Chapter 5
Protecting Security of Assets

THE CISSP TOPICS COVERED IN THIS CHAPTER INCLUDE:

✓ **Domain 2.0: Asset Security**

- 2.1 Identify and classify information and assets
  - 2.1.1 Data classification
  - 2.1.2 Asset classification
- 2.2 Establish information and asset handling requirements
- 2.4 Manage data lifecycle
  - 2.4.1 Data roles (i.e., owners, controllers, custodians, processors, users/subjects)
  - 2.4.2 Data collection
  - 2.4.3 Data location
  - 2.4.4 Data maintenance
  - 2.4.5 Data retention
  - 2.4.6 Data remanence
  - 2.4.7 Data destruction
- 2.5 Ensure appropriate asset retention (e.g., end of life (EOL), end of support)
- 2.6 Determine data security controls and compliance requirements
  - 2.6.1 Data states (e.g., in use, in transit, at rest)
  - 2.6.2 Scoping and tailoring
  - 2.6.3 Standards selection
  - 2.6.4 Data protection methods (e.g., Digital Rights Management (DRM), data loss prevention (DLP), cloud access security broker (CASB))

The Asset Security domain focuses on collecting, handling, and protecting information throughout its life cycle. A primary step in this domain is classifying information based on its value to the organization. All follow-on actions vary depending on the classification. For example, highly classified data requires stringent security controls. In contrast, unclassified data uses fewer security controls.

Identifying and Classifying Information and Assets

Managing the data life cycle refers to protecting it from the cradle to the grave. Steps need to be taken to protect the data when it is first created until it is destroyed.

One of the first steps in the life cycle is identifying and classifying information and assets. Organizations often include classification definitions within a security policy. Personnel then label assets appropriately based on the security policy requirements. In this context, assets include sensitive data, the hardware used to process it, and the media used to hold it.

Defining Sensitive Data

Sensitive data is any information that isn't public or unclassified. It can include confidential, proprietary, protected, or any other type of data that an organization needs to protect due to its value to the organization, or to comply with existing laws and regulations.

Personally Identifiable Information

Personally identifiable information (PII) is any information that can identify an individual. National Institute of Standards and Technology (NIST) Special Publication (SP) 800-122 provides a more formal definition:

Any information about an individual maintained by an agency, including

1. any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and
2. any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.

The key is that organizations have a responsibility to protect PII. This includes PII related to employees and customers. Many laws require organizations to notify individuals if a data breach results in a compromise of PII.

Protection for personally identifiable information (PII) drives privacy and confidentiality requirements for rules, regulations, and legislation worldwide (especially in North America and the European Union). NIST SP 800-122—Guide to Protecting the Confidentiality of Personally Identifiable Information (PII), provides more information on how to protect PII. It is available from the NIST Special Publications (800 Series) download page: http://csrc.nist.gov/publications/sp800.

Protected Health Information

In the United States, the Health Insurance Portability and Accountability Act (HIPAA) mandates the protection of some health records. To fully understand what information is covered by HIPAA, we need to look at a few definitions. First, the general definition of health information is:

> Health information means any information, whether oral or recorded in any form or medium, that—
>
> (A) is created or received by a health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse; and
>
> (B) relates to the past, present, or future physical or mental health or condition of any individual, the provision of health care to an individual, or the past, present, or future payment for the provision of health care to an individual.

Protected health information (PHI) is any health information that is transmitted in electronic form, maintained in electronic media, or transmitted or maintained in any other form or media. Education records, employment records of a covered entity, and records relating to individuals who have been deceased more than 50 years are excluded from the definition of PHI.

Some people think that only medical care providers, such as doctors and hospitals, need to protect PHI. However, HIPAA defines PHI much more broadly. The law applies to healthcare providers, health insurers, and health information clearinghouses, as well as business associates of those organizations that handle PHI. Employers that provide health insurance may handle PHI, so HIPAA applies to a large percentage of organizations in the United States.

Proprietary Data

Proprietary data refers to any data that helps an organization maintain a competitive edge. It could be software code it developed, technical plans for products, internal processes, intellectual property, or trade secrets. If competitors gain access to the proprietary data, it can seriously affect the primary mission of an organization.

Although copyright, patent, and trade secret laws provide a level of protection for proprietary data, this isn't always enough. Many criminals ignore copyrights, patents, and laws. Similarly, foreign entities have stolen a significant amount of proprietary data.

Defining Data Classifications

Organizations typically include data classifications in their security policy or a data policy. A *data classification* identifies the value of the data to the organization and is critical to protect data confidentiality and integrity. The policy identifies classification labels used within the organization. It also identifies how data owners can determine the proper classification and how personnel should protect data based on its classification.

As an example, government data classifications include top secret, secret, confidential, and unclassified. Anything above unclassified is sensitive data, but clearly, these have different values. The U.S. government provides clear definitions for these classifications. As you read them, note that the wording of each definition is close except for a few key words. *Top secret* uses the phrase "exceptionally grave damage," *secret* uses the phrase "serious damage," and *confidential* uses "damage":

Top Secret The top secret label is "applied to information, the unauthorized disclosure of which reasonably could be expected to cause exceptionally grave damage to the national security that the original classification authority is able to identify or describe."

Secret The secret label is "applied to information, the unauthorized disclosure of which reasonably could be expected to cause serious damage to the national security that the original classification authority is able to identify or describe."

Confidential The confidential label is "applied to information, the unauthorized disclosure of which reasonably could be expected to cause damage to the national security that the original classification authority is able to identify or describe."

Unclassified Unclassified refers to any data that doesn't meet one of the descriptions for top secret, secret, or confidential data. Within the United States, unclassified data is available to anyone, though it often requires individuals to

request the information using procedures identified in the Freedom of Information Act (FOIA).

There are additional subclassifications of unclassified, such as for official use only (FOUO), sensitive but unclassified (SBU), and controlled unclassified information (CUI). Documents with these designations have strict controls limiting their distribution. As an example, the U.S. Internal Revenue Service (IRS) uses SBU for individual tax records, restricting access to these records.

A classification authority is the entity that applies the original classification to the sensitive data, and there are strict rules that identify who can do so. For example, the U.S. president, vice president, and agency heads can classify data in the United States. Additionally, individuals in any of these positions can delegate permission for others to classify data.

Although the focus of classifications is often on data, these classifications also apply to hardware assets. This includes any computing system or media that processes or holds this data.

Nongovernmental organizations rarely need to classify their data based on potential damage to national security. However, management is concerned about potential damage to the organization. For example, if attackers accessed the organization's data, what is the potential adverse impact? In other words, an organization doesn't just consider the sensitivity of the data but also the criticality of the data. They could use the same phrases of “exceptionally grave damage,” “serious damage,” and “damage” that the U.S. government uses when describing top secret, secret, and confidential data.

Some nongovernmental organizations use labels such as Class 3, Class 2, Class 1, and Class 0. Other organizations use more meaningful labels such as confidential (or proprietary), private,

sensitive, and public. Figure 5.1 shows the relationship between these different classifications, with the government classifications on the left and the nongovernment (or civilian) classifications on the right. Just as the government can define the data based on the potential adverse impact from a data breach, organizations can use similar descriptions.

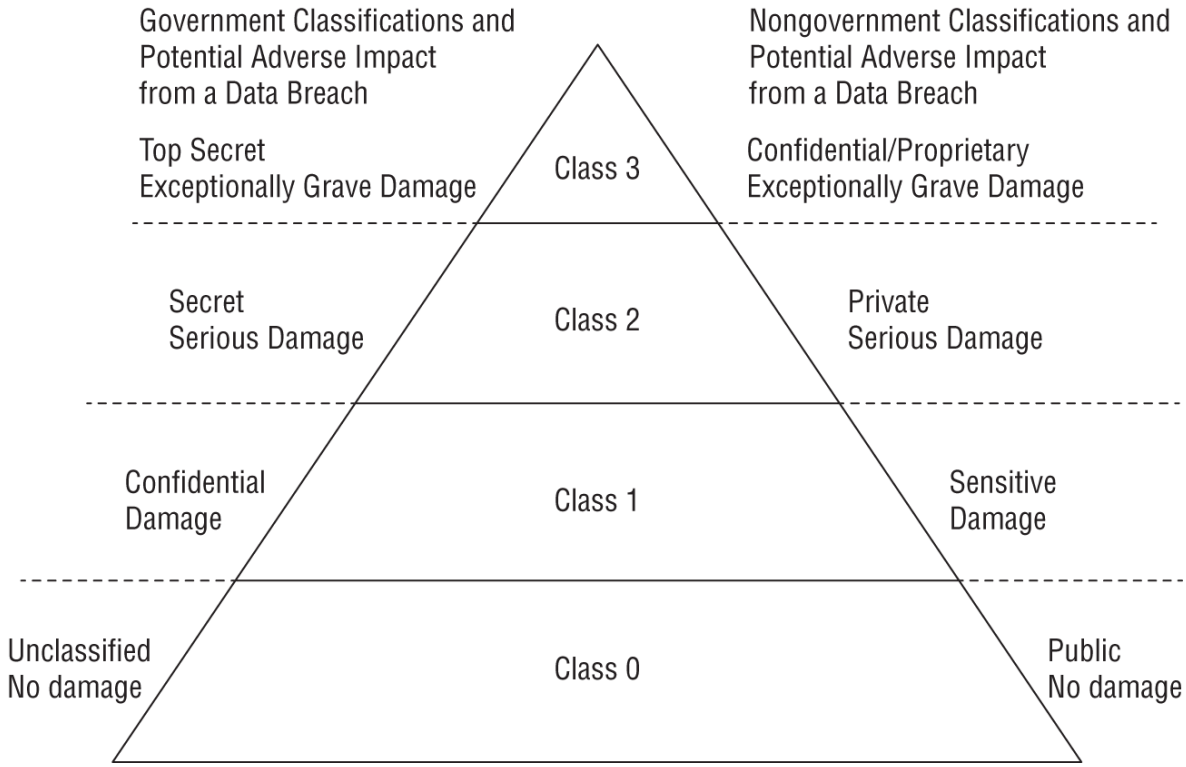

FIGURE 5.1 Data classifications

Both government and civilian classifications identify the relative value of the data to the organization, with top secret representing the highest classification for governments and confidential representing the highest classification for organizations in Figure 5.1. However, it's important to remember that organizations can use any labels they desire. The following sections identify the meaning of some common nongovernment classifications. Remember, even though these are commonly used, there is no standard that all private organizations must use.

Confidential or Proprietary The confidential or *proprietary* label typically refers to the highest level of classified data. In this context, a data breach would cause exceptionally grave damage to the mission of the organization. As an example, attackers have

repeatedly attacked Sony, stealing more than 100 terabytes of data, including full-length versions of unreleased movies. These quickly showed up on file-sharing sites, and security experts estimate that people downloaded these movies up to a million times. With pirated versions of the movies available, many people skipped seeing them when Sony ultimately released them. This directly affected Sony's bottom line. The movies were proprietary, and the organization might have considered it exceptionally grave damage. In retrospect, they may choose to label movies as confidential or proprietary and use the strongest access controls to protect them.

Private The *private* label refers to data that should stay private within the organization but that doesn't meet the definition of confidential or proprietary data. In this context, a data breach would cause serious damage to the mission of the organization. Many organizations label PII and PHI data as private. It's also common to label internal employee data and some financial data as private. As an example, the payroll department of a company would have access to payroll data, but this data is not available to regular employees.

Sensitive *Sensitive data* is similar to confidential data. In this context, a data breach would cause damage to the mission of the organization. As an example, IT personnel within an organization might have extensive data about the internal network, including the layout, devices, operating systems, software, Internet Protocol (IP) addresses, and more. If attackers have easy access to this data, it makes it much easier for them to launch attacks. Management may decide they don't want this information available to the public, so they might label it as sensitive.

Public *Public data* is similar to unclassified data. It includes information posted on websites, in brochures, or any other public source. Although an organization doesn't protect the confidentiality of public data, it does take steps to protect its integrity. For example, anyone can view public data posted on a website. However, an organization doesn't want attackers to modify this data, so it takes steps to protect it.

Civilian organizations aren't required to use any specific classification labels. However, it is important to classify data in some manner and ensure personnel understand the classifications. No matter what labels an organization uses, it still has an obligation to protect sensitive information.

After classifying the data, an organization takes additional steps to manage it based on its classification. Unauthorized access to sensitive information can result in significant losses to an organization. However, basic security practices, such as properly marking, handling, storing, and destroying data and hardware assets based on classifications, helps prevent losses.

Defining Asset Classifications

Asset classifications should match the data classifications. In other words, if a computer is processing top secret data, the computer should also be classified as a top secret asset. Similarly, if media such as internal or external drives hold top secret data, the media should also be classified as top secret.

It is common to use clear marking on the hardware assets so that personnel are reminded of data that can be processed or stored on the asset. For example, if a computer is used to process top secret data, the computer and the monitor will have clear and prominent labels reminding users of the classification of data that can be processed on the computer.

Understanding Data States

It's important to protect data in all *data states*, including while it is at rest, in transit, and in use.

Data at Rest Data at rest (sometimes called data on storage) is any data stored on media such as system hard drives, solid-state drives (SSDs), external USB drives, storage area networks (SANs), and backup tapes. Strong symmetric encryption protects data at rest.

Data in Transit Data in transit (sometimes called data in motion or being communicated) is any data being transmitted

over a network. This includes data being transmitted over an internal network using wired or wireless methods and data being transmitted over public networks such as the Internet. A combination of symmetric and asymmetric encryption protects data in transit.

Data in Use Data in use (also known as data being processed) refers to data in memory or temporary storage buffers while an application is using it. Applications often decrypt encrypted data before placing it in memory. This allows the application to work on it, but it's important to flush these buffers when the data is no longer needed. In some cases, it's possible for an application to work on encrypted data using homomorphic encryption. This limits the risk because memory doesn't hold unencrypted data.

The best way to protect the confidentiality of data is to use strong encryption protocols, discussed extensively in Chapter 6, "Cryptography and Symmetric Key Algorithms." Additionally, strong authentication and authorization controls help prevent unauthorized access.

As an example, consider a web application that retrieves credit card data for quick access and reuse with the user's permission for an e-commerce transaction. The credit card data is stored in a database server and protected while at rest, while in transit, and while in use.

Database administrators take steps to encrypt sensitive data stored in the database server (data at rest). They would typically encrypt columns holding sensitive data such as credit card data. Additionally, they would implement strong authentication and authorization controls to prevent unauthorized entities from accessing the database.

When the web application sends a request for data from the web server, the database server verifies that the web application is authorized to retrieve the data and, if so, the database server sends it. However, this entails several steps. For example, the database management system first retrieves and decrypts the data and formats it in a way that the web application can read it. The database server then uses a transport encryption algorithm

to encrypt the data before transmitting it. This ensures that the data in transit is secure.

The web application server receives the data in an encrypted format. It decrypts the data and sends it to the web application. The web application stores the data in temporary memory buffers while it uses it to authorize the transaction. When the web application no longer needs the data, it takes steps to purge memory buffers, ensuring the complete removal of all residual sensitive data.

The Identity Theft Resource Center (ITRC) routinely tracks data breaches. They post reports through their website (http://idtheftcenter.org) that are free to anyone. In 2023, they tracked 3,205 data breaches, exposing the information of more than 353 million people.

Determining Compliance Requirements

Every organization has a responsibility to learn what legal requirements apply to them and ensure they meet all the compliance requirements. This is especially important if an organization handles PII in different countries. Chapter 4, "Laws, Regulations, and Compliance," covers a wide assortment of laws and regulations that apply to organizations around the world. For any organization involved in e-commerce, this can get complex very quickly. An important point to remember is that an organization needs to determine what laws apply to it.

Imagine a group of college students work together and create an app that solves a problem for them. On a whim, they start selling the app from the Apple App Store and it goes viral. People around the world are buying the app, bringing cash windfalls to these students. It also brings major headaches. Suddenly these college students need to be knowledgeable about laws around the world that apply to them.

Some organizations have created a formal position called a compliance officer. The person filling this role ensures that the organization is conducting all business activities by following the laws and regulations that apply to the organization. Of course, this starts by first determining everywhere the organization operates and what compliance requirements apply.

Determining Data Security Controls

After defining data and asset classifications, you must define the security requirements and identify security controls to implement those requirements. Imagine that your organization has decided to use the data labels Confidential/Proprietary, Private, Sensitive, and Public, as described earlier. Management then decides on a data security policy dictating the use of specific security controls to protect data in these categories. The policy will likely address data stored in files, in databases, on servers such as email servers, on user systems, sent via email, and stored in the cloud.

For this example, we're limiting the type of data to email only. Your organization has defined how it wants to protect email in each of the data categories. They've decided that any email in the Public category doesn't need to be encrypted. However, email in all other categories (Confidential/Proprietary, Private, and Sensitive) must be encrypted when being sent (data in transit) and while stored on an email server (data at rest).

Encryption converts cleartext data into scrambled ciphertext and makes it more difficult to read. Using strong encryption methods such as the Advanced Encryption Standard with 256-bit keys (AES 256) makes it almost impossible for unauthorized personnel to read the text.

Table 5.1 shows other security requirements for email that management has defined in their data security policy. Notice that data in the highest level of classification category (Confidential/Proprietary in this example) has the most security requirements defined in the security policy.

TABLE 5.1 Securing email data

Classification	Security requirements for email
Confidential/Proprietary (highest level of protection for any data)	Email and attachments must be encrypted with AES 256. Email and attachments remain encrypted except when viewed. Email can be sent only to recipients within the organization. Email can be opened and viewed only by recipients (forwarded emails cannot be opened). Attachments can be opened and viewed, but not saved. Email content cannot be copied and pasted into other documents. Email cannot be printed.
Private (examples include PII and PHI)	Email and attachments must be encrypted with AES 256. Email and attachments remain encrypted except when viewed. Email can be sent only to recipients within the organization.
Sensitive (lowest level of protection for classified data)	Email and attachments must be encrypted with AES 256.
Public	Email and attachments can be sent in cleartext.

The requirements listed in Table 5.1 are provided as an example only. Any organization could use these requirements or define other requirements that work for them.

Security administrators use the requirements defined in the security policy to identify security controls. For Table 5.1, the primary security control is strong encryption using AES 256. Administrators should identify methodologies, making it easy for employees to meet the requirements.

Although it's possible to meet all the requirements for securing email shown in Table 5.1, doing so might require implementing other solutions. For example, several software companies sell a range of products that organizations can use to automate these tasks. Users apply relevant labels (such as confidential, private, sensitive, and public) to emails before sending them. These emails pass through a data loss prevention (DLP) server that detects the labels and applies the required protection. The settings for these DLP solutions can be configured for an organization's specific needs.

Table 5.1 shows possible requirements that your organization might want to apply to email. However, you shouldn't stop there. Any type of data that your organization wants to protect needs similar security definitions. For example, you should define requirements for data stored on assets such as servers, data backups stored on-site and off-site, and proprietary data.

Additionally, identity and access management security controls help ensure that only authorized personnel can access resources. Chapter 13, “Managing Identity and Authentication,” and Chapter 14, “Controlling and Monitoring Access,” cover identity and access management security controls in more depth.

Establishing Information and Asset Handling Requirements

A key goal of managing sensitive data is to prevent data breaches. A data breach is an event in which an unauthorized entity can view or access sensitive data. If you pay attention to the news, you probably hear about data breaches quite often. Large data breaches such as the Marriott data breach of 2020 hit the mainstream news. Marriott reported that attackers stole personal data, including names, addresses, email addresses, employer information, and phone numbers, of approximately 5.2 million guests.

The following sections identify basic steps people within an organization should follow to limit the possibility of data breaches.

Data Maintenance

Data maintenance refers to ongoing efforts to organize and care for data throughout its lifetime. In general, if an organization stores all sensitive data on one server, it is relatively easy to apply all the appropriate controls to this one server. In contrast, if sensitive data is stored throughout an organization on multiple servers and end-user computers and mixed with nonsensitive data, it becomes much harder to protect it.

One option would be for one network to process only unclassified data while another network processes classified data. Techniques such as air gaps ensure the two networks never physically touch each other. An *air gap* is a physical security control and means that systems and cables from the classified network never physically touch systems and cables from the unclassified network. Additionally, the classified network can't access the Internet, and Internet attackers can't access it.

Still, there are times when personnel need to add data to the classified network, such as when devices, systems, and applications need updates. One way is manual; personnel copy the data from the unclassified network to a USB device and carry

it to the classified network. Another method is to use a unidirectional network bridge; this connects the two networks but allows the data to travel in only one direction, from the unclassified network to the classified network. A third method is to use a technical guard solution, which is a combination of hardware and software placed between the two networks. A guard solution allows properly marked data to travel between the two networks.

Additionally, an organization should routinely review data policies to ensure that they are kept up-to-date and that personnel are following the policies. It's often a good practice to review the causes of recent data breaches and ensure that similar mistakes are not causing needless vulnerabilities.

Data Loss Prevention

Data loss prevention (DLP) solutions attempt to detect and block data exfiltration attempts. These solutions have the capability of scanning unencrypted data looking for keywords and data patterns. For example, imagine that your organization uses data classifications of Confidential, Proprietary, Private, and Sensitive. A DLP system can scan files for these words and detect them.

Pattern-matching DLP systems look for specific patterns. For example, U.S. Social Security numbers have a pattern of *nnn-nn-nnnn* (three numbers, a dash, two numbers, a dash, and four numbers). The DLP can look for this pattern and detect it. Administrators can set up a DLP system to look for any patterns based on their needs. Cloud DLP solutions can look for the same keywords or patterns.

There are three types of DLP solutions:

Network DLP A network DLP scans all outgoing data in a traditional network looking for specific data. Administrators place it on the edge of the network to scan all data leaving the organization. If a user sends out a file containing restricted data, the DLP system will detect it and prevent it from leaving the organization. The DLP system will send an alert, such as an email to an administrator.

Endpoint DLP An endpoint DLP can scan files stored on a system as well as files sent to external devices, such as printers. For example, an organization's endpoint DLP can prevent users from copying sensitive data to USB flash drives or sending sensitive data to a printer. Administrators configure the DLP to scan the files with the appropriate keywords, and if it detects files with these keywords, it will block the copy or print job. It's also possible to configure an endpoint DLP solution to regularly scan files (such as on a file server) for files containing specific keywords or patterns, or even for unauthorized file types, such as MP3 files.

Cloud DLP Cloud DLP is a subset of network DLP designed and tailored for cloud-native environments.

DLP solutions typically can perform deep-level examinations. For example, if users embed the files in compressed zip files, a DLP solution can still detect the keywords and patterns. However, a DLP solution can't decrypt data or examine encrypted data.

Most DLP solutions also include discovery capabilities. The goal is to discover the location of valuable data within an internal network. When security administrators know where the data is, they can take additional steps to protect it. As an example, a database server may include unencrypted credit card numbers. When the DLP discovers and reports this, database administrators can ensure the numbers are encrypted. As another example, company policy may dictate that employee laptops do not contain any PII data. A DLP content discovery system can search these and discover any unauthorized data. Additionally, many content discovery systems can search cloud resources used by an organization.

Labeling Sensitive Data and Assets

Labeling (often called security labeling) sensitive information ensures that users can easily identify the classification level of any data. The most important information that a tag or a label provides is the classification of the data. For example, a label of top secret makes it clear to anyone who sees the label that the

information or asset is classified top secret. When users know the value of the data or asset, they are more likely to take appropriate steps to control and protect it based on the classification. Security labeling includes both physical and electronic tags and labels.

Physical labels indicate the security classification for the data stored on assets such as media or processed on a system. For example, if a backup tape includes secret data, a physical label attached to the tape makes it clear to users that it holds secret data.

Similarly, if a computer processes sensitive information, the computer would have a label indicating the highest classification of information that it processes. A computer used to process confidential, secret, and top secret data should be marked with a label indicating that it processes top secret data. Physical labels remain on the system or media throughout its lifetime.

Security labeling also includes using digital tags or labels. A simple method is to include the classification as a header or footer in a document or embed it as a watermark. A benefit of these methods is that they also appear on printouts. Even when users include headers and footers on printouts, most organizations require users to place printed sensitive documents within a folder that includes a label or cover page clearly indicating the classification. Headers aren't limited to files. Backup tapes often include header information, and the classification can be included in this header.

Another benefit of headers, footers, and watermarks is that DLP systems can identify documents that include sensitive information and apply the appropriate security controls. Some DLP systems will also add metadata tags to the document when they detect that the document is classified. These tags provide insight into the document's contents and help the DLP system handle it appropriately.

Similarly, some organizations mandate specific desktop backgrounds on their computers. For example, a system used to process proprietary data might have a black desktop background with the word *Proprietary* in white and a wide orange border. The

background could also include statements such as “This computer processes proprietary data” and statements reminding users of their responsibilities to protect the data.

In many secure environments, personnel also use labels for unclassified media and equipment. This prevents an error of omission where sensitive information isn't marked. For example, if a backup tape holding sensitive data isn't marked, a user might assume it only holds unclassified data. However, if the organization marks unclassified data, too, unlabeled media would be easily noticeable, and the user would view an unmarked tape with suspicion.

Organizations often identify procedures to downgrade media. For example, if a backup tape includes confidential information, an administrator might want to downgrade the tape to unclassified. The organization would identify trusted procedures that will purge the tape of all usable data. After administrators purge the tape, they can then downgrade it and replace the labels.

However, many organizations prohibit downgrading media at all. For example, a data policy might prohibit downgrading a backup tape that contains top secret data. Instead, the policy might mandate destroying this tape when it reaches the end of its life cycle. Similarly, it is rare to downgrade a system. In other words, if a system has been processing top secret data, it would be rare to downgrade it and relabel it as an unclassified system. In any event, approved procedures would need to be created to inform personnel what can be downgraded and what should be destroyed.

If media or a computing system needs to be downgraded to a less sensitive classification, it must be sanitized using appropriate procedures, as described in the section “Data Destruction,” later in this chapter. However, it's often safer and easier just to purchase new media or equipment rather than follow through with the sanitization steps for reuse.

Handling Sensitive Information and Assets

Handling refers to the secure transportation of media through its lifetime. Personnel handle data differently based on its value and classification, and as you'd expect, highly classified information needs much greater protection. Even though this is common sense, people still make mistakes. Many times, people get accustomed to handling sensitive information and become lackadaisical about protecting it.

A common occurrence is the loss of control of backup tapes. Backup tapes should be protected with the same level of protection as the data that they contain. In other words, if confidential information is on a backup tape, the backup tape should be protected as a confidential asset.

Similarly, data stored in the cloud needs to be protected with the same level of protection with which it is protected on-site. Amazon Web Services (AWS) Simple Storage Service (S3) is a cloud-based object storage service. Data is stored in S3 *buckets*, which are like folders on Windows systems. Just as you set permissions on any folder, you set permissions on AWS buckets. Unfortunately, this concept eludes many AWS users. As an example, a bucket owned by THSuite, a cannabis retailer, exposed the PII of more than 30,000 individuals in early 2020. Another example from 2020 involved 900,000 before and after cosmetic surgery images and videos stored in an unsecured bucket. Many

of these included clear views of the patients' faces, along with all parts of their bodies.

Policies and procedures need to be in place to ensure that people understand how to handle sensitive data. This starts by ensuring that systems and media are labeled appropriately. Additionally, as President Reagan famously said when discussing relations with the Soviet Union, “Trust, but verify.” Chapter 17, “Preventing and Responding to Incidents,” discusses the importance of logging, monitoring, and auditing. These controls verify that sensitive information is handled appropriately before a significant loss occurs. If a loss does occur, investigators use audit trails to help discover what went wrong. Any incidents that occur because personnel didn't handle data appropriately should be quickly investigated and actions taken to prevent a reoccurrence.

Data Collection Limitation

One of the easiest ways to prevent the loss of data is to simply not collect it. As an example, consider a small e-commerce company that allows customers to make purchases with a credit card. It uses a credit card processor to process credit card payments. If the company just passes the credit card data to the processor for approval and never stores it on a company server, the company cannot lose the credit card data in a later breach.

In contrast, imagine a different e-commerce company sells products online. Every time a customer makes a purchase, the company collects as much information as possible on the customer, such as the name, email address, physical address, phone number, credit card data, and more. It suffers a data breach and all this data is exposed, resulting in significant liabilities for the company.

The guideline is clear. If the data doesn't have a clear purpose for use, don't collect it and store it. This is also why many privacy regulations mention limiting data collection.

Data Location

Data location refers to the location of data backups or data copies. Imagine a small organization's primary business location is in Norfolk, Virginia. The organization stores all the data on-site. However, they regularly perform backups of the data.

A best practice is to keep a backup copy on-site and another backup copy off-site. If a disaster, such as a fire, destroys the primary business location, the organization would still have a backup copy stored off-site.

The decision of how far off-site to store the backup needs to be considered. If it's stored in a business located in the same building, it could be destroyed in the same fire. Even if the backup was stored 5 miles away, it is possible a hurricane or flood could destroy both locations.

Some organizations maintain data in large data centers. It's common to replicate this data to one or more other data centers to maintain the availability of the critical data. These data centers are typically located in separate geographical locations. When using cloud storage for backups, some organizations may need to verify the location of the cloud storage to ensure it is in a separate geographical location.

Storing Sensitive Data

Sensitive data should be stored in such a way that it is protected against any type of loss. Encryption methods prevent unauthorized entities from accessing the data even if they obtain databases or hardware assets.

If sensitive data is stored on physical media such as portable disk drives or backup tapes, personnel should follow basic physical security practices to prevent losses due to theft. This includes storing these devices in locked safes or vaults, or within a secure room that includes several additional physical security controls. For example, a server room includes physical security measures to prevent unauthorized access, so storing portable media within

a locked cabinet in a server room would provide strong protection.

Additionally, environmental controls protect the media. This includes temperature and humidity controls such as heating, ventilation, and air conditioning (HVAC) systems.

Here's a point that end users often forget: the value of any sensitive data is much greater than the value of the media holding the sensitive data. In other words, it's cost-effective to purchase high-quality media, especially if the data will be stored for a long time, such as on backup tapes. Similarly, the purchase of high-quality USB flash drives with built-in encryption is worth the cost. Some of these USB flash drives include biometric authentication mechanisms using fingerprints, which provide added protection.

Encryption of sensitive data provides an additional layer of protection and should be considered for any data at rest. If data is encrypted, it becomes much more difficult for an attacker to access it, even if it is stolen.

Data Destruction

When an organization no longer needs sensitive data, personnel should destroy it. Proper destruction ensures that it cannot fall into the wrong hands and result in unauthorized disclosure. Highly classified data requires different steps to destroy it than data classified at a lower level. An organization's security policy or data policy should define the acceptable methods of destroying data based on the data's classification. For example, an organization may require the complete destruction of media holding highly classified data, but allow personnel to use software tools to overwrite data files classified at a lower level.

NIST SP 800-88, Rev. 1—Guides for Media Sanitization provides comprehensive details on different sanitization methods. Sanitization methods (such as clearing, purging, and destroying)

help ensure that data cannot be recovered. Proper sanitization steps remove all sensitive data before disposing of a computer. This includes removing or destroying data on nonvolatile memory, internal hard drives, and solid-state drives (SSDs). It also includes removing all CDs/DVDs and Universal Serial Bus (USB) drives.

Sanitization can refer to the destruction of media or using a trusted method to purge classified data from the media without destroying it.

Eliminating Data Remanence

Data remanence is the data that remains on media after the data was supposedly erased. It typically refers to data on a hard drive as residual magnetic flux or slack space. If media includes any type of private and sensitive data, it is important to eliminate data remanence.

Slack space is the unused space within a disk cluster. Operating systems store files on hard disk drives in clusters, which are groups of sectors (the smallest storage unit on a hard disk drive). Sector and cluster sizes vary, but for this example, imagine a cluster size of 4,096 bytes and a file size of 1,024 bytes. After storing the file, the cluster would have 3,072 bytes of unused space or slack space.

Some operating systems fill this slack space with data from memory. If a user was working on a top secret file a moment ago and then creates a small unclassified file, the small file might contain top secret data pulled from memory. This is one of the reasons why personnel should never process classified data on unclassified systems.

Using system tools to delete data generally leaves much of the data remaining on the media, and widely available tools can easily undelete it. Even when you use sophisticated tools to overwrite the media, traces of the original data may remain as less perceptible magnetic fields. This is like a ghost image that can remain on some older TV and computer monitors if the same data is displayed for long periods of time. Forensics experts and

attackers have tools they can use to retrieve this data even after it has been supposedly overwritten.

One way to remove data remanence is with a degausser. A degausser generates a heavy magnetic field, which realigns the magnetic fields in magnetic media such as traditional hard drives, magnetic tape, and floppy disk drives. Degaussers using power will reliably rewrite these magnetic fields and remove data remanence. However, they are only effective on magnetic media.

In contrast, SSDs use integrated circuitry instead of magnetic flux on spinning platters. Because of this, degaussing SSDs won't remove data. However, even when using other methods to remove data from SSDs, data remnants often remain.

Some SSDs include built-in erase commands to sanitize the entire disk, but unfortunately, these weren't effective on some SSDs from different manufacturers. Due to these risks, the best method of sanitizing SSDs is destruction. The U.S. National Security Agency (NSA) requires the destruction of SSDs using an approved disintegrator. Approved disintegrators shred the SSDs to a size of 2 millimeters (mm) or smaller. Many organizations sell multiple information destruction and sanitization solutions used by government agencies and organizations in the private sector that the NSA has approved.

Another method of protecting SSDs is to ensure that all stored data is encrypted. If a sanitization method fails to remove all the data remnants, the remaining data would be unreadable.

Be careful when performing any type of clearing, purging, or sanitization process. The human operator or the tool involved in the activity may not properly perform the task of completely removing data from the media. Software can be flawed, magnets can be faulty, and either can be used improperly. Always verify that the desired result is achieved after performing any sanitization process.

Common Data Destruction Methods

The following list includes some common terms associated with destroying data:

Erasing *Erasing* media is simply performing a delete operation against a file, a selection of files, or the entire media. In most cases, the deletion or removal process removes only the directory or catalog link to the data. The actual data remains on the drive. As new files are written to the media, the system eventually overwrites the erased data, but depending on the size of the drive, how much free space it has, and several other factors, the data may not be overwritten for months. Anyone can typically retrieve the data using widely available undelete tools.

Clearing *Clearing*, or *overwriting*, is a process of preparing media for reuse and ensuring that the cleared data cannot be recovered using traditional recovery tools. When media is cleared, unclassified data is written over all addressable locations on the media. One method writes a single character, or a specific bit pattern, over the entire media. A more thorough method writes a single character over the entire media, writes the character's complement over the entire media, and finishes by writing random bits over the entire media. It repeats this in three separate passes, as shown in Figure 5.2. Although this sounds like the original data is lost forever, it may be possible to retrieve some of the original data using sophisticated laboratory or forensics techniques. Additionally, not all types of data storage respond well to clearing techniques. For example, spare sectors on hard drives, sectors labeled as “bad,” and areas on many modern SSDs are not necessarily cleared and may still retain data.

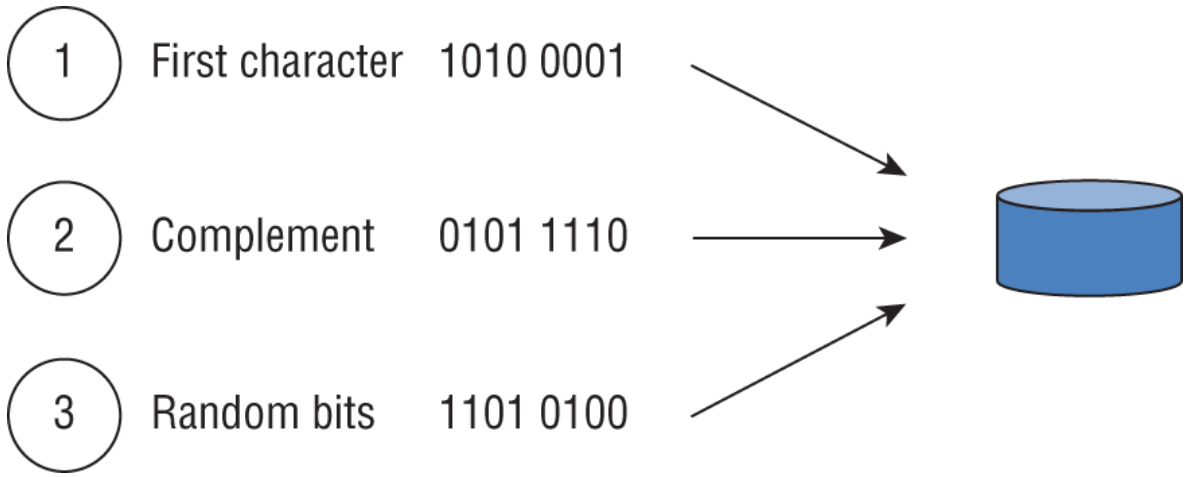

FIGURE 5.2 Clearing a hard drive

Purging *Purging* is a more intense form of clearing that prepares media for reuse in less secure environments. It provides a level of assurance that the original data is not recoverable using any known methods. A purging process will repeat the clearing process multiple times in order to completely remove the data. Even though purging is intended to remove all data remnants, it isn't always trusted. For example, the U.S. government doesn't consider any purging method acceptable to purge top secret data. Media labeled top secret will always remain top secret until it is destroyed.

Degaussing A degausser creates a strong magnetic field that erases data on some types of media in a process called *degaussing*. Technicians commonly use degaussing methods to remove data from magnetic tapes and magnetic hard disk drives (HDDs) with the goal of removing data from that media. Degaussing may render a hard drive unusable so it is not a good option when you intend to reuse the media.

Degaussing does not affect optical discs (CDs, DVDs, Blu-rays) or flash storage media (SD cards, USB flash drives, SSDs).

Destruction Destruction is the final stage in the life cycle of media and is the most secure method of sanitizing media. When destroying media, ensure that the media cannot be reused or repaired and that data cannot be extracted from the destroyed media. Methods of destruction include incineration, shredding, disintegration, pulverizing, and melting. Some organizations

remove the platters in highly classified disk drives and destroy them separately.

When organizations donate or sell used computer equipment, they often remove and destroy storage devices that hold sensitive data rather than attempting to purge them. This eliminates the risk that the purging process wasn't complete, which would have resulted in a loss of confidentiality.

Declassification involves any process that purges media or a system in preparation for reuse in an unclassified environment. Sanitization methods can be used to prepare media for declassification, but often the efforts required to securely declassify media are significantly greater than the cost of new media for a less secure environment. Additionally, even though purged data is not recoverable using any known methods, there is a remote possibility that an unknown method is or becomes available. Instead of taking the risk, many organizations choose not to declassify any media and instead destroy it when it is no longer needed.

Cryptographic Erasure

If data is encrypted on a device, it's possible to use cryptographic erasure or cryptoshredding to destroy the data. However, these terms are misleading. They don't erase or shred the data. Instead, they destroy the associated keys. With the cryptographic keys erased, data remains encrypted and can't be accessed.

When using this method, you should use another method to overwrite the data. If the original encryption isn't strong, someone may be able to decrypt it without the key. Additionally, there are often backups of cryptographic keys, and if someone discovers a backup key, they can still access the data.

When using cloud storage, destroying the cryptographic keys may be the only form of secure deletion available to an organization.

Ensuring Appropriate Data and Asset Retention

Retention requirements apply to data or records, media holding sensitive data, systems that process sensitive data, and personnel who have access to sensitive data. Record retention and media retention are the most important elements of asset retention. Chapter 3, "Business Continuity Planning," covers a vital records program, which can be referenced to identify records to retain.

Record retention involves retaining and maintaining important information as long as it is needed and destroying it when it is no longer needed. An organization's security policy or data policy typically identifies retention time frames. Some laws and regulations dictate the length of time that an organization should retain data, such as three years, seven years, or even indefinitely. Organizations have the responsibility of identifying laws and regulations that apply and complying with them. However, even in the absence of external requirements, an organization should still identify how long to retain data.

As an example, many organizations require the retention of all audit logs for a specific amount of time. The period can be dictated by laws, regulations, requirements related to partnerships with other organizations, or internal management decisions. These audit logs allow the organization to reconstruct the details of past security incidents. When an organization doesn't have a retention policy, administrators may delete valuable data earlier than management expects them to or attempt to keep data indefinitely. The longer an organization retains data, the more it costs in terms of media, locations to store it, and personnel to protect it.

End of life (EOL) and *end of support* can apply to either software or hardware. In the context of asset retention, they apply directly to hardware assets. Most vendors refer to EOL as the time when they stop producing and offering a product for sale. However, they will still support the products they've sold, at least for a while. End of support refers to the time when this support ends. Most hardware is on a refresh cycle based on the EOL and end-of-

support time frames. Organizations sometimes retain legacy hardware to access older data, such as data on tape drives.

Real World Scenario

Retention Policies Can Reduce Liabilities

Saving data longer than necessary also presents unnecessary legal issues. As an example, aircraft manufacturer Boeing was once the target of a class action lawsuit. Attorneys for the claimants learned that Boeing had a warehouse filled with 14,000 email backup tapes and demanded the relevant tapes. Not all the tapes were relevant to the lawsuit, but Boeing had to first restore the 14,000 tapes and examine the content before they could turn them over. Boeing ended up settling the lawsuit for $92.5 million, and analysts speculated that there would have been a different outcome if those 14,000 tapes hadn't existed.

The Boeing lawsuit is an extreme example, but it's not the only one. These events have prompted many companies to implement aggressive email retention policies. It is not uncommon for an email policy to require the deletion of all emails older than six months. These policies are often implemented using automated tools that search for old emails and delete them without any user or administrator intervention.

It is important, however, to understand that companies may *never* delete data when they can reasonably anticipate litigation. In fact, companies who believe that legal action may be forthcoming have a proactive obligation to preserve data and suspend any automated processes that might delete data.

Data Protection Methods

One of the primary methods of protecting the confidentiality of data is encryption, as discussed in the "Understanding Data States" section, earlier in this chapter. DLP methods (discussed in the "Data Loss Prevention" section, earlier in this chapter) help prevent data from leaving the network or even leaving a computer system. This section covers some additional data protection methods.

Digital Rights Management

Digital rights management (DRM) methods attempt to provide copyright protection for copyrighted works. The purpose is to prevent the unauthorized use, modification, and distribution of copyrighted works such as intellectual property. Here are some methods associated with DRM solutions:

DRM License A license grants access to a product and defines the terms of use. A DRM license is typically a small file that includes the terms of use, along with a decryption key that unlocks access to the product.

Persistent Online Authentication Persistent online authentication (also known as always-on DRM) requires a system to be connected with the Internet to use a product. The system periodically connects with an authentication server, and if the connection or authentication fails, DRM blocks the use of the product.

Continuous Audit Trail A continuous audit trail tracks all use of a copyrighted product. When combined with persistence, it can detect abuse, such as concurrent use of a product simultaneously but in two geographically different locations.

Automatic Expiration Many products are sold on a subscription basis. For example, you can often rent new streaming movies, but these are only available for a limited time, such as 30 days. When the subscription period ends, an automatic expiration function blocks any further access.

As an example, imagine you dreamed up a fantastic idea for a book. When you awoke, you vigorously wrote down everything you remembered. In the following year, you spent every free moment you had developing the idea and eventually published your book. To make it easy for some people to read your book, you included a Portable Document Format (PDF) version of the book. You were grateful to see it skyrocket onto bestseller lists. You're on track for financial freedom to develop another great idea that came to you in another dream.

Unfortunately, someone copied the PDF file and posted it on the dark web. People from around the world found it and then began selling it online for next to nothing, claiming that they had your permission to do so. Of course, you didn't give them permission. Instead, they were collecting money from your year of work, while your revenue sales began to tumble.

This type of copying and distribution, commonly called pirating, has enriched criminals for years. Not only do they sell books they didn't write, but they also copy and sell music, videos, video games, software, and more.

Some DRM methods attempt to prevent the copying, printing, and forwarding of protected materials. Digital watermarks are sometimes placed within audio or video files using steganography. They don't prevent copying but can be used to detect the unauthorized copying of a file. They can also be used for copyright enforcement and prosecution. Similarly, metadata is sometimes placed into files to identify the buyer.

Many organizations and individuals are opposed to DRM. They claim it restricts the fair use of materials they purchase. For example, after paying for some songs, they want to copy them onto both an MP3 player and a smartphone. Additionally, people against DRM claim it isn't effective against people that want to bypass it but instead complicates the usage for legitimate users.

Chapter 4 covers intellectual property, copyrights, trademarks, patents, and trade secrets in more depth. DRM methods are used to protect copyrighted data, but they aren't used to protect trademarks, patents, or trade secrets.

Cloud Access Security Broker

A cloud access security broker (CASB) is software placed logically between users and cloud-based resources. It can be on-premises or within the cloud. Anyone who accesses the cloud goes through the CASB software. It monitors all activity and enforces administrator-defined security policies.

As a simple example, imagine a company has decided to use a cloud provider for data storage but management wants all data stored in the cloud to be encrypted. The CASB can monitor all data going to the cloud and ensure that it arrives and is stored in an encrypted format.

A CASB would typically include authentication and authorization controls and ensure only authorized users can access the cloud resources. The CASB can also log all access, monitor activity, and send alerts on suspicious activity. In general, any security controls that an organization has created internally can be replicated to a CASB. This includes any DLP functions implemented by an organization.

CASB solutions can also be effective at detecting *shadow IT*. Shadow IT is the use of IT resources (such as cloud services) without the approval of, or even the knowledge of, the IT department. If the IT department doesn't know about the usage, it can't manage it. One way a CASB solution can detect shadow IT is by collecting and analyzing logs from network firewalls and web proxies. Chapter 16, “Managing Security Operations,” covers other cloud topics.

Pseudonymization

Pseudonymization refers to the process of using pseudonyms to represent other data. When pseudonymization is performed effectively, it can result in less stringent requirements that would otherwise apply under the European Union (EU) General Data Protection Regulation (GDPR), covered in Chapter 4.

The EU GDPR replaced the European Data Protection Directive (Directive 95/46/EC), and it became enforceable on May 25, 2018. It applies to all EU member states and to all countries transferring data to and from the EU and anyone residing in the EU.

A pseudonym is an alias. As an example, *Harry Potter* author J. K. Rowling published a book titled *The Cuckoo's Calling* under the pseudonym of Robert Galbraith. No one knew it was her, at least for a few months. Someone leaked that Galbraith was a pseudonym, and her agent later confirmed the rumor. Now, if you know the pseudonym, you'll know that any books attributed to Robert Galbraith are written by J. K. Rowling.

Similarly, pseudonymization can prevent data from directly identifying an entity, such as a person. As an example, consider a medical record held by a doctor's office. Instead of including personal information such as the patient's name, address, and phone number, it could just refer to the patient as Patient 23456 in the medical record. The doctor's office still needs this personal information, and it could be held in another database linking it to the patient pseudonym (Patient 23456).

Note that in the example, the pseudonym (Patient 23456) refers to several pieces of information on the person. It's also possible for a pseudonym to refer to a single piece of information. For example, you can use one pseudonym for a first name and another pseudonym for a last name. The key is to have another resource (such as another database) that allows you to identify the original data using the pseudonym.

The doctor's office can release pseudonymized data to medical researchers without compromising patients' privacy information. However, the doctor's office can still reverse the process to discover the original data if necessary.

The GDPR refers to pseudonymization as replacing data with artificial identifiers. These artificial identifiers are pseudonyms.

Tokenization

Tokenization is the use of a token, typically a random string of characters, to replace other data. It is often used with credit card transactions.

As an example, imagine Becky Smith has associated a credit card with her smartphone. Tokenization with a credit card typically works like this:

Registration When she first associated the credit card with her smartphone's digital wallet, the digital wallet's service provider securely sent the actual credit card information to a payment network (Visa, Mastercard, American Express) or the issuing bank. The payment network sent the credit card number to a tokenization vault controlled by the payment network or a third-party service provider. The vault created a token (a string of characters) and recorded the token along with the encrypted credit card number, and associated it with the user's phone. The token was then sent to the digital wallet provider, which saved it to her smartphone.

Usage Later, Becky goes to a Starbucks and buys a cup of coffee with her smartphone. Her smartphone passes the token to the point-of-sale (POS) system. The POS system sends the token to the credit card processor to authorize the charge.

Validation The credit card processor sends the token to the tokenization vault. The vault answers with the unencrypted credit card data, and the credit card processor then processes the charge.

Completing the Sale The credit card processor sends only a reply to the POS system indicating the charge is approved or declined and, if approved, credits the seller for the purchase.

In the past, credit card data has been intercepted and stolen at the POS system. However, when tokenization is used, the credit card number is never used or known to the POS system. The user transfers it once to the payment network, and the payment network stores an encrypted copy of the credit card number along with a token matched to this credit card. Later, the user presents

the token, and the payment processor validates the token through the tokenization vault.

E-commerce sites that have recurring charges also use tokenization. Instead of the e-commerce site collecting and storing credit card data, the site obtains a token from the payment gateway or processor. The token is created by a tokenization service, which stores an encrypted copy of the credit card data and sends the token to the e-commerce site. The site processes a charge the same way as it does for a POS system. However, the e-commerce site doesn't hold any sensitive data. Even if an attacker obtained a token and tried to make a charge with it, it would fail because the charges are only accepted from the e-commerce site.

Tokenization is similar to pseudonymization. Pseudonymization uses pseudonyms to represent other data. Tokenization uses tokens to represent other data. Neither the pseudonym nor the token has any meaning or value outside the process that creates them and links them to the other data. Pseudonymization is most useful when releasing a dataset to a third party (such as researchers aggregating data) without releasing any privacy-related data to the third party. Tokenization allows a third party (such as a payment network) to know the token and the original data. However, no one else knows both the token and the original data.

Anonymization

If you don't need personal data, another option is to use anonymization. *Anonymization* is the process of removing all relevant data so that it is theoretically impossible to identify the original subject or person. If done effectively, the GDPR is no longer relevant for the anonymized data. However, it can be difficult to truly anonymize the data. Data inference techniques may be able to identify individuals, even if personal data is

removed. This is sometimes referred to as reidentification of anonymized data.

As an example, consider a database that includes a listing of all the actors who have starred or co-starred in movies in the last 75 years, along with the money they earned for each movie. The database has three tables. The Actor table includes the actor names, the Movie table lists the movie names, and the Payment table reports the amount of money each actor earned for each movie. The three tables are linked so that you can query the database and easily identify how much money any actor earned for any movie.

If you removed the names from the Actor table, it no longer includes personal data, but it is not truly anonymized. For example, Gene Hackman has been in more than 70 movies, and no other actor has been in all the same movies. If you identify those movies, you can now query the database and learn exactly how much he earned for each of those movies. Even though his name was removed from the database, and that was the only obvious personal data in the database, data inference techniques can identify records applying to him.

Randomized masking can be an effective method of anonymizing data. Randomized masking swaps (shuffles) data in individual data columns so that records no longer represent the actual data. However, the data still maintains aggregate values that can be used for other purposes, such as scientific purposes. As an example, Table 5.2 shows four records in a database with the original values. An example of aggregated data is the average age of the four people, which is 29.

TABLE 5.2 Unmodified data within a database

First Name	Last Name	Age
Joe	Smith	25
Sally	Jones	28
Bob	Johnson	37
Maria	Doe	26

Table 5.3 shows the records after data has been swapped around, effectively masking the original data. Notice that this becomes a random set of first names, a random set of last names, and a random set of ages. It looks like real data, but none of the columns relate to each other. However, it is still possible to retrieve aggregated data from the table. The average age is still 29.

TABLE 5.3 Masked data

First Name	Last Name	Age
Sally	Doe	37
Maria	Johnson	25
Bob	Smith	28
Joe	Jones	26

Someone familiar with the dataset may be able to reconstruct some of the data if the table has only three columns and only four records. However, this is an effective method of anonymizing data if the table has a dozen columns and thousands of records.

Unlike pseudonymization and tokenization, anonymization cannot be reversed. After the data is randomized using an anonymization process, it cannot be returned to the original state.

Understanding Data Roles

Many people within an organization manage, handle, and use data, and they have different requirements based on their roles. Different documentation refers to these roles a little differently. Some of the terms you may see match the terminology used in some NIST documents, and other terms match some of the terminology used in the EU GDPR. When appropriate, we've listed the source so that you can dig into these terms a little deeper if desired.

One of the most important concepts here is ensuring that personnel know who owns information and assets. The owners have a primary responsibility of protecting the data and assets.

Data Owners

The *data owner* is the person who has ultimate organizational responsibility for data. The owner is typically the chief executive officer (CEO), president, or a department head. Data owners identify the classification of data and ensure that it is labeled properly. They also ensure that it has adequate security controls based on the classification and the organization's security policy requirements. Owners may be liable for negligence if they fail to perform due diligence in establishing and enforcing security policies to protect and sustain sensitive data.

NIST SP 800-18, Rev. 1—Guide for Developing Security Plans for Federal Information Systems outlines the following responsibilities for the information owner, which can be interpreted the same as the data owner:

- Establishes the rules for appropriate use and protection of the subject data/information (rules of behavior)
- Provides input to information system owners regarding the security requirements and security controls for the information system(s) where the information resides
- Decides who has access to the information system and with what types of privileges or access rights
- Assists in the identification and assessment of the common security controls where the information resides

NIST SP 800-18 frequently uses the phrase "rules of behavior," which is effectively the same as an acceptable use policy (AUP). Both outline the responsibilities and expected behavior of individuals and state the consequences of not complying with the rules or AUP. Additionally, individuals are required to periodically acknowledge that they have read, understand, and agree to abide by the rules or AUP. Many organizations post these on a website and allow users to acknowledge that they understand and agree to abide by them using an online electronic digital signature.

Data Controllers and Processors

Data controllers are the persons and organizations responsible for the collection and use of data. In the language of GDPR, "the data controller determines the purposes for which and the means by which personal data is processed." In other words, the data controller is the entity that determines the "how" and the "why" of personal data collection and use. This is true even if the data controller doesn't handle the data themselves.

In many cases, data controllers outsource some data handling tasks to other organizations. These organizations are known as *data processors*. Under GDPR, a data processor is "a natural or legal person, public authority, agency, or other body, which processes personal data solely on behalf of the data controller."

As an example, an employer that collects personal information on employees for payroll is a data controller. If they pass this information to a third-party company to process payroll, the payroll company is the data processor. In this example, the payroll company (the data processor) must not use the data for anything other than processing payroll at the direction of the data controller.

The GDPR restricts data transfers to countries outside the EU. Companies that violate privacy rules in the GDPR may face fines

of up to 4 percent of their global revenue or 20 million Euros, whichever is higher. Unfortunately, the GDPR is filled with legalese, presenting many challenges for organizations. As an example, clause 107 includes this single sentence statement:

> Consequently the transfer of personal data to that third country or international organisation should be prohibited, unless the requirements in this Regulation relating to transfers subject to appropriate safeguards, including binding corporate rules, and derogations for specific situations are fulfilled.

As a result, many organizations have created dedicated roles, such as a data privacy officer (DPO), to oversee the control of data and ensure the organization follows all relevant laws and regulations. The GDPR has mandated the role of a data protection officer for any organization that must comply with the GDPR. The person in this role is responsible for ensuring the organization applies the laws to protect individuals' private data.

Data Custodians

Data owners often delegate day-to-day tasks to a *data custodian*. A custodian helps protect the integrity and security of data by ensuring that it is properly stored and protected. For example, custodians would ensure that the data is backed up by following guidelines in a backup policy. If administrators have configured auditing on the data, custodians would also maintain these logs.

In practice, personnel within an IT department or system security administrators would typically be the custodians. They might be the same administrators responsible for assigning permissions to data.

Users and Subjects

A *user* is any person who accesses data via a computing system to accomplish work tasks. Users should have access only to the data they need to perform their work tasks. You can also think of users as employees or end users.

The GDPR defines a *data subject* as a person who can be identified through an identifier, such as a name, an identification number, location data, an online identifier, or other means. As an example, if a file includes PII on Sally Smith, Sally Smith is the data subject.

Using Security Baselines

Once an organization has identified and classified its assets, it will typically want to secure them. That's where security baselines come in. Baselines provide a starting point and ensure a minimum security standard. One common baseline that organizations use is imaging. Chapter 16 covers system imaging in the context of configuration management in more depth. As an introduction, administrators configure a single system with desired settings, capture it as an image, and then deploy the image to other systems. This ensures that systems are deployed in a similar secure state, which helps to protect the privacy of data.

After deploying systems in a secure state, auditing processes periodically check the systems to ensure they remain in a secure state. For example, Microsoft Group Policy can periodically check systems and reapply settings to match the security baseline.

NIST SP 800-53, Rev. 5—Security and Privacy Controls for Information Systems and Organizations mentions *security control baseline* and identifies it as the set of minimum security controls defined for an information system. It stresses that a single set of security controls does not apply to all situations. Still, any organization can select a set of baseline security controls and tailor the baseline to its needs. NIST SP 800-53B—Control Baselines for Information Systems and Organizations includes a comprehensive list of security controls and has identified many of them to include in various baselines. Specifically, they present three security control baselines (determined by the impact level of the system) and a privacy control baseline. These are based on the potential impact to an

organization's mission if there is a loss of confidentiality, integrity, or availability of a system. The baselines are as follows:

Low-Impact System Controls in this baseline are recommended if any loss of confidentiality, integrity, and/or availability will have a low impact on the organization's mission.

Moderate-Impact System Controls in this baseline are recommended if it is possible that a loss of confidentiality, integrity, or availability will have a moderate impact on the organization's mission.

High-Impact System Controls in this baseline are recommended if it is possible that a loss of confidentiality, integrity, or availability will have a high impact on the organization's mission.

Privacy Control Baseline This baseline provides an initial baseline for any systems that process PII. Organizations may combine this baseline with one of the other baselines.

These refer to the worst-case potential impact if a system is compromised and a data breach occurs. For example, imagine a system is compromised. You would try to predict the impact of the compromise on the confidentiality, integrity, or availability of the system and any data it holds:

- If the compromise would cause privacy data to be compromised, you would consider adding the security controls identified as privacy control baseline items to your baseline.
- If the impact is low for all three of the security objectives, you would consider adding the security controls identified as low-impact controls to your baseline.
- If the impact of this compromise is moderate for any one of the security objectives, you would consider adding the security controls identified as moderate-impact, in addition to the low-impact controls.
- If the impact is high for any one of the security objectives, you would consider adding all the controls listed as high-

impact in addition to the low-impact and moderate-impact controls.

It's worth noting that many of the items in these lists are basic security practices. Additionally, implementing basic security principles such as the least privilege principle shouldn't surprise anyone. Of course, just because these are basic security practices, it doesn't mean organizations implement them. Unfortunately, many organizations have yet to discover or enforce the basics.

Comparing Tailoring and Scoping

After selecting a control baseline, organizations fine-tune it with tailoring and scoping processes. A big part of the tailoring process is aligning the controls with an organization's specific security requirements. As a comparison, think of a clothes tailor who alters or repairs clothes. If a person buys a suit at a high-end retailer, a tailor modifies the suit to fit the person perfectly. Similarly, tailoring a baseline ensures it is a good fit for the organization.

Tailoring refers to modifying the list of security controls within a baseline to align with the organization's mission. NIST SP 800-53B formally defines it as “part of an organization-wide risk management process that includes framing, assessing, responding to, and monitoring information security and privacy risks” and indicates it includes the following activities:

- Identifying and designating common controls
- Applying scoping considerations
- Selecting compensating controls
- Assigning values to organization-defined control parameters via explicit assignment and selection operations
- Supplementing baselines with additional controls and control enhancements
- Providing specification information for control implementation

A selected baseline may not include commonly implemented controls. However, just because a security control isn't included in the baseline doesn't mean it should be removed. For example, imagine that a data center includes video cameras covering the external entry, the internal exit, and every row of servers, but the baseline only recommends a video camera cover the external entry. During the tailoring process, personnel will evaluate these extra cameras and determine if they are needed. They may decide to remove some to save costs or keep them.

An organization might decide that a set of baseline controls applies perfectly to computers in their central location but that some controls aren't appropriate or feasible in a remote office location. In this situation, the organization can select compensating security controls to tailor the baseline to the remote site. For example, imagine the account lockout policy is set to lock out users if they enter an incorrect password five times. In this example, the control value is 5, but the tailoring process may change it to 3.

Scoping is a part of the tailoring process and refers to reviewing a list of baseline security and privacy controls and selecting only those security and privacy controls that apply to the IT systems you're trying to protect. Or, in the simplest terms, scoping processes eliminate controls that are recommended in a baseline. For example, if a system doesn't allow any two people to log on to it simultaneously, there's no need to apply a concurrent session control. During this part of the tailoring process, the organization looks at every control in the baseline and vigorously defends (in writing) any decision to omit a control from the baseline.

Standards Selection

When selecting security controls within a baseline, or otherwise, organizations need to ensure that the controls comply with external security standards. External elements typically define compulsory requirements for an organization. For example, the Payment Card Industry Data Security Standard (PCI DSS) defines requirements that businesses must follow to process major credit cards. Similarly, organizations that collect or process data

belonging to EU citizens must abide by the requirements in the GDPR.

Obviously, not all organizations have to comply with these standards. Organizations that don't store, process, or transmit payment card transactions do not need to comply with PCI DSS. Similarly, organizations that do not collect or process EU citizens' data do not need to comply with GDPR requirements. Organizations need to identify the standards that apply and ensure that the security and privacy controls they select fully comply with those standards.

Even if your organization isn't legally required to comply with a specific standard, using a well-designed community standard can be helpful. For example, U.S. government organizations are required to comply with many of the standards published by NIST SP 800 documents. These same documents are used by many organizations in the private sector to help them develop and implement their own security standards.

Summary

Asset security focuses on collecting, handling, and protecting information throughout its life cycle. This includes sensitive information stored or processed on computing systems or transferred over a network and the assets used in these processes. Sensitive information is any information that an organization keeps private and can include multiple levels of classifications. Proper destruction methods ensure that data can't be retrieved after destruction.

Data protection methods include digital rights management (DRM) and using cloud access security brokers (CASBs) when using cloud resources. DRM methods attempt to protect copyrighted materials. A CASB is software placed logically between users and cloud-based resources. It can ensure that cloud resources have the same protections as resources within a network. Entities that must comply with the EU GDPR use additional data protection methods such as pseudonymization, tokenization, and anonymization.

Personnel can fulfill many different roles when handling data. Data owners are ultimately responsible for classifying, labeling, and protecting data. System owners are responsible for the systems that process the data. The GDPR defines data controllers, data processors, and data custodians. Data controllers decide what data to process, the purpose of data collection, and how to process it. A data controller can hire a third party to process data, and in this context, the third party is the data processor. Data processors have a responsibility to protect the privacy of the data and not use it for any purpose other than directed by the data controller. A custodian is delegated day-to-day responsibilities for properly storing and protecting data.

Security baselines provide a set of security controls that an organization can implement as a secure starting point. Some publications (such as NIST SP 800-53B) identify security control baselines. However, these baselines don't apply equally to all organizations. Instead, organizations use scoping and tailoring techniques to identify the security controls to implement after selecting baselines. Additionally, organizations ensure that they implement security controls mandated by external standards that apply to their organization.

Study Essentials

Understand the importance of data and asset classifications. Data owners are responsible for defining data and asset classifications and ensuring that data and systems are properly tagged. Additionally, data owners define requirements to protect data at different classifications, such as encrypting sensitive data at rest, in transit, and in use. Data classifications are typically defined within security policies or data policies.

Define PII and PHI. Personally identifiable information (PII) is any information that can identify an individual. Protected health information (PHI) is any health-related information that can be related to a specific person and is subject to HIPAA. Many laws and regulations mandate the protection of PII and PHI.

Know how to manage sensitive information. Sensitive information is any type of classified information, and proper management helps prevent unauthorized disclosure resulting in a loss of confidentiality. Proper management includes tagging, handling, storing, and destroying sensitive information. The two areas where organizations often miss the mark are adequately protecting backup media holding sensitive information and sanitizing media or equipment when it is at the end of its life cycle.

Describe the three data states. The three data states are at rest, in transit, and in use. Data at rest is any data stored on media such as hard drives or external media. Data in transit is any data transmitted over a network. Encryption methods protect data at rest and in transit. Data in use refers to data in memory and used by an application. Applications should flush memory buffers to remove data after it is no longer needed.

Define DLP. Data loss prevention (DLP) solutions detect and block data exfiltration attempts by scanning unencrypted files and looking for keywords and data patterns. Network DLP solutions (including cloud DLP solutions) scan files before they leave the network. Endpoint DLP solutions prevent users from copying or printing some files.

Compare data destruction methods. Erasing a file doesn't delete it. Clearing media overwrites it with characters or bits. Purging repeats the clearing process multiple times and removes data so that the media can be reused. Degaussing removes data from tapes and magnetic hard disk drives, but it does not affect optical media or SSDs. Destruction methods include incineration, shredding, and disintegration, pulverizing, and melting.

Describe data remanence. Data remanence is the data that remains on media after it should have been removed. Hard disk drives sometimes retain residual magnetic flux that can be read with advanced tools. Advanced tools can read slack space on a disk, which is unused space in clusters. Erasing data on a disk leaves data remanence. For solid-state drives (SSDs), data remanence can persist due to the wear-leveling algorithms they employ, making traditional data erasure methods less effective

and potentially allowing remnants of data to remain on unaddressed memory cells.

Understand record retention policies. Record retention policies ensure that data is kept in a usable state while it is needed and destroyed when it is no longer needed. Many laws and regulations mandate keeping data for a specific amount of time, but in the absence of formal regulations, organizations specify the retention period within a policy. Audit trail data needs to be kept long enough to reconstruct past incidents, but the organization must identify how far back they want to investigate. A current trend in many organizations is to reduce legal liabilities by implementing short retention policies with email.

Know the difference between end of life and end of support. End of life (EOL) is the date announced by a vendor when production and sales of a product stop. However, the vendor still supports the product after EOL. End of support identifies the date when a vendor will no longer support a product.

Explain DRM. Digital rights management (DRM) methods provide copyright protection for copyrighted works. The purpose is to prevent the unauthorized use, modification, and distribution of copyrighted works.

Explain CASB. A cloud access security broker (CASB) is a solution placed logically between users and cloud resources. It can apply internal security controls to cloud resources. The CASB solution can be placed on-premises or in the cloud.

Define pseudonymization. Pseudonymization is the process of replacing some data elements with pseudonyms or aliases. It removes privacy data so that a dataset can be shared. However, the original data remains available in a separate dataset.

Define tokenization. Tokenization replaces data elements with a string of characters or a token. Credit card processors replace credit card data with a token, and a third party holds the mapping to the original data and the token.

Define anonymization. Anonymization replaces privacy data with useful but inaccurate data. The dataset can be shared and used for analysis purposes, but anonymization removes individual identities. Anonymization is permanent.

Know the responsibilities of data roles. The data owner is the person responsible for classifying, labeling, and protecting data. Data controllers decide what data to process, the purpose of data collection, and how to process data. Data processors are third-party entities that process data for an organization at the direction of the data controller. A user accesses data while performing work tasks. The data subject is the person described in the PII. A custodian has day-to-day responsibilities for protecting and storing data.

Know about security control baselines. Security control baselines provide a listing of controls that an organization can apply as a baseline. Not all baselines apply to all organizations. Organizations apply scoping and tailoring techniques to adapt a baseline to their needs.

Written Lab

1. Describe sensitive data.
2. Identify the difference between EOL and EOS.
3. Identify common uses of pseudonymization, tokenization, and anonymization.
4. Describe the difference between scoping and tailoring.

Review Questions

1. Which of the following provides the best protection against the loss of confidentiality for sensitive data?

    A. Data labels

    B. Data classifications

    C. Data handling