: Determine and review other information security processes impacted by data classification

Step 1: Determine the need and/or requirements for data classification

Sub-Step	Tips	Resource	Resource Type
1.1 There are laws, regulations, rules, or policies (federal, state and/or institutional) that *require* classification of data. a. These might be **Federal** regulations, such as:			
• Health Insurance Portability and Accountability Act (HIPAA) Security Regulations 45 CFR Parts 160, 162, and 164		HIPAA Security Standard - Final Rule	Government
• Health Information Technology for Economic and Clinical Health (HITECH) Act Section 13405(c), which expands an individual's right under HIPAA		HITECH Act Section 13405(c)	Government
• Family Educational Government Rights and Privacy Act (FERPA) Regulations CFR Part 99		FERPA	Government
• Department of Health and Human Government Services (HHS) Title 45 CFR Part 46 Protection of Human Subjects.		HHS Title 45 CFR Part 46	Government

b. There are probably **State** laws with regard to personal information, such as SSN, which, if compromised, would lead to consumer notification; there may also be rules regarding performance of periodic risk assessments, and the protection of such data accordingly.			
• Texas Administrative Code (TAC) Part 10 Chapter 202		TAC 202	Government
• Michigan Social Security Privacy Act		Michigan Social Security Privacy Act	Government
• New York State Breach Notification Law		State of New York Information Security Breach and Notification Act	Government
• Massachusetts Standards for the Protection of Personal Information of Residents of the Commonwealth (201 CMR 17.00)		Massachusetts 201 CMR 17.00	Government
c. These might be **Institutional** policies.			
1.2 In higher education institutions, where data stewardship is usually decentralized and where data is held by autonomous units, more guidance is needed than would be in an organization where data is completely centrally held and controlled.	Policies and guidelines need to be very clearly worded for very different populations. For example: Faculty members and colleges may hold student grades. Researchers, who typically work with their own data and equipment, may need to meet information security requirements from granting agencies. In addition, data classification applies to all data at the institution, including not only centrally collected and managed data, but also data collected and managed within schools and departments.	University of Florida Data Classification Guidelines	Higher Education
1.3 Data classification assists in risk management.	Risk assessments help organizations and/or departments determine the levels of security needed to safeguard their data appropriately and the best way to allocate scarce budget and staff resources. Almost no organization can afford to apply the highest levels of security to all data. Although risk assessment can be performed without formal data classification, formally classifying data enables organizations to prioritize which systems receive the most security resources, and thus manage risk appropriately.	Risk Management Framework, Phase 0: Strategic Risk Assessment Planning	EDUCAUSE

Top of page

Step 2: Determine the roles involved in data classification

Sub-Step	Tips	Resource	Resource Type
2.1 A number of terms are used to define the various roles and responsibilities.	It is common for people to assume that since IT manages the system, IT owns the data, but this is dangerous since IT is not responsible for the function that uses the data. However, it can be difficult to get the appropriate owners to take on the responsibility.		
a. At the governance level, terms such as "Data Trustee," "Data Steward," or "Data Owner" are common.	Start by asking these questions: 1) Who is the manager or agent responsible for the function that uses the resource? 2) Who is responsible for determining levels of protection for the data? 3) Who is responsible for making decisions about appropriate use of the data? 4) Who is responsible for classifying the data? 5) Who is responsible for business results of the system or the business use of the information? Most universities declare that the data is owned by the unit responsible for creating the data itself, such as Registrar, HR, etc. But not all data has such a clear "creating" unit, since in many cases, individuals actually "create" the data themselves (e.g., online student systems where students enter much of their own data, online human resource systems where employees enter much of their own data, course management systems where professors and students create much of the data, and email systems where the user or sender creates the data.)	EDUCAUSE Confidential Data Handling Blueprint: Step 3.1, Data Stewardship Roles and Responsibilities	EDUCAUSE
		Presentation: Who Owns the Data Anyway? Defining Data Stewardship	EDUCAUSE
		Indiana University Data Management Policies and Guidelines	Higher Education

		University of Texas at Austin Glossary Entry for Data Owner	Higher Education
b. At the management level, terms such as "Data Custodian" are common.	These are the persons who are responsible for implementing the controls the owner identifies. Many places now distinguish between the Dean/Director/AVP who is ultimately responsible for the data and the person who supervises the data entry personnel (and is thus quite a bit lower down the hierarchy).	Cornell University: Policy 4.12 Data Stewardship and Custodianship	Higher Education
c. At the operational level, terms such as "Data Custodian" or "Data User" are common.	Identify those who actually "touch" the data (enter, delete, even read).	The Ohio State University Institutional Data Procedures: Roles and Responsibilities For Data Trustees, Data Stewards, Data Custodians, and Data Users	Higher Education}

Top of page

Step 3: Determine your institution's classification levels

Sub-Step	Tips	Resource	Resource Type
3.1 Typically a number of "data classification levels" are identified by the institution.	Keep it as simple as possible - don't create any more levels than you have to. Each level should be differentiated from the other by the different actions required to appropriately handle the data.	Data Classification, Security, and Compliance: Helping Users Help Themselves (University of Michigan)	Higher Education
a. The levels are given appropriate names and definitions, and then each data element is classified into the proper level. Universities differ on how many levels are defined, although the most common number is three, four, or five.	Use names that are very clear to users, for example, "restricted" and "sensitive" are very similar terms and would cause confusion if used for a medium and high level, respectively. Keep the highest level very high, because this level will cost a lot to secure. Examples (from lowest to highest level): 1) Public, Restricted, and Private 2) Public, Sensitive, and Confidential 3) Category III, Category II, and Category I 4) Public, Official Use Only, and Confidential (Note that the word "public" used in data classification is defined differently than the word "public" used in the phrase "public records request" as used in state open records acts.)	University of Washington Privacy Brief: Data Classification	Higher Education
		The Ohio State University Data Element Classification Assignments	Higher Education
		Stanford Data Classification, Access, Transmittal, and Storage Guidelines and Chart	Higher Education
		University of Texas at Austin Data Classification Policy	Higher Education
3.2 Check for state statutes that may already define some or all levels for you, and what words to use to describe the levels. State guidelines will most likely apply to state schools.		Minnesota Government Data Practices Act	Government
3.3 Check for recognized standards that may already define some or all levels for you, and what words to use to describe the levels.		FIPS 199: Standards for Security Categorization of Federal Information and Information Systems	Government
3.4 Consider using Confidentiality, Integrity, and Availability (CIA) as criteria to classify data.		Presentation: Data Classification and Privacy: A Foundation for Compliance	Higher Education

Top of page

Step 4: Determine the methodology and procedures for classifying data

Sub-Step	Tips	Resource	Resource Type

4.1 How will you get started on your classification activities?	This appears to be a huge project, if you consider all the data elements collected and used by a higher education institution. Institutions have used a number of techniques to make the task more manageable.	Speaking the Same Language: Building a Data Governance Program for Institutional Impact (University of Notre Dame)	Higher Education
a. Establish a project team and start with a select few data areas.	Many institutions start with centrally-held administrative data. Those who are responsible for establishing and maintaining appropriate data classification levels for centrally-held data are trained in how to do so; experience is gained with that project; and then this team provides training and guidance subsequently to each data owner to use for all other types of data elements, both centrally managed and managed within decentralized units. *Note: Special projects can be planned to address research data, institutional data NOT in central control, and personal data held by staff such as contact data.*	Risk Management Framework, Phase 0: Strategic Risk Assessment Planning	EDUCAUSE
b. Assign a default classification	Data Classification Policy could state that all data is classified at a particular level as the default. Then, only data that falls outside of this default level needs to be formally classified.	Duke University: Data Classification at Duke	Higher Education
c. Provide basic policy and procedure documents and tools, and ask each data owner to work independently.		Carnegie Mellon University: Guidelines for Data Classification	Higher Education

Top of page

Step 5: Determine and review other information security processes impacted by data classification

Once you are done creating your classification scheme, what comes next? Data classification is usually one of the first steps in a long progression of activities to safeguard your data. The list below provides a checklist of other information security processes that may be impacted by data classification activities, but it does not attempt to provide full information on each of these other processes.

Sub-Step	Tips	Resource	Resource Type
5.1 Access Management	Determining who can access the data, and what they can do with it.	Identity and Access Management	EDUCAUSE
5.2 Physical Security		Physical and Environmental Controls	EDUCAUSE
5.3 Risk Assessment		Risk Management	EDUCAUSE
5.4 Change Management Requirements		University of Texas at Austin Change Management Guidelines	Higher Education
5.5 Training		Awareness and Training	EDUCAUSE
		Kansas State University: Data Classification and Security Policy – Data Security Standards – Training	Higher Education
5.6 Need for Policy and Procedures		Security Policies	EDUCAUSE
5.7 Need for Encryption	Determining how data is appropriately secured both while at rest (in storage) and in transmission.	Encryption 101	EDUCAUSE
5.8 Records Retention	Determining how long each type of data should be stored.	Boston University: Data Protection Requirements	Higher Education
5.9 Data Incident Handling and Response	Determining what happens if/when data is lost, stolen, or compromised.	Confidential Data Handling Blueprint Data Incident Notification Toolkit Incident Management and Response	EDUCAUSE

Top of page

Questions or comments? Contact us.