This page intentionally left blank

PART II

Asset Security

This page intentionally left blank

CHAPTER 5

Assets

This chapter presents the following:

- Identification and classification of information and assets
- Information and asset handling requirements
- Secure resource provisioning
- The data life cycle
- Data compliance requirements

You don't know what you've got till it's gone.

—Joni Mitchell

An asset is, by definition, anything of worth to an organization. This includes people, partners, equipment, facilities, reputation, and information. We already touched on the importance of some of these assets when we addressed risk in Chapter 2. While every asset needs to be protected, our coverage of the second CISSP domain in this chapter and the next one focuses a bit more narrowly on protecting information assets. This is because, apart from people, information is typically the most valuable asset to an organization. It lies at the heart of every information system, so precision focus on its protection makes a lot of sense.

Information, of course, exists in context; it is acquired or created at a particular point in time through a specific process and (usually) for a purpose. It moves through an organization's information systems, sometimes adding value to processes and sometimes waiting to be useful. Eventually, the information outlives its utility (or becomes a liability) and must be disposed of appropriately. We start off our discussion of asset security by addressing two fundamental questions: "What do we have?" and "Why should we care?" The first question is probably rather obvious, since we cannot protect that of which we're not aware. The second question may sound flippant, but it really gets to the heart of how important an asset is to the organization. We've already tackled this (at least with regard to data) in Chapter 4 in our discussion of the categorize step of the NIST Risk Management Framework. Data and asset classification, as we will shortly see, is very similar to the categorization we've already explored. Let's get to it!

EXAM TIP An information asset can be either the data, the device on which it is stored and used, or both. In the exam, when you see the term *asset* by itself, it typically means only the device.

Information and Assets

An *asset* can be defined as anything that is useful or valuable. In the context of products and services, this value is usually considered financially: how much would someone pay for it minus how much does the thing cost. If that value is positive, we call the thing an asset. However, if that value is negative (that is, the thing costs more than what someone would pay for it), then we call the thing a liability. Clearly, assets can be both tangible things like computers and firewalls and intangible things like data or reputation. It is important to narrow down the definition for purposes of the CISSP exam, so in this domain, we consider assets as tangible things and we deal with data separately.

Information is a set of data items, placed in a context, and having some meaning. Data is just an item. It could be the word "yes," the time "9:00," or the name "Fernando's Café" and, by itself, has no meaning. Put this data together in the context of an answer to the question "Would you like to have coffee tomorrow morning?" and now we have information. Namely, that we'll be sharing a beverage tomorrow morning at a particular place. Data processing yields information, and this is why we often use these two terms interchangeably when talking about security issues.

Identification

Whether we are concerned with data security or asset security (or both), we first have to know what we have. Identification is simply establishing what something is. When you look at a computing device occupying a slot in your server rack, you may want to know what it is. You may want to identify it. The most common way of doing this is by placing tags on our assets and data. These tags can be physical (e.g., stickers), electronic (e.g., radio frequency identification [RFID] tags), or logical (e.g., software license keys). Using tags is critically important to establishing and maintaining accurate inventories of our assets.

But what about data? Do we need to identify it and track it like we do with our more tangible assets? The answer is: it depends. Most organizations have at least some data that is so critical that, were it to become lost or corrupted or even made public, the impact would be severe. Think of financial records at a bank, or patient data at a healthcare provider. These organizations would have a very bad day indeed if any of those records were lost, inaccurate, or posted on the dark web. To prevent this, they go to great lengths to identify and track their sensitive information, usually by using metadata embedded in files or records.

While it may not be critical (or even feasible) for many organizations to identify all their information, it is critical to most of us to at least decide how much effort should be put into protecting different types of data (or assets, for that matter). This is where classification comes in handy.

Classification

Classification just means saying that something belongs to a certain class. We could say, for example, that your personnel file belongs to the class named "private" and that your organization's marketing brochure for the latest appliance belongs to the class "public." Right away, we would have a sense that your file has more value to your organization than the brochure. The rationale behind assigning values to different assets and data is that this enables an organization to gauge the amount of funds and resources that should go toward protecting each class, because not all assets and data have the same value to an organization. After identifying all important data, it should be properly classified. An organization copies and creates a lot of data that it must maintain, so classification is an ongoing process and not a one-time effort.

Data Classification

An important metadata item that should be attached to all our information is a classification level. This classification tag, which remains attached (and perhaps updated) throughout the life cycle of the data, is important to determining the protective controls we apply to the data.

Information can be classified by sensitivity, criticality, or both. Either way, the classification aims to quantify how much loss an organization would likely suffer if the information was lost. The *sensitivity* of information is commensurate with the losses to an organization if that information was revealed to unauthorized individuals. This kind of compromise has made headlines in recent years with the losses of information suffered by organizations such as Equifax, Sina Weibo, and Marriott International. In each case, the organizations lost trust and had to undertake expensive responses because sensitive data was compromised.

The *criticality* of information, on the other hand, is an indicator of how the loss of the information would impact the fundamental business processes of the organization. In other words, critical information is that which is essential for the organization to continue operations. For example, Code Spaces, a company that provided code repository services, was forced to shut down in 2014 after an unidentified individual or group deleted its code repositories. This data was critical to the operations of the company and, without it, the corporation had no choice but to go out of business.

Once data is segmented according to its sensitivity or criticality level, the organization can decide what security controls are necessary to protect different types of data. This ensures that information assets receive the appropriate level of protection, and classifications indicate the priority of that security protection. The primary purpose of data classification is to indicate the level of confidentiality, integrity, and availability protection that is required for each type of data set. Many people mistakenly only consider the confidentiality aspects of data protection, but we need to make sure our data is not modified in an unauthorized manner and that it is available when needed.

Data classification helps ensure that data is protected in the most cost-effective manner. Protecting and maintaining data costs money, but spending money for the information that actually requires protection is important. If you were in charge of making sure Russia does not know the encryption algorithms used when transmitting information to and

from U.S. spy satellites, you would use more extreme (and expensive) security measures than you would use to protect your peanut butter and banana sandwich recipe from your next-door neighbor.

Each classification should have separate handling requirements and procedures pertaining to how that data is accessed, used, and destroyed. For example, in a corporation, confidential information may be accessed only by senior management and a select few trusted employees throughout the company. Accessing the information may require two or more people to enter their access codes. Auditing could be very detailed and its results monitored daily, and paper copies of the information may be kept in a vault. To properly erase this data from the media, degaussing or overwriting procedures may be required. Other information in this company may be classified as sensitive, allowing a slightly larger group of people to view it. Access control on the information classified as sensitive may require only one set of credentials. Auditing happens but is only reviewed weekly, paper copies are kept in locked file cabinets, and the data can be deleted using regular measures when it is time to do so. Then, the rest of the information is marked public. All employees can access it, and no special auditing or destruction methods are required.

EXAM TIP Each classification level should have its own handling and destruction requirements.

Classification Levels There are no hard and fast rules on the classification levels that an organization should use. Table 5-1 explains the types of classifications available. An organization could choose to use any of the classification levels presented in Table 5-1. One organization may choose to use only two layers of classifications, while another organization may choose to use four. Note that some classifications are more commonly used for commercial businesses, whereas others are military classifications.

The following are the common levels of sensitivity from the highest to the lowest for commercial business:

- Confidential
- Private
- Sensitive
- Public

And here are the levels of sensitivity from the highest to the lowest for military purposes:

- Top secret
- Secret
- Confidential
- Controlled unclassified information
- Unclassified

Classification	Definition	Example	Organizations That Would Use This
Public	• Disclosure is not welcome, but it would not cause an adverse impact to company or personnel.	• How many people are working on a specific project • Upcoming projects	Commercial business
Sensitive	• Requires special precautions to ensure the integrity and confidentiality of the data by protecting it from unauthorized modification or deletion. • Requires higher-than-normal assurance of accuracy and completeness.	• Financial information • Details of projects • Profit earnings and forecasts	Commercial business
Private	• Personal information for use within a company. • Unauthorized disclosure could adversely affect personnel or the company.	• Work history • Human resources information • Medical information	Commercial business
Confidential	• For use within the company only. • Data exempt from disclosure under the Freedom of Information Act or other laws and regulations. • Unauthorized disclosure could seriously affect a company.	• Trade secrets • Healthcare information • Programming code • Information that keeps the company competitive	Commercial business Military
Unclassified	• Data is not sensitive or classified.	• Computer manual and warranty information • Recruiting information	Military
Controlled unclassified information (CUI)	• Sensitive, but not secret. • Information that cannot legally be made public.	• Health records • Answers to test scores	Military
Secret	• If disclosed, it could cause serious damage to national security.	• Deployment plans for troops • Unit readiness information	Military
Top secret	• If disclosed, it could cause grave damage to national security.	• Blueprints of new weapons • Spy satellite information • Espionage data	Military

Table 5-1 Commercial Business and Military Data Classifications

The classifications listed in Table 5-1 are *commonly* used in the industry, but there is a lot of variance. An organization first must decide the number of data classifications that best fit its security needs, then choose the classification naming scheme, and then define what the names in those schemes represent. Company A might use the classification level "confidential," which represents its most sensitive information. Company B might use "top secret," "secret," and "confidential," where confidential represents its least sensitive information. Each organization must develop an information classification scheme that best fits its business and security needs.

EXAM TIP The terms "unclassified," "secret," and "top secret" are usually associated with governmental organizations. The terms "private," "proprietary," and "sensitive" are usually associated with nongovernmental organizations.

It is important to not go overboard and come up with a long list of classifications, which will only cause confusion and frustration for the individuals who will use the system. The classifications should not be too restrictive either, because many types of data may need to be classified. As with every other issue in security, we must balance our business and security needs.

Each classification should be unique and separate from the others and not have any overlapping effects. The classification process should also outline how information is controlled and handled through its life cycle (from creation to termination).

NOTE An organization must make sure that whoever is backing up classified data—and whoever has access to backed-up data—has the necessary clearance level. A large security risk can be introduced if low-level technicians with no security clearance have access to this information during their tasks.

Once the scheme is decided upon, the organization must develop the criteria it will use to decide what information goes into which classification. The following list shows some criteria parameters an organization may use to determine the sensitivity of data:

- The usefulness of data
- The value of data
- The age of data
- The level of damage that could be caused if the data were disclosed
- The level of damage that could be caused if the data were modified or corrupted
- Legal, regulatory, or contractual responsibility to protect the data
- Effects the data has on security
- Who should be able to access the data
- Who should maintain the data
- Who should be able to reproduce the data
- Lost opportunity costs that could be incurred if the data were not available or were corrupted

Applications and sometimes whole systems may need to be classified. The applications that hold and process classified information should be evaluated for the level of protection they provide. You do not want a program filled with security vulnerabilities to process and "protect" your most sensitive information. The application classifications should be based on the assurance (confidence level) the organization has in the software and the type of information it can store and process.

CAUTION The classification rules must apply to data no matter what format it is in: digital, paper, video, fax, audio, and so on.

Asset Classification

Information is not the only thing we should classify. Consider that information must reside somewhere. If a confidential file is stored and processed in the CEO's laptop, then that device (and its hard drive if it is removed) should also be considered worthy of more protection. Typically, the classification of an asset (like a removable drive or a laptop) used to store or process information should be as high as the classification of the most valuable data in it. If an asset has public, sensitive, and confidential information, then that asset should be classified as private (the highest of the three classifications) and protected accordingly.

Classification Procedures

The following outlines the necessary steps for a proper classification program:

1. Define classification levels.
2. Specify the criteria that will determine how data is classified.
3. Identify data owners who will be responsible for classifying data.
4. Identify the data custodian who will be responsible for maintaining data and its security level.
5. Indicate the security controls, or protection mechanisms, required for each classification level.
6. Document any exceptions to the previous classification issues.
7. Indicate the methods that can be used to transfer custody of the information to a different data owner.
8. Create a procedure to periodically review the classification and ownership. Communicate any changes to the data custodian.
9. Indicate procedures for declassifying the data.
10. Integrate these issues into the security awareness program so all employees understand how to handle data at different classification levels.

Physical Security Considerations

We discuss data security in detail in Chapter 10. However, that data lives physically in devices and printed documents, both of which require protection also. The main threats that physical security components combat are theft, interruptions to services, physical damage, compromised system and environment integrity, and unauthorized access. Real loss is determined by the cost to replace the stolen items, the negative effect on productivity, the negative effect on reputation and customer confidence, fees for consultants that may need to be brought in, and the cost to restore lost data and production levels. Many times, organizations just perform an inventory of their hardware and provide value estimates that are plugged into risk analysis to determine what the cost to the organization would be if the equipment were stolen or destroyed. However, the data held within the equipment may be much more valuable than the equipment itself, and proper recovery mechanisms and procedures also need to be plugged into the risk assessment for a more realistic and fair assessment of cost. Let's take a look at some of the controls we can use in order to mitigate risks to our data and to the media on which it resides.

Protecting Mobile Devices

Mobile devices are almost indispensable. For most of us, significant chunks of our personal and work lives are chronicled in our smartphones or tablets. Employees who use these devices as they travel for work may have extremely sensitive company or customer data on their systems that can easily fall into the wrong hands. This problem can be mitigated to a point by ensuring our employees use company devices for their work, so we can implement policies and controls to protect them. Still, many organizations allow their staff members to bring their own devices (BYOD) to the workplace and/or use them for work functions. In these cases, it is not only security but also privacy that should receive serious attention.

There is no one-size-fits-all solution to protecting company, let alone personal, mobile devices. Still, the following list provides some of the mechanisms that can be used to protect these devices and the data they hold:

- Inventory all mobile devices, including serial numbers, so they can be properly identified if they are stolen and then recovered.
- Harden the operating system by applying baseline secure configurations.
- Stay current with the latest security updates and patches.
- Ensure mobile devices have strong authentication.
- Register all devices with their respective vendors, and file a report with the vendor when a device is stolen. If a stolen device is sent in for repairs after it is stolen, it will be flagged by the vendor if you have reported the theft.
- Do not check mobile devices as luggage when flying. Always carry them on with you.
- Never leave a mobile device unattended, and carry it in a nondescript carrying case.

- Engrave the device with a symbol or number for proper identification.
- Back up all data on mobile devices to an organizationally controlled repository.
- Encrypt all data on a mobile device.
- Enable remote wiping of data on the device.

Tracing software can be installed so that your device can "phone home" if it is taken from you. Several products offer this tracing capability. Once installed and configured, the software periodically sends in a signal to a tracking center or allows you to track it through a website or application. If you report that your device has been stolen, the vendor of this software may work with service providers and law enforcement to track down and return your device.

Paper Records

It is easy to forget that many organizations still process information on paper records. The fact that this is relatively rare compared to the volume of their electronic counterparts is little consolation when a printed e-mail with sensitive information finds its way into the wrong hands and potentially causes just as much damage. Here are some principles to consider when protecting paper records:

- Educate your staff on proper handling of paper records.
- Minimize the use of paper records.
- Ensure workspaces are kept tidy so it is easy to tell when sensitive papers are left exposed, and routinely audit workspaces to ensure sensitive documents are not exposed.
- Lock away all sensitive paperwork as soon as you are done with it.
- Prohibit taking sensitive paperwork home.
- Label all paperwork with its classification level. Ideally, also include its owner's name and disposition (e.g., retention) instructions.
- Conduct random searches of employees' bags as they leave the office to ensure sensitive materials are not being taken home.
- Destroy unneeded sensitive papers using a crosscut shredder, or consider contracting a document destruction company.

Safes

An organization may have need for a safe. Safes are commonly used to store backup data tapes, original contracts, or other types of valuables. The safe should be penetration resistant and provide fire protection. The types of safes an organization can choose from are

- **Wall safe** Embedded into the wall and easily hidden
- **Floor safe** Embedded into the floor and easily hidden

- **Chests** Stand-alone safes
- **Depositories** Safes with slots, which allow the valuables to be easily slipped in
- **Vaults** Safes that are large enough to provide walk-in access

If a safe has a combination lock, it should be changed periodically, and only a small subset of people should have access to the combination or key. The safe should be in a visible location, so anyone who is interacting with the safe can be seen. It should also be covered by a video surveillance system that records any activity around it. The goal is to uncover any unauthorized access attempts. Some safes have passive or thermal relocking functionality. If the safe has a *passive relocking* function, it can detect when someone attempts to tamper with it, in which case extra internal bolts will fall into place to ensure it cannot be compromised. If a safe has a *thermal relocking* function, when a certain temperature is met (possibly from drilling), an extra lock is implemented to ensure the valuables are properly protected.

Managing the Life Cycle of Assets

A life-cycle model describes the changes that an entity experiences during its lifetime. While it may seem odd to refer to assets as having a "life," the fact is that their utility for (and presence within) organizations can be described with clear start and end points. That is the lifetime of the asset within that organization (even if it gets refurbished and used elsewhere). After the asset departs, its utility is oftentimes transferred to its replacement even if the new asset is different than the original in meaningful ways. That new asset will, in turn, be replaced by something else, and so on.

The life cycle, which is shown in Figure 5-1, starts with the identification of a new requirement. Whoever identifies the new requirement either becomes its champion or

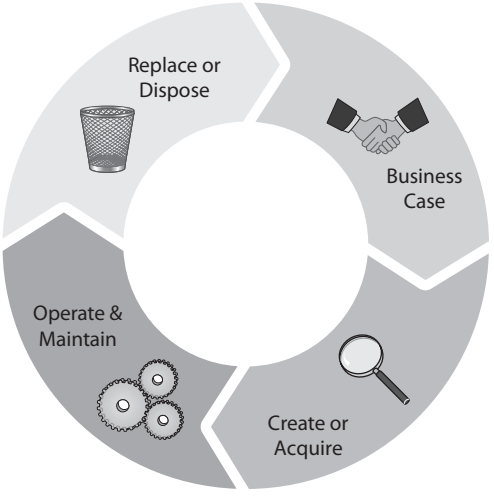


Figure 5-1 The IT asset life cycle

finds someone else to do so. The champion for this requirement then makes a business case for it that shows that the existing assets are unable to satisfy this need. The champion also explains why the organization really should get a new asset, which typically includes a conversation about risks and return on investment (ROI). If the champion is successful, senior management validates the requirement and identifies the needed resources (people, money, time).

The validated requirement then goes to a change management board, giving the different organizational stakeholders a say in what, how, and when the asset will be acquired. This board's goal is to ensure that this new asset doesn't break any processes, introduce undue risks, or derail any ongoing projects. In mature organizations, the change management process also attempts to look over the horizon and see what the long-term ramifications of this asset might be. After the board determines how to proceed, the new asset is either developed in-house or acquired from a vendor.

The third phase of asset management is also the longest one: operation and maintenance (O&M). Before the asset is put into operation, the IT and security operations teams configure it to balance three (sometimes competing) goals: it must be able to do whatever it was acquired to do, it must be able to do it without interfering or breaking anything else, and it must be secure. This configuration will almost certainly need to change over time, which is why we discuss configuration management in Chapter 20.

NOTE This initial part of the O&M phase is usually the most problematic for a new asset and is a major driver for the use of an integrated product team (IPT) such as DevOps, which we discuss in Chapter 24.

Eventually, the asset is no longer effective (in terms of function or cost) or required. At this point, it moves out of O&M and is retired. This move, as you may have already guessed, triggers another review by the change management board, because retiring the asset is likely to have effects on other resources or processes. Once the process of retirement is hashed out, the asset is removed from production. At this point, the organization needs to figure out what to do with the thing. If the asset stored any data, the data probably has to be purged. If the asset has any environmentally hazardous materials, it has to be properly discarded. If it might be useful to someone else, it might be donated or sold. At any rate, the loss of this asset may result in a new requirement being identified, which starts the whole asset management life cycle again, as shown in Figure 5-1.

Ownership

In most cases, whoever makes the business case for an asset ultimately owns it, but this is not always the case. Asset *ownership*, once the asset shows up and as long as it remains in the organization, entails responsibility for the effective management of the asset over its whole life cycle. Ownership in this sense is somewhat different than ownership in a strictly legal sense. The legal owner of a server could be the corporation that buys it, while the life cycle owner would be whatever employee or department is responsible for it on a day-to-day basis.

Inventories

One of the fundamental responsibilities for asset owners is to keep track of their assets. Though the approaches to tracking hardware and software vary, they are both widely recognized as critical controls. At the very least, it is very difficult to defend an asset that you don't know you have. As obvious as this sounds, many organizations lack an accurate and timely inventory of their hardware and software.

Tracking Hardware

Seemingly, maintaining awareness of which devices are in your organization should be an easier task than tracking your software. A hardware device can be seen, touched, and bar-scanned. It can also be sensed electronically once it is connected to the network. If you have the right tools and processes available, tracking hardware should not be all that difficult, right? Not so fast. It turns out that the set of problems ranges from supply chain security to insider threats and everything in between.

Let's start with the basics. How do you ensure that a new device you've ordered is the right one and free of back doors or piracy issues? There have been multiple reports in the news media recently of confirmed or suspected back doors installed in hardware assets by either manufacturers (e.g., pirated hardware) or by third parties (e.g., government spy agencies) before the assets get to the organization that acquired them. In response to these and other threats, the International Organization for Standardization published ISO 28000:2007 as a means for organizations to use a consistent approach to securing their supply chains. In essence, we want to ensure we purchase from trusted sources, use a trusted transportation network, and have effective inspection processes to mitigate the risk of pirated, tampered, or stolen hardware.

But even if we can assure ourselves that all the hardware we acquire is legitimate, how would we know if someone else were to add devices to our networks? Asset monitoring includes not only tracking our known devices but also identifying unknown ones that may occasionally pop up in our enclaves. Examples that come to mind from personal experience include rogue wireless access points, personal mobile devices, and even (believe it or not) telephone modems. Each introduces unknown (and thus unmitigated) risks. The solution is to have a comprehensive monitoring process that actively searches for these devices and ensures compliance with your organization's security policies.

In many cases, monitoring devices on the premises can be as simple as having a member of the security or IT team randomly walk through every space in the organization looking for things that are out of place. This becomes even more effective if this person does this after work hours and also looks for wireless networks as part of these walks. Alternatively, much of this monitoring can be done using device management platforms and a variety of sensors.

Tracking Software

Obviously, we can't just walk around and inventory our software. The unique challenges of tracking software are similar to those of managing hardware, but with a few important differences. Unlike hardware, software assets can be copied or installed multiple times. This could be a problem from a licensing perspective. Commercial applications typically

have limits on how many times you can install a single license. The terms of these licensing agreements vary wildly from single-use to enterprise-wide. It bears pointing out that tracking what software is installed on which systems, and for which users, is an important part of software asset management. Otherwise, you risk violating software licenses.

Using unlicensed software not only is unethical but also exposes an organization to financial liability from the legitimate product vendors. This liability can manifest in a number of ways, including having the organization reported to the vendor by a disgruntled employee. It could also come up when certain software packages "phone home" to the vendors' servers or when downloading software patches and updates. Depending on the number and types of licenses, this could end up costing significant amounts of money in retroactive licensing fees.

Pirated software is even more problematic because many forms of it include back doors installed by the pirates or are Trojan horses. Even if this were not the case, it would almost certainly be impossible to update or patch this software, which makes it inherently more insecure. Since no IT staff in their right mind would seriously consider using pirated software as an organizational policy, its presence on a network would suggest that at least some users have privileges that are being abused and to which they may not be entitled.

Another problem created by the fact that you can copy and install software on multiple systems, apart from unlicensed or pirated software, is security. If you lose track of how many copies of which software are on your systems, it is harder to ensure they are all updated and patched. Vulnerability scanners and patch management systems are helpful in this regard, but depending on how these systems operate, you could end up with periods (perhaps indefinitely long) of vulnerability.

The solution to the software tracking problem is multifaceted. It starts with an assessment of the legitimate application requirements of the organization. Perhaps some users need an expensive photo editing software suite, but its provisioning should be carefully controlled and only available to that set of users in order to minimize the licensing costs. Once the requirements are known and broken down by class of user, there are several ways to keep a handle on what software exists on which systems. Here are some of the most widely accepted best practices:

- **Application whitelisting** A whitelist is a list of software that is allowed to execute on a device or set of devices. Implementing this approach not only prevents unlicensed or unauthorized software from being installed but also protects against many classes of malware.
- **Using Gold Masters** A Gold Master is a standard image workstation or server that includes properly configured and authorized software. Organizations may have multiple images representing different sets of users. The use of Gold Masters simplifies new device provisioning and configuration, particularly if the users are not allowed to modify them.
- **Enforcing the principle of least privilege** If the typical users are not able to install any software on their devices, then it becomes a lot harder for rogue applications to show up in our networks. Furthermore, if we apply this approach, we mitigate risks from a very large set of attacks.

- **Device management software** Unified endpoint management (UEM) systems allow you to fully and remotely manage most devices, including smartphones, tablets, laptops, printers, and even Internet of Things (IoT) devices.
- **Automated scanning** Every device on your network should be periodically scanned to ensure it is running only approved software with proper configurations. Deviations from this policy should be logged and investigated by the IT or security team.

Licensing Issues

Companies have the ethical obligation to use only legitimately purchased software applications. Software makers and their industry representation groups such as The Software Alliance (BSA) use aggressive tactics to target companies that use pirated (illegal) copies of software.

Companies are responsible for ensuring that software in the corporate environment is not pirated and that the licenses (that is, license counts) are being abided by. An operations or configuration management department is often where this capability is located in a company. Automated asset management systems, or more general system management systems, may be able to report on the software installed throughout an environment, including a count of installations of each. These counts should be compared regularly (perhaps quarterly) against the inventory of licensed applications and counts of licenses purchased for each application. Applications that are found in the environment and for which no license is known to have been purchased by the company, or applications found in excess of the number of licenses known to have been purchased, should be investigated.

When applications are found in the environment for which the authorized change control and supply chain processes were not followed, they need to be brought under control, and the business area that acquired the application outside of the approved processes must be educated as to the legal and information security risks their actions may pose to the company. Many times, the business unit manager would need to sign a document indicating he understands this risk and is personally accepting it.

An application for which no valid business need can be found should be removed, and the person who installed the application should be educated and warned that future such actions may result in more severe consequences—like termination. This may sound extreme, but installing pirated software is not only an ethical violation but also both a liability risk and a potential vector for introducing malware. Organizations that use or tolerate unlicensed products are sometimes turned in by disgruntled employees as an act of revenge.

Companies should have an acceptable use policy (AUP) that indicates what software users can install and informs users that the environment will be surveyed from time to time to verify compliance. Technical controls should be emplaced to prevent unauthorized users from being able to install unauthorized software in the environment.

A fundamental best practice in software asset management is to prevent users from installing software and requiring them to submit a request for a system administrator to do so instead. This allows the administrator to ensure the software is properly licensed and added to the appropriate management systems. It also enables effective configuration management across the enterprise.

Controlling the existing hardware and software on our networks should be a precondition to provisioning new services and capabilities. To do otherwise risks making an already untenable position even worse.

Secure Provisioning

The term "provisioning" is overloaded in the technology world, which is to say that it means different actions to different people. To a telecommunications service provider, it could mean the process of running wires, installing customer premises equipment, configuring services, and setting up accounts to provide a given service (e.g., DSL). To an IT department, it could mean the acquisition, configuration, and deployment of an information system (e.g., a new server) within a broader enterprise environment. Finally, to a cloud services provider, provisioning could mean automatically spinning up a new instance of that physical server that the IT department delivered to us.

For the purpose of the CISSP exam, *provisioning* is the set of all activities required to provide one or more new information services to a user or group of users ("new" meaning previously not available to that user or group). Though this definition is admittedly broad, it does subsume all that the overloaded term means. As you will see in the following sections, the specific actions included in various types of provisioning vary significantly, while remaining squarely within our given definition.

At the heart of provisioning is the imperative to provide these information services in a secure manner. In other words, we must ensure that both the services and the devices on which they rely are secure. We already discussed supply chain risks in asset acquisition in Chapter 2. So, assuming you have a trusted supply chain, you would want to start with a Gold Master image applied to your devices as soon as you receive them. Ideally, you would then configure them according to the needs defined in the business and adapted to whatever classes of user they will support. Finally, you scan for vulnerabilities (just to be sure) and deploy it on the network. Easy, right?

Well, it gets a bit trickier when you deal with remote employees, which for many organizations are an increasing portion of their workforce. Some of the added concerns to consider are listed here:

- Securely shipping the devices to users
- Securely sending credentials to users
- Requirements for virtual private network (VPN) connectivity
- Remote monitoring of whether or not the device is on the VPN
- Making remote configuration changes
- Multifactor authentication while the device is disconnected

Obviously, the list of issues will very much depend on your particular situation. You may not have any remote users but perhaps you have a data center or hosting provider who owns the physical environment in which your assets reside. That presents its own set of concerns you need to think through in terms of secure provisioning. Finally, and perhaps inescapably, many of us have to consider unique issues when dealing with cloud assets.

Provisioning Cloud Assets

Generally, cloud provisioning is the set of all activities required to provide one or more new cloud assets to a user or group of users. So what exactly are these cloud assets? As we will see in Chapter 7, cloud computing is generally divided into three types of service: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). The provisioning of each type of service presents its own set of issues.

When we are dealing with provisioning IaaS assets, our user population is limited to the IT department. To see why this is true, we need only consider a noncloud (that is, physical) equivalent: provisioning a new server or router. Because these assets typically impact a large number of users in the organization, we must be very careful in planning and testing their provisioning. Accordingly, these provisioning actions often require the approval of the senior leadership or of the change control committee. Only a very small group of IT personnel should be able to perform such provisioning.

PaaS is similar to IaaS in terms of organizational impact, but oftentimes has a more limited scope. A platform, in this context, is typically a service such as a web or database management service. Though the IT team typically handles the provisioning, in some cases someone else in the organization may handle it. Consider, for example, the case of a development (intranet-only) web service that is being provisioned to test a web application that a team of coders is developing. Depending on the scope, context, and accessibility, this provisioning could be delegated to any one of the developers, though someone in IT would first constrain the platform to ensure it is accessible only to that team.

Finally, SaaS could be provisioned by a larger pool of users within the constraints established by the IT team in accordance with the organizational policy. If a given group of users is authorized to use the customer relationship management (CRM) system, then those users should be able to log into their accounts and self-provision that and any other applications to which they are authorized.

As you can see, the provisioning of cloud assets should be increasingly more controlled depending on the organizational impact and the risk profile of the specific asset. The key to secure provisioning is carefully setting up the cloud computing environment so that properly configured applications, platforms, and infrastructure are rapidly available to authorized users when and where they need them. After all, one of the benefits of cloud computing is the promise of self-service provisioning in near real time.

Asset Retention

Assets typically remain in use until they are no longer required, they become obsolete, or their O&M costs exceed their value to the organization. If they are no longer required, they may still be retained for some time in anticipation of future needs or perhaps for emergency use. Asset retention should be a deliberate decision that is documented and periodically revisited. Ideally, this is done as part of the change management process to ensure the retained (and no longer in use) assets don't pose undue risks.

Suppose your organization has a policy of refreshing laptops for its workforce every three years. After the latest refresh, you end up with a dozen laptops that are no longer required. Someone suggests you keep them around in case of an emergency, so you do. A couple of refresh cycles later, you end up with dozens of laptops (some of them potentially unable to run modern software) clogging up your storage spaces. This is a problem for at least four reasons. Firstly, you've run out of storage space. Secondly, there is a risk of theft since nobody is paying much attention to the laptops in the closet. Thirdly, they may no longer work when that emergency finally happens and you decide to pull them out and use them. Finally, and perhaps most seriously, unless they were properly decommissioned, they could have sensitive data in their disk drives that nobody is aware of.

Your asset retention decision-making should consider the fact that your asset life cycle may differ from its manufacturer's intended one. Original equipment manufacturers (OEMs) sell a particular product only for a specific period of time, typically one to three years. After that, they'll move on to the next version or may stop making it altogether. Either way, the product is no longer sold. OEMs will, however, continue to support their product after this point for some time, usually another three to six years. Replacement parts may still be sold and customer support resources will remain available to registered owners. *End-of-life (EOL)* for an asset is that point in time when its OEM is neither manufacturing nor sustaining it. In other words, you can't send it in for repairs, buy spare parts, or get technical assistance from the OEM. The risk in using assets after their announced EOL is that hardware failures will be much more difficult to address at reasonable costs.

There is a related term, *end-of-support (EOS)*, which is sometimes also called end-of-service-life (EOSL), that means that the manufacturer is no longer patching bugs or vulnerabilities on the product. Typically, manufacturers will continue issuing patches after a product reaches EOL for another few years. Sometimes, however, EOL and EOS coincide. Either way, we face significant risk after the product reaches EOS because whatever vulnerabilities are discovered will remain unpatched, meaning the asset is much more likely to be exploited.

Whether the business needs change or the asset reaches EOL or EOS, eventually it's time to retire it, which may drive a new business case. Before throwing an asset in the recycling bin, however, we need to properly decommission it.

Decommissioning Assets

Once an asset has reached the end of its useful life in your organization, it's important to follow a thorough process to decommission it. *Decommissioning* is the set of all activities required to permanently remove an existing asset from an operational environment. In a way, it is the opposite of provisioning.

The specific tasks required to decommission assets vary greatly depending on what the asset is. However, there are some overarching thoughts to consider before pulling the proverbial plug. These include the following:

- *Decommission only within the change management process.* The only way to minimize the risk of unintended (adverse) consequences when you pull the plug is to ensure that everyone who may have a stake in the asset is part of the decision.

- *Ensure that the asset is no longer in use.* It may seem obvious, but there may be unknown users (or uses) of the asset that were never properly documented. You'd hate to pull the plug, only to find out you killed a critical business process.
- *Review the impact on data retention.* We'll discuss data retention later in this chapter, but you have to ensure that there isn't any data in the asset (and only in that asset) that needs to be preserved.
- *Securely wipe any data on the asset.* It seems like just about every asset has the potential to hold sensitive data in nonvolatile memory or disk. Be sure you understand the persistent data storage capabilities in the asset, and you wipe them.
- *Safely dispose of the hardware.* Many assets have hazardous components such as lithium batteries that require special handling. Don't just toss that old computer into the dumpster before checking for environmental or safety hazards first.

Data Life Cycle

The data life cycle differs from the asset life cycle in some important ways. First, it usually doesn't cost anything to acquire most of the data our organizations use. Sure, there are notable exceptions, but, overall, we don't really have to demonstrate the ROI or get the chief financial officer (CFO) to agree that we need to know what each customer buys on an e-commerce site. (Actually, a CFO should be justifiably worried if that data is *not* being collected.) Another significant difference is that we can share our data with as many others as we'd like without losing it. Finally, data tends to be archived rather than disposed of when it is no longer immediately useful. Sure, we can put a workstation in a storage room in case we need it later, but this is the exception rather than the norm when dealing with tangible assets.

There are a number of data life-cycle models out there. The one we will use for our discussion is fairly simple but still effective when considering the changing nature of data and the security implications of those dynamics. At a macro level, we can divide the life of our data into six phases: acquisition, storage, use, sharing, archival, and destruction, as shown in Figure 5-2.

Data Acquisition

Generally speaking, data is acquired by an organization in one of three ways: collected directly, copied from elsewhere, or created from scratch. Collection is possible when an organization has sensors in an environment of interest. For example, an e-commerce site has a web server that can *collect* the IP address of visitors and what page referred them to the site. The application server can further collect the identity of each customer, which products they explored, and what they eventually bought. All this data can be enhanced by buying customer data from ad agencies and having it *copied* into a local data store. Finally, the marketing department can analyze all that data and *create* reports and forecasts.

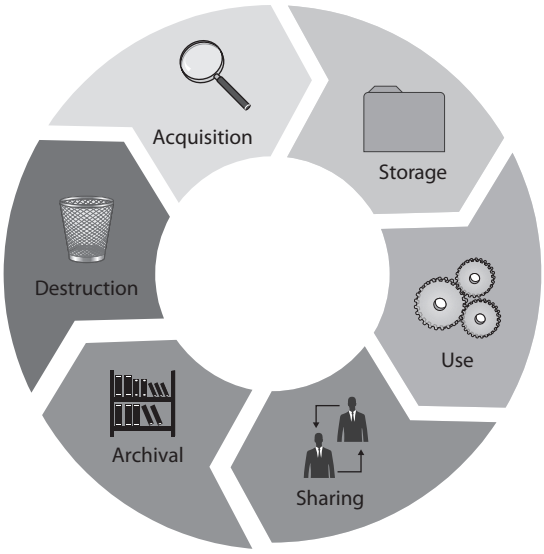


Figure 5-2
The data life cycle

Data Collection

We must ensure that the data we collect, particularly when it is personal in nature, is necessary for our jobs. Generally speaking, organizations should collect the least amount of private personal data required for the performance of their business functions. In many cases, this is not a matter of choice but of law. As of 2020, over 128 countries have enacted privacy protection laws that affect organizations within their jurisdictions. It is important to note that privacy protections vary widely among countries. The European Union is one of the most restrictive regions with respect to privacy, while China effectively has no restrictions, and therefore no real privacy protections. The United States has very few restrictions on the collection of private data by nongovernmental organizations at the national level, but has states such as California with protections similar to those of the EU. The point is that you have to be aware of the specific privacy laws that pertain to the places in which your organization stores or uses its data. This is particularly important when you outsource services (which may require access to your data) to third parties in a different country.

Apart from applicable laws and regulations, the types of personal data that your organization collects, as well as its life-cycle considerations, must be a matter of explicit written policy. Your privacy policy needs to cover your organization's collection, use, disclosure, and protection of employee and client data. Many organizations break their privacy policy into two documents: an internal document that covers employee data, and an external document that covers customer information. At a minimum, you want to answer the following questions when writing your policy:

- What personal data is collected (e.g., name, website visits, e-mail messages, etc.)?
- Why do we collect this data and how do we use it (e.g., to provide a service, for security)?

- With whom do we share this data (e.g., third-party providers, law enforcement agencies)?
- Who owns the collected data (e.g., subject, organization)?
- What rights does the subject of this data have with regard to it (e.g., opt out, restrictions)?
- When do we destroy the data (e.g., after five years, never)?
- What specific laws or regulations exist that pertain to this data (e.g., HIPAA, GDPR)?

Data Storage

After data is acquired, but before it can be used, it must be stored somewhere. There are also other steps we must take to make the information useful. Typically, we attach both system metadata (e.g., author, date/time of creation, and permissions) and business process metadata (e.g., classification, project, and owner) to it. Finally, the data is indexed to facilitate searching and assigned to one or more data stores. In smaller organizations, much of this process is invisible to the user. All that person knows is that when they create a contact in the CRM system, an order in the purchasing system, or a ticket in the workflow system, the entry is magically available to everyone in the organization who needs to access the information. In larger organizations, the process needs to be carefully architected.

Finally, there are policy controls that we have to apply. For instance, we have to encrypt credit card numbers and certain other personally identifiable information (PII) wherever

Where in the World Is My Data?

Data location can be a particularly important issue, especially when dealing with personal, healthcare, or national security data. As we discussed in Chapter 3, some countries have *data localization* laws that require certain types of data to be stored and processed in that country (examples include China and Russia). Other countries have enacted *data sovereignty* laws that stipulate that anyone who stores or processes certain types of data (typically personal data on their citizens), whether or not they do so locally, must comply with those countries' laws. Meeting these requirements can be impossible without data classification. It can also be either enabled or hindered by cloud services. Used properly, cloud service providers can help ensure data localization requirements are met by restricting certain classifications of data to a region or even a specific country. If, on the other hand, data location is not considered when architecting a cloud solution, it is very likely that sensitive data will end up in some random location at some point, potentially causing no shortage of headaches (and perhaps legal and financial liability) to its owners.

we store them. We also have to implement strict controls on who gets to access sensitive information. Additionally, we may have to provide some sort of rollback capability to revert data to a previous state, particularly if users or processes may be able to corrupt it. These and many other important considerations must be deliberately addressed as we store the data and not as an afterthought.

Data Retention

There is no universal agreement on how long an organization should retain data. Legal and regulatory requirements (where they exist) vary among countries and business sectors. What is universal is the need to ensure your organization has and follows a documented data retention policy. Doing otherwise is flirting with disaster, particularly when dealing with pending or ongoing litigation. It is not enough, of course, to simply have a policy; you must ensure it is being followed, and you must document this through regular audits.

NOTE When outsourcing data storage, it is important to specify in the contract language how long the storage provider will retain your data after you stop doing business with them and what process they will use to eradicate your data from their systems.

A very straightforward and perhaps tempting approach would be to look at the lengthiest legal or regulatory retention requirement imposed on your organization and then apply that timeframe to all your data retention. The problem with this approach is that it will probably make your retained data set orders of magnitude greater than it needs to be. Not only does this impose additional storage costs, but it also makes it more difficult to comply with electronic discovery (e-discovery) orders. When you receive an e-discovery order from a court, you are typically required to produce a specific amount of data (usually pretty large) within a given timeframe (usually very short). Obviously, the more data you retain, the more difficult and expensive this process will be.

A better approach is to segregate the specific data sets that have mandated retention requirements and handle those accordingly. Everything else should have a retention period that minimally satisfies the business requirements. Commonly, different business units within medium and large organizations have different retention requirements. For instance, a company may want to keep data from its research and development (R&D) division for a much longer period than it keeps data from its customer service division. R&D projects that are not particularly helpful today may be so at a later date, but audio recordings of customer service calls probably don't have to hang around for several years.

NOTE Be sure to get buy-in from your legal counsel when developing or modifying data retention and privacy policies.

Developing a Retention Policy

At its core, every data retention policy answers three fundamental questions:

- What data do we keep?
- How long do we keep this data?
- Where do we keep this data?

Most security professionals understand the first two questions. After all, many of us are used to keeping tax records for three years in case we get audited. The "what" and the "how long" are easy. The last question, however, surprises more than a few of us. The twist is that the question is not so much about the location per se, but rather the manner in which the data is kept at that location. In order to be useful to us, retained data must be easy to locate and retrieve.

Think about it this way. Suppose your organization had a business transaction with Acme Corporation in which you learned that Acme was involved in the sale of a particular service to a client in another country. Two years later, you receive a third-party subpoena asking for any data you may have regarding that sale. You know you retain all your data for three years, but you have no idea where the relevant data may be. Was it an e-mail, a recording of a phone conversation, the minutes from a meeting, or something else? Where would you go looking for it? Alternatively, how could you make a case to the court that locating and providing the data would be too costly for your organization?

What Data We Retain There are many reasons to retain data. Among the more common ones are data analysis (to plot trends and make predictions), historical knowledge (how did we deal with this in the past?), and regulatory requirements. Again, legal counsel must be involved in this process to ensure all legal obligations are being met. Beyond these obligations, there will be specific information that is important to the business for a variety of reasons. It is also worth considering what data might be valuable in light of business arrangements, partnerships, or third-party dealings.

The decision to retain data must be deliberate, specific, and enforceable. We want to keep only the data that we consciously decide to keep, and then we want to ensure that we can enforce that retention. Importantly, there should be a way for us to ensure that data that should not be retained is promptly and properly disposed of. If this sounds painful, we need only consider the consequences of not getting this process right. Many companies have endured undue hardships because they couldn't develop, implement, and enforce a proper retention policy. Among the biggest challenges in this realm is the balance between business needs and employee or customer privacy.

How Long We Retain Once upon a time, there were two main data retention longevity approaches: the "keep nothing" camp and the "keep everything" camp. As the legal processes caught up with modern computer technology, it became clear that (except in very limited cases) these approaches were not acceptable. For starters, whether they

Data Retention in the Age of Big Data

The term *big data* refers to collections of data that exhibit five characteristics: volume, velocity, variety, veracity, and value. Volume refers to the sheer size of the data collection, which exceeds what can reasonably be stored in traditional systems like a regular data server or a conventional database management system. Velocity describes the high speed with which new data is added, while variety means that the data is not all in the same format or even concerning the same things. Because the data comes from a multitude of sources, its veracity is difficult to establish, but we oftentimes deal with this by looking for trends and clusters rather than individual data points. Finally, there is an expectation that all this data adds value to our organizations, which justifies the costs of storing and processing it in the first place.

This last point is the crux of data retention in the age of big data: just because we *can* keep every data point from every business unit and occasionally get valuable insights is not sufficient reason to keep the data. It is far easier (and way more cost effective) to develop a retention policy that allows us to build big data stores as needed, but does so in a way that balances risks, costs, and value. Are there privacy or confidentiality issues concerning any of the data? Could any data create a legal liability for the organization? Is any of the data likely to be subject to e-discovery? If so, how difficult would it be to comply with an e-discovery order?

Apart from any legal or regulatory concerns, there's also the practical one of deciding what data is useful and what is just taking up storage space. Even if the price tag of storage doesn't seem excessive now, left unchecked, we can get there quicker than expected if we keep pumping data in. And when we get there, how would we go about removing the data we no longer want or need?

This all underscores the importance of being deliberate about building our big data stores and having policies and procedures that support valid organizational requirements, while mitigating risks at a reasonable cost.

retained nothing or everything, organizations following one of these extreme approaches found out it was difficult to defend themselves in lawsuits. The first group had nothing with which to show due diligence, for instance, while those in the second group had too much information that plaintiffs could use against them. So what is the right data retention policy? Ask your legal counsel. Seriously.

There are myriads of statutory and regulatory retention requirements, which vary from jurisdiction to jurisdiction (sometimes even within the same country). There are also best practices and case law to consider, so we won't attempt to get too specific here. Still, Table 5-2 provides some general guidelines sufficient to start the conversation with your attorneys.

Type of Data	General Period of Retention
Business documents (e.g., meeting minutes)	7 years
Invoices	5 years
Accounts payable and receivable	7 years
Human resource files	7 years (for employees who leave) or 3 years (for candidates who were not hired)
Tax records	3 years after taxes were paid
Legal correspondence	Permanently

Table 5-2 Typical Retention Periods for Different Types of Data

How We Retain Data In order for retained data to be useful, it must be accessible in a timely manner. It really does us no good to have data that takes an inordinate (and perhaps prohibitive) amount of effort to query. To ensure this accessibility, we need to consider various issues, including the ones listed here.

- **Taxonomy** A taxonomy is a scheme for classifying data. This classification can be made using a variety of categories, including functional (e.g., human resources, product development), chronological (e.g., 2020), organizational (e.g., executives, union employees), or any combination of these or other categories.
- **Classification** The sensitivity classification of the data determines the controls we place on it both while it is in use and when it gets archived. This is particularly important because many organizations protect sensitive information while in use, but not so much after it goes into the archives.
- **Normalization** Retained data comes in a variety of formats, including word processing documents, database records, flat files, images, PDF files, video, and so on. Simply storing the data in its original format is not sufficient in any but the most trivial cases. Instead, we need to develop tagging schemas that make the data searchable.
- **Indexing** Retained data must be searchable if we are to quickly pull out specific items of interest. The most common approach to making data searchable is to build indexes for it. Many archiving systems implement this feature, but others do not. Either way, the indexing approach must support the likely future queries on the archived data.

Ideally, archiving occurs in a centralized, regimented, and homogenous manner. We all know, however, that this is seldom the case. We may have to compromise in order to arrive at solutions that meet our minimum requirements within our resource constraints. Still, as we plan and execute our retention strategies, we must remain focused on how we will efficiently access archived data many months or years later.

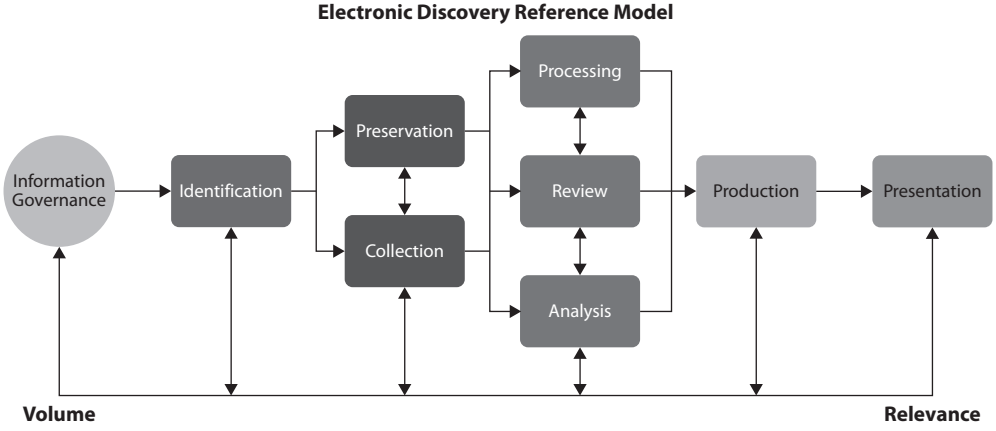
E-Discovery

Discovery of electronically stored information (ESI), or *e-discovery*, is the process of producing for a court or external attorney all ESI pertinent to a legal proceeding. For example, if your company is being sued for damages resulting from a faulty product,

the plaintiff's attorney could get an e-discovery order compelling you to produce all e-mail between the QA team and senior executives in which the product's faults are discussed. If your data retention policy and procedures are adequate, e-discovery should not require excessive efforts. If, on the other hand, you have been slack about retention, such an order could cripple the organization.

The Electronic Discovery Reference Model (EDRM) identifies eight steps, though they are not necessarily all required, nor are they performed in a linear manner:

1. **Identification** of data required under the order.
2. **Preservation** of this data to ensure it is not accidentally or routinely destroyed while complying with the order.
3. **Collection** of the data from the various stores in which it may be.
4. **Processing** to ensure the correct format is used for both the data and its metadata.
5. **Review** of the data to ensure it is relevant.
6. **Analysis** of the data for proper context.
7. **Production** of the final data set to those requesting it.
8. **Presentation** of the data to external audiences to prove or disprove a claim.



(Source: EDRM; www.edrm.net)

Data Use

After data is acquired and stored, it will spend much of its time being used. That is to say it will be read and modified by a variety of users with the necessary access level. From a security perspective, this stage in the data life cycle presents the most challenges in terms of ensuring confidentiality, integrity, and availability. You want the information available, but only to the right people who should then be able to modify it in authorized ways.

Consistency is also an issue with regard to policy and regulatory compliance. As the information is used and aggregated, it may trigger requirements that must be automatically enforced. For example, a document that refers to a project using a code word or name

may be unclassified and freely available, but if that word/name is used in conjunction with other details (a place, purpose, or team members' names), then it would make the entire document classified. Changes in the information as it is in use must be mapped to the appropriate internal policies, and perhaps to regulations or laws.

Data Maintenance

As data is being used, we have to ensure that it remains accurate and internally consistent. Suppose that Sally is a salesperson in our organization. She meets a prospective customer named Charlie and enters his contact information and other details into a CRM system. E-mails are exchanged, meetings are scheduled, and documents are filed with Charlie's data. One day, Charlie gets a promotion and moves to corporate headquarters. Just like that, his title, phone number, and address all change. How do we ensure that we update this data and that we do it across the entire organization? Sure, the CRM piece is easy, but what about the myriad of other places in which the now obsolete data exists? We need to have a plan for maintaining the accuracy of data that is being used and may be critical to our business processes.

We must also consider what happens when the data is incorrect when it is first acquired. There was a recent story in the news about a police clerk who incorrectly entered the personal information of a convicted murderer who had just been transferred to his station. The information was actually that of an innocent citizen who had, earlier that day, applied for a permit. The erroneous information was shared across the country with local, national, and even private organizations. By the time the error was discovered, there was no way to globally correct the entry. To this day, that innocent man is periodically denied employment or services because some system shows that he is a convicted murderer. For most of our organizations, this scenario would likely result in hefty fines or a major lawsuit unless we had an effective way to maintain our data.

Another case for data maintenance deals with corruption and inconsistencies. For instance, if we have multiple data stores for performance or reliability purposes, we must ensure that modifications to the data are replicated. We also need to have mechanisms for automatically resolving inconsistencies, such as those that would occur from a server having a power outage after data has been modified but before it has been replicated. This is particularly important in very dynamic systems that have rollback capabilities.

Data Sharing

Gone are the days when any of us could accomplish anything significant solely on our own. Virtually every organization in the world, particularly those with information systems, is part of a supply chain. Information sharing is a key enabler of modern supply chains. Without it, we wouldn't be able to log into our systems (especially if you have a third-party identity management service like Google or Facebook), send or receive e-mail, or sell widgets online (it's hard to sell something without sharing payment card information with a payment processor).

While we all have some data sharing requirements imposed by our IT infrastructure, we also willingly share data with others for specific business reasons. For example, an e-commerce site will almost certainly partner with a digital advertising firm to drum up

business and with a logistics company to deliver tangible goods. It may also partner with other companies that offer complementary goods or services and collect referral fees from each other. There are many other reasons to share data, but the important concept here is that this sharing needs to be deliberate. If you share the wrong data, or do so in the wrong way, you could lose competitive advantage or even break the law.

To avoid data sharing nightmares, be sure to involve all the necessary staff (business, IT, security, legal) in the conversation early. Discuss the business need to share data and restrict that data to the minimum essential to satisfy that need. Document the agreement in a legally binding contract that's been approved by your legal counsel. This agreement needs to specify the obligations of each party with regard to the entire shared data life cycle. For example, what data will be shared, how it will be stored and used by each party, with whom it may be shared, how it will be archived and for how long, and, finally, when and how it will be destroyed.

Data Archival

The data in our systems will likely stop being used regularly (or at all) at some point. When this happens, but before we get rid of it, we probably want to retain it for a variety of reasons. Maybe we anticipate that it will again be useful at a later time, or maybe we are required to keep it around for a certain period of time, as is the case with certain financial information. Whatever the reason for moving this data off to the side, the fact that it is no longer regularly used could mean that unauthorized or accidental access and changes to it could go undetected for a long time if we don't implement appropriate controls. Of course, the same lack of use could make it easier to detect this threat if we do have the right controls.

Another driver for retention is the need for backups. Whether we're talking about user or back-end backups, it is important to consider our risk assessment when deciding which backups are protected and how. To the extent that end-user backups are performed to removable disk drives, it is difficult to imagine a scenario in which these backups should not be encrypted. Every major operating system provides a means to perform automatic backups as well as encrypt those backups. Let's take advantage of this.

This all leads us to the question of how long we need to retain data. If we discard it too soon, we risk not being able to recover from a failure or an attack. We also risk not being able to comply with e-discovery requests or subpoenas. If we keep the data for too long,

Backup vs. Archive

The terms backup and archive are sometimes used interchangeably. In reality, they have different meanings that are best illustrated using the life-cycle model described in this section. A data *backup* is a copy of a data set currently in use that is made for the purpose of recovering from the loss of the original data. Backup data normally becomes less useful as it gets older.

A data *archive* is a copy of a data set that is no longer in use, but is kept in case it is needed at some future point. When data is archived, it is usually removed from its original location so that the storage space is available for data in use.

we risk excessive costs as well as increased liabilities. The answer, once again, is that this is all part of our risk management process and needs to be codified in policies.

Data Destruction

Sooner or later, every organization will have to dispose of data. This usually, but not always, means data destruction. Old mailboxes, former employee records, and past financial transactions are all examples of data sets that must, at some point, be destroyed. When this time comes, there are two important issues to consider: that the data does in fact get destroyed, and that it is destroyed correctly. When we discuss roles and responsibilities later in this chapter, we'll see who is responsible for ensuring that both of these issues are taken care of.

A twist on the data destruction issue is when we need to transfer the data to another party and then destroy it on our data stores. For instance, organizations hosting services for their clients typically have to deal with requests to do a bulk export of their data when they migrate to another provider. Companies sometimes sell accounts (e.g., home mortgages) to each other, in which case the data is transferred and eventually (after the mandatory retention period) destroyed on the original company's systems.

No matter the reason, we have to ensure that the data is properly destroyed. How this is done is, again, tied to our risk management. The bottom line is that the data must be rendered sufficiently difficult for an adversary to recover so that the risk of such recovery is acceptable to our organization. This is not hard to do when we are dealing with physical devices such as hard disk drives that can be wiped, degaussed, or shredded (or all of these in particularly risk-adverse organizations such as certain government entities). Data destruction can be a bit more complicated when we deal with individual files (or parts thereof) or database records (such as many e-mail systems use for mailbox storage). Further complicating matters, it is very common for multiple copies of each data item to exist across our information systems. How can you ensure that all versions are gone? The point is that the technical details of how and where the data is stored are critical to ensuring its proper destruction.

Data Remanence

Even when policies exist (and are enforced and audited) to ensure the protection of privacy, it is possible for technical issues to threaten this privacy. It is a well-known fact that most data deletion operations do not, in fact, erase anything; normally, they simply mark the memory as available for other data, without wiping (or even erasing) the original data. This is true not only of file systems but also of databases. Since it is difficult to imagine a data store that would not fit in either of these two constructs, it should be clear that simply "deleting" data will likely result in data remanence issues.

NOTE NIST Special Publication 800-88, Revision 1, *Guidelines for Media Sanitization* (December 2014), describes the best practices for combating data remanence.

Let's consider what happens when we create a text file using the File Allocation Table (FAT) file system. Though this original form of FAT is antiquated, its core constructs

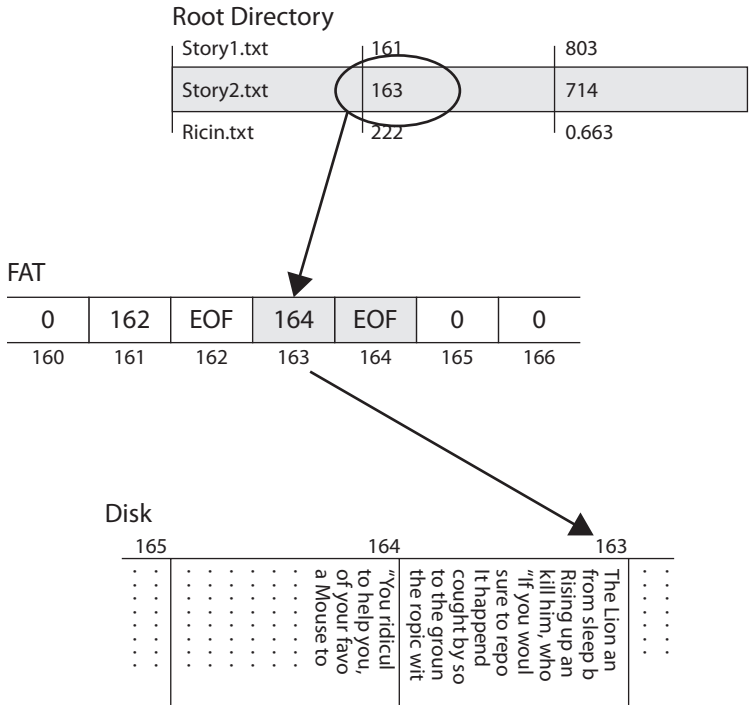


Figure 5-3
Writing a text file to disk

(e.g., disk blocks, free block list/table, file metadata table) are also found at the heart of all other modern file systems. Its simplicity makes it a wonderful training tool for the purpose of explaining file creation and deletion.

Suppose we type up the famous Aesop fable titled "The Lion and the Mouse" in a text editor and save it to disk. The operating system will ask us for a filename, which will be Story2.txt for this example. The system will then check the File Allocation Table for available blocks on which to store the text file. As shown in Figure 5-3, the system creates a directory entry for the file containing the name (Story2.txt), location of the first block (163), and the file size in bytes (714). In our simplistic example, each block is 512 bytes in size, so we'll need two of them. Fortunately, block 164 is right next to the start block and is also free. The system will use the entry for block 163 (the first block of the file) to point to the next block containing it (164). This allows files to occupy discontinuous blocks if the disk is heavily fragmented. That chain of blocks could be quite long if the file was big enough and we didn't run out of disk space first. In our simple example, however, we just need two blocks, so block 164 is the final one in use and gets a special label of EOF to denote the end of the file.

Suppose we decide to delete the file. Instead of cleaning up the table, the FAT file system will simply replace the first character of the filename in the directory table with a reserved character (shown in Figure 5-4 as a question mark) to indicate that the file was deleted. The starting block will be preserved in the directory, but the corresponding entries in the File Allocation Table are zeroed out to show that those blocks are available

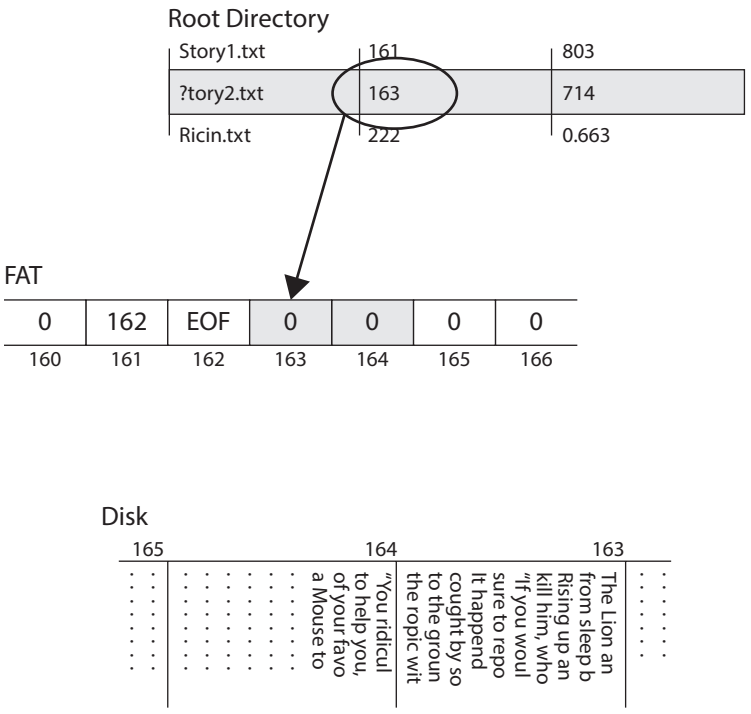


Figure 5-4
Deleting a file

for other files. As you can see in Figure 5-4, the contents of the file on the disk remain intact. This is why data remanence is such a big problem: because file systems almost never securely wipe data when deleting files.

At some point, however, users will create new files and save them to disk, which could result in our original data being partly or completely overwritten. This is shown in Figure 5-5. In this case, the new file requires only one block of disk space because it only contains the text "Hello World!" Suppose the user calls this file "hello.txt" and the system stores it in block 163, which used to be the start block for the previous Story2.txt file. That block will be overwritten with the new file's content and almost certainly padded with empty characters to fill out the block. The next block, however, contains the remainder of the deleted file, so partial contents are still available to anyone with the right recovery tools. Note also that the original file's metadata is preserved in the directory table until that block is needed for another file.

This example, though simplistic, illustrates the process used by almost every file system when creating and deleting files. The data structures may be named differently in modern versions of Windows, Linux, and macOS, but their purpose and behavior remain essentially the same. In fact, many databases use a similar approach to "deleting" entries by simply marking them as deleted without wiping the original data.

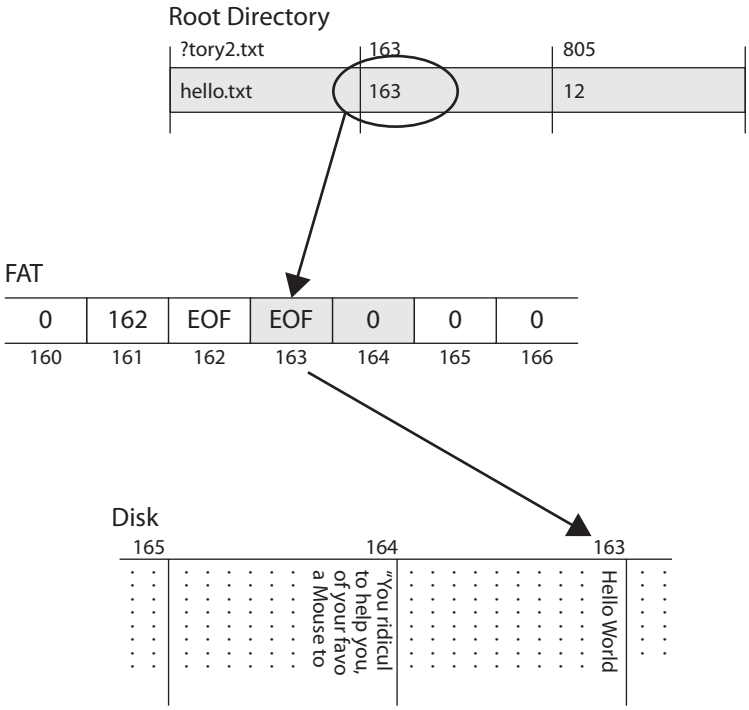


Figure 5-5 Partially overwriting a file

To counter data remanence, it is important to identify procedures for ensuring that private data is properly removed. Generally speaking, there are four approaches to eliminating data remanence:

- **Overwriting** Overwriting data entails replacing the 1's and 0's that represent it on storage media with random or fixed patterns of 1's and 0's in order to render the original data unrecoverable. This should be done at least once (e.g., overwriting the medium with 1's, 0's, or a pattern of these), but may have to be done more than that. For many years the U.S. Department of Defense (DoD) standard 5220.22-M required that media be overwritten seven times. This standard has since been superseded. DoD systems with sensitive information must now be degaussed.
- **Degaussing** This is the process of removing or reducing the magnetic field patterns on conventional disk drives or tapes. In essence, a powerful magnetic force is applied to the media, which results in the wiping of the data and sometimes the destruction of the motors that drive the platters. While it may still be possible to recover the data, it is typically cost prohibitive to do so.
- **Encryption** Many mobile devices take this approach to quickly and securely render data unusable. The premise is that the data is stored on the medium in encrypted format using a strong key. To render the data unrecoverable, the system simply needs to securely delete the encryption key, which is many times faster than deleting the encrypted data. Recovering the data in this scenario is typically computationally infeasible.

- **Physical destruction** Perhaps the best way to combat data remanence is to simply destroy the physical media. The two most commonly used approaches to destroying media are to shred it or expose it to caustic or corrosive chemicals that render it unusable. Another approach is incineration.

Data Roles

The data life cycle and, just as importantly, its protection, is driven by responsible and accountable individuals within each organization. We've already seen how data breaches can wreak havoc on otherwise successful companies and even drive them (or their key leaders) out of business. While this is not an exhaustive list, the following sections describe some of the key responsibilities by role when it comes to protecting data.

Data Controllers

Data controllers decide why and how different types of data will be processed. These are the senior managers that set policies with regard to the management of the data life cycle, particularly with regard to sensitive data such as personal data. Once these controllers set the policy, it is up to the rest of the organization to abide by it.

Data Owners

Data owners are responsible for the life cycle management of a set of data. Among the responsibilities of the data owners are data classification and the approval of disclosure requests. The data owners, therefore, indirectly or directly decide who gets access to specific data. This is particularly important given that these individuals typically are senior managers within the organization. In reality, the majority of these decisions should be codified in formal written policies. Any exceptions to policy should be just that—exceptions—and must be properly documented.

Data Custodians

It is good and well to have policies addressing the life cycle of your data, but someone needs to implement them at the technical level. These individuals are the data custodians, who are responsible for controlling access to the data, implementing the required security controls, and ensuring that both the data and manner in which it is used can be audited. Data custodians also participate in the change management process for all matters pertaining to the data life cycle.

Data Processors

The group of users best positioned to protect (or compromise) data consists of those who deal with that data on a routine basis: *data processors*. These individuals can be found in a variety of places within the organization depending on what particular data is of concern. The critical issue here is that these individuals understand the boundaries of what acceptable behavior is and (just as importantly) know what to do when data is accidentally or intentionally handled in a manner that does not conform to applicable policies. The

best ways to address this issue are through training and auditing. On the one hand, data processors must be properly trained to handle their duties and responsibilities. On the other hand, there must be routine inspections to ensure their behavior complies with all applicable laws, regulations, and policies.

Data Subjects

All personal data concerns a real individual. The person about whom the data is concerned is the data subject. While data subjects are seldom involved in the organizational data life cycle, we all have a solemn duty to protect them and their privacy as we use their data for our own purposes. Respect for the data subjects is foundational to ensuring the protection and privacy of their data.

Chapter Review

Protecting assets, particularly information, is critical to any organization and must be incorporated into the comprehensive risk management process described in Chapter 2. This protection will probably require different controls at different phases in the data life cycle, so it is important to consider phase-specific risks when selecting controls. Rather than trying to protect all information equally, our organizations need classification standards that help us identify, handle, and protect data according to its sensitivity and criticality. We must also consider the roles played by various people in the organization. From the senior executives to the newest and most junior member of the team, everyone who interacts with our information has (and should understand) specific responsibilities with regard to protecting our assets.

A key responsibility is the protection of privacy of personal information. For various legal, regulatory, and operational reasons, we want to limit how long we hold on to personal information. There is no one-size-fits-all approach to data retention, so it is incumbent on the organization's leadership to consider a multitude of factors when developing privacy and data retention policies. These policies, in turn, should drive risk-based controls, baselines, and standards applied to the protection of our data. A key element in applying controls needs to be the proper use of strong cryptography.

Quick Review

- Data goes through a life cycle that starts with its acquisition and ends with its disposal.
- Each phase of the data life cycle requires different considerations when assessing risks and selecting controls.
- New information is prepared for use by adding metadata, including classification labels.

- Ensuring the consistency of data must be a deliberate process in organizations that use data replication.
- Cryptography can be an effective control at all phases of the data life cycle.
- The data retention policy drives the timeframe at which data transitions from the archival phase to the disposal phase of its life cycle.
- Information classification corresponds to the information's value to the organization.
- Each classification should have separate handling requirements and procedures pertaining to how that data is accessed, used, and destroyed.
- Senior executives are ultimately responsible to the shareholders for the successes and failures of their corporations, including security issues.
- The data owner is the manager in charge of a specific business unit and is ultimately responsible for the protection and use of a specific subset of information.
- Data owners specify the classification of data, and data custodians implement and maintain controls to enforce the set classification levels.
- The data retention policy must consider legal, regulatory, and operational requirements.
- The data retention policy should address what data is to be retained, where, how, and for how long.
- Electronic discovery (e-discovery) is the process of producing for a court or external attorney all electronically stored information (ESI) pertinent to a legal proceeding.
- Normal deletion of a file does not permanently remove it from media.
- NIST SP 800-88, Revision 1, *Guidelines for Media Sanitization*, describes the best practices for combating data remanence.
- Overwriting data entails replacing the 1's and 0's that represent it on storage media with random or fixed patterns of 1's and 0's to render the original data unrecoverable.
- Degaussing is the process of removing or reducing the magnetic field patterns on conventional disk drives or tapes.
- Privacy pertains to personal information, both from your employees and your customers.
- Generally speaking, organizations should collect the least amount of private personal data required for the performance of their business functions.
- Mobile devices are easily lost or stolen and should proactively be configured to mitigate the risks of data loss or leakage.
- Paper products oftentimes contain information that deserves controls commensurate to the sensitivity and criticality of that information.

Questions

Please remember that these questions are formatted and asked in a certain way for a reason. Keep in mind that the CISSP exam is asking questions at a conceptual level. Questions may not always have the perfect answer, and the candidate is advised against always looking for the perfect answer. Instead, the candidate should look for the best answer in the list.

1. Which of the following statements is true about the data life cycle?
   - **A.** The data life cycle begins with its archival and ends with its classification.
   - **B.** Most data must be retained indefinitely.
   - **C.** The data life cycle begins with its acquisition/creation and ends with its disposal/destruction.
   - **D.** Preparing data for use does not typically involve adding metadata to it.
2. Ensuring data consistency is important for all the following reasons, *except*
   - **A.** Replicated data sets can become desynchronized.
   - **B.** Multiple data items are commonly needed to perform a transaction.
   - **C.** Data may exist in multiple locations within our information systems.
   - **D.** Multiple users could attempt to modify data simultaneously.
3. Which of the following makes the most sense for a single organization's classification levels for data?
   - **A.** Unclassified, Secret, Top Secret
   - **B.** Public, Releasable, Unclassified
   - **C.** Sensitive, Controlled unclassified information (CUI), Proprietary
   - **D.** Proprietary, Trade Secret, Private
4. Which of the following is the most important criterion in determining the classification of data?
   - **A.** The level of damage that could be caused if the data were disclosed
   - **B.** The likelihood that the data will be accidentally or maliciously disclosed
   - **C.** Regulatory requirements in jurisdictions within which the organization is not operating
   - **D.** The cost of implementing controls for the data
5. Who bears ultimate responsibility for the protection of assets within the organization?
   - **A.** Data owners
   - **B.** Cyber insurance providers
   - **C.** Senior management
   - **D.** Security professionals

6. During which phase or phases of the data life cycle can cryptography be an effective control?
   - **A.** Use
   - **B.** Archival
   - **C.** Disposal
   - **D.** All the above
7. A transition into the disposal phase of the data life cycle is most commonly triggered by
   - **A.** Senior management
   - **B.** Insufficient storage
   - **C.** Acceptable use policies
   - **D.** Data retention policies
8. Information classification is most closely related to which of the following?
   - **A.** The source of the information
   - **B.** The information's destination
   - **C.** The information's value
   - **D.** The information's age
9. The data owner is most often described by all of the following *except*
   - **A.** Manager in charge of a business unit
   - **B.** Ultimately responsible for the protection of the data
   - **C.** Financially liable for the loss of the data
   - **D.** Ultimately responsible for the use of the data
10. Who has the primary responsibility of determining the classification level for information?
    - **A.** The functional manager
    - **B.** Senior management
    - **C.** The owner
    - **D.** The user
11. If different user groups with different security access levels need to access the same information, which of the following actions should management take?
    - **A.** Decrease the security level on the information to ensure accessibility and usability of the information.
    - **B.** Require specific written approval each time an individual needs to access the information.
    - **C.** Increase the security controls on the information.
    - **D.** Decrease the classification label on the information.

12. What should management consider the most when classifying data?

- **A.** The type of employees, contractors, and customers who will be accessing the data
- **B.** Availability, integrity, and confidentiality
- **C.** Assessing the risk level and disabling countermeasures
- **D.** The access controls that will be protecting the data

13. Which of the following requirements should the data retention policy address?

- **A.** Legal
- **B.** Regulatory
- **C.** Operational
- **D.** All the above

14. Which of the following is *not* addressed by the data retention policy?

- **A.** What data to keep
- **B.** For whom data is kept
- **C.** How long data is kept
- **D.** Where data is kept

15. Which of the following best describes the mitigation of data remanence by a physical destruction process?

- **A.** Replacing the 1's and 0's that represent data on storage media with random or fixed patterns of 1's and 0's
- **B.** Converting the 1's and 0's that represent data with the output of a cryptographic function
- **C.** Removing or reducing the magnetic field patterns on conventional disk drives or tapes
- **D.** Exposing storage media to caustic or corrosive chemicals that render it unusable

16. Which of the following best describes the mitigation of data remanence by a degaussing destruction process?

- **A.** Replacing the 1's and 0's that represent data on storage media with random or fixed patterns of 1's and 0's
- **B.** Converting the 1's and 0's that represent data with the output of a cryptographic function
- **C.** Removing or reducing the magnetic field patterns on conventional disk drives or tapes
- **D.** Exposing storage media to caustic or corrosive chemicals that render it unusable

17. Which of the following best describes the mitigation of data remanence by an overwriting process?
    - **A.** Replacing the 1's and 0's that represent data on storage media with random or fixed patterns of 1's and 0's
    - **B.** Converting the 1's and 0's that represent data with the output of a cryptographic function
    - **C.** Removing or reducing the magnetic field patterns on conventional disk drives or tapes
    - **D.** Exposing storage media to caustic or corrosive chemicals that render it unusable

Answers

1. **C.** Although various data life-cycle models exist, they all begin with the creation or acquisition of the data and end with its ultimate disposal (typically destruction).
2. **B.** Although it is typically true that multiple data items are needed for a transaction, this has much less to do with the need for data consistency than do the other three options. Consistency is important because we oftentimes keep multiple copies of a given data item.
3. **A.** This is a typical set of classification levels for government and military organizations. Each of the other options has at least two terms that are synonymous or nearly synonymous.
4. **A.** There are many criteria for classifying data, but it is most important to focus on the value of the data or the potential loss from its disclosure. The likelihood of disclosure, irrelevant jurisdictions, and cost considerations should not be central to the classification process.
5. **C.** Senior management always carries the ultimate responsibility for the organization.
6. **D.** Cryptography can be an effective control at every phase in the data life cycle. During data acquisition, a cryptographic hash can certify its integrity. When sensitive data is in use or in archives, encryption can protect it from unauthorized access. Finally, encryption can be an effective means of destroying the data.
7. **D.** Data retention policies should be the primary reason for the disposal of most of our information. Senior management or lack of resources should seldom, if ever, be the reason we dispose of data, while acceptable use policies have little, if anything, to do with it.
8. **C.** Information classification is very strongly related to the information's value and/or risk. For instance, trade secrets that are the key to a business's success are highly valuable, which will lead to a higher classification level. Similarly, information that could severely damage a company's reputation presents a high level of risk and is similarly classified at a higher level.

9. **C.** The data owner is the manager in charge of a specific business unit, and is ultimately responsible for the protection and use of a specific subset of information. In most situations, this person is not financially liable for the loss of his or her data.
10. **C.** A company can have one specific data owner or different data owners who have been delegated the responsibility of protecting specific sets of data. One of the responsibilities that goes into protecting this information is properly classifying it.
11. **C.** If data is going to be available to a wide range of people, more granular security should be implemented to ensure that only the necessary people access the data and that the operations they carry out are controlled. The security implemented can come in the form of authentication and authorization technologies, encryption, and specific access control mechanisms.
12. **B.** The best answer to this question is B, because to properly classify data, the data owner must evaluate the availability, integrity, and confidentiality requirements of the data. Once this evaluation is done, it will dictate which employees, contractors, and users can access the data, which is expressed in answer A. This assessment will also help determine the controls that should be put into place.
13. **D.** The data retention policy should follow the laws of any jurisdiction within which the organization's data resides. It must similarly comply with any regulatory requirements. Finally, the policy must address the organization's operational requirements.
14. **B.** The data retention policy should address what data to keep, where to keep it, how to store it, and for how long to keep it. The policy is not concerned with "for whom" the data is kept.
15. **D.** Two of the most common approaches to destroying data physically involve shredding the storage media or exposing it to corrosive or caustic chemicals. In certain highly sensitive government organizations, these approaches are used in tandem to make the risk of data remanence negligible.
16. **C.** Degaussing is typically accomplished by exposing magnetic media (such as hard disk drives or magnetic tapes) to powerful magnetic fields in order to change the orientation of the particles that physically represent 1's and 0's.
17. **A.** Data remanence can be mitigated by overwriting every bit on the storage medium. This is normally accomplished by writing all 0's, or all 1's, or a fixed pattern of them, or a random sequence of them. Better results can be obtained by repeating the process with different patterns multiple times.

This page intentionally left blank

CHAPTER 6

Data Security

This chapter presents the following:

- Data states
- Data security controls
- Data protection methods

Data is a precious thing and will last longer than the systems themselves.

—Tim Berners-Lee

Having addressed assets in general in the previous chapter, we now turn our attention to specific ways in which we go about protecting one of our most precious assets: data. One of the facts that makes securing data so difficult is that it can seemingly flow and rest anywhere in the world, literally. Even that virtual sticky note on your home computer's desktop reminding you to pick up some milk can be backed up automatically and its contents stored almost anywhere in the world unless you take steps to control it. The same issue arises, though with more significant consequences, when we consider data in our organizations' IT systems.

Clearly, the manner in which we protect our data depends on where it is and what it is doing (or having done to it). That sticky note on your desktop has different security implications than a confidential message being transmitted between two government organizations. Part of the decision deals with the data classification we discussed in Chapter 5, but another part deals with whether the data is just sitting somewhere, moving between places, or actively being worked on. These are the data states, and they determine what security controls make sense over time.

Data Security Controls

As described in Chapter 5, which types of controls should be implemented per classification depends upon the level of protection that management and the security team have determined is needed. The numerous types of controls available are discussed throughout this book. But some considerations pertaining to sensitive data and applications are common across most organizations:

- Strict and granular access control for all levels of sensitive data and programs
- Encryption of data while stored and while in transit

- Auditing and monitoring (determine what level of auditing is required and how long logs are to be retained)
- Separation of duties (determine whether two or more people must be involved in accessing sensitive information to protect against fraudulent activities; if so, define and document procedures)
- Periodic reviews (review classification levels, and the data and programs that adhere to them, to ensure they are still in alignment with business needs; data or applications may also need to be reclassified or declassified, depending upon the situation)
- Backup and recovery procedures (define and document)
- Change control procedures (define and document)
- Physical security protection (define and document)
- Information flow channels (where does the sensitive data reside and how does it traverse the network)
- Proper disposal actions, such as shredding, degaussing, and so on (define and document)
- Marking, labeling, and handling procedures

Clearly, this is not an exhaustive list. Still, it should be a good start as you delve into whatever specific compliance requirements apply to your organization. Keep in mind that the controls that constitute adequate data protections vary greatly between jurisdictions. When it comes to compliance, always be sure to consult your legal counsel.

Data States

Which controls we choose to use to mitigate risks to our information depend not only on the value we assign to that information but also on the dynamic state of that information. Generally speaking, data exists in one of three states: at rest, in motion, or in use. These states and their interrelations are shown in Figure 6-1. The risks to each state are different in significant ways, as described next.

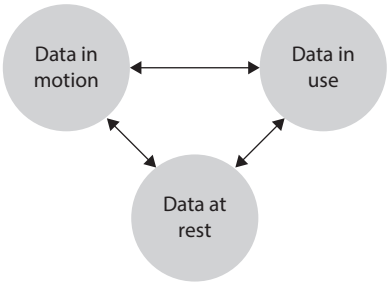


Figure 6-1
The states of data

Data at Rest

Information in an information system spends most of its time waiting to be used. The term *data at rest* refers to data that resides in external or auxiliary storage devices, such as hard disk drives (HDDs), solid-state drives (SSDs), optical discs (CD/DVD), or even on magnetic tape. A challenge with protecting data in this state is that it is vulnerable, not only to threat actors attempting to reach it over our systems and networks but also to anyone who can gain physical access to the device. It is not uncommon to hear of data breaches caused by laptops or mobile devices being stolen. In fact, one of the largest personal health information (PHI) breaches occurred in San Antonio, Texas, in September 2009 when an employee left unattended in his car backup tapes containing PHI on some 4.9 million patients. A thief broke into the vehicle and made off with the data. The solution to protecting data in such scenarios is as simple as it is ubiquitous: encryption.

Every major operating system now provides means to encrypt individual files or entire volumes in a way that is almost completely transparent to the user. Third-party software is also available to encrypt compressed files or perform whole-disk encryption. What's more, the current state of processor power means that there is no noticeable decrease in the performance of computers that use encryption to protect their data. Unfortunately, encryption is not yet the default configuration in any major operation system. The process of enabling it, however, is so simple that it borders on the trivial.

Many medium and large organizations now have policies that require certain information to be encrypted whenever it is stored in an information system. While typically this applies to PII, PHI, or other regulated information, some organizations are taking the proactive step of requiring whole-disk encryption to be used on all portable computing devices such as laptops and external hard drives. Beyond what are clearly easily pilfered devices, we should also consider computers we don't normally think of as mobile. Another major breach of PHI was reported by Sutter Health of California in 2011 when a thief broke a window and stole a desktop computer containing the unencrypted records on more than 4 million patients. We should resolve to encrypt all data being stored anywhere, and modern technology makes this easier than ever. This approach to "encrypt everywhere" reduces the risk of users accidentally storing sensitive information in unencrypted volumes.

NOTE NIST Special Publication 800-111, *Guide to Storage Encryption Technologies for End User Devices*, provides a good, if somewhat dated (2007), approach to this topic.

Data in Motion

Data in motion is data that is moving between computing nodes over a data network such as the Internet. This is perhaps the riskiest time for our data: when it leaves the confines of our protected enclaves and ventures into that Wild West that is the Internet. Fortunately, encryption once again rises to the challenge. The single best protection for our data while it is in motion (whether within or without our protected networks) is strong encryption such as that offered by Transport Layer Security (TLS version 1.2 and later)

or IPSec. We will discuss strong (and weak) encryption in Chapter 8, but for now you should be aware that TLS and IPSec support multiple cipher suites and that some of these are not as strong as others. Weaknesses typically are caused by attempts to ensure backward compatibility, but result in unnecessary (or perhaps unknown) risks.

NOTE The terms data in motion, data in transit, and data in flight are all used interchangeably.

By and large, TLS relies on digital certificates (more on those in Chapter 8) to certify the identity of one or both endpoints. Typically, the server uses a certificate but the client doesn't. This one-way authentication can be problematic because it relies on the user to detect a potential impostor. A common exploit for this vulnerability is known as a man-in-the-middle (MitM) attack. The attacker intercepts the request from the client to the server and impersonates the server, pretending to be, say, Facebook. The attacker presents to the client a fake web page that looks exactly like Facebook and requests the user's credentials. Once the user provides that information, the attacker can forward the log-in request to Facebook and then continue to relay information back and forth between the client and the server over secure connections, intercepting all traffic in the process. A savvy client would detect this by noticing that the web browser reports a problem with the server's certificate. (It is extremely difficult for all but certain nation-states to spoof a legitimate certificate.) Most users, however, simply click through any such warnings without thinking of the consequences. This tendency to ignore the warnings underscores the importance of security awareness in our overall efforts to protect our information and systems.

Another approach to protecting our data in motion is to use trusted channels between critical nodes. Virtual private networks (VPNs) are frequently used to provide secure connections between remote users and corporate resources. VPNs are also used to securely connect campuses or other nodes that are physically distant from each other. The trusted channels we thus create allow secure communications over shared or untrusted network infrastructure.

Data in Use

Data in use is the term for data residing in primary storage devices, such as volatile memory (e.g., RAM), memory caches, or CPU registers. Typically, data remains in primary storage for short periods of time while a process is using it. Note, however, that anything stored in volatile memory could persist there for extended periods (until power is shut down) in some cases. The point is that data in use is being touched by the CPU or ALU in the computer system and will eventually go back to being data at rest, or end up being deleted.

As discussed earlier, data at rest should be encrypted. The challenge is that, in most operating systems today, the data must be decrypted before it is used. In other words, data in use generally cannot be protected by encrypting it. Many people think this is safe, the thought process being, "If I'm encrypting my data at rest and in transit already,

why would I worry about protecting it during the brief period in which it is being used by the CPU? After all, if someone can get to my volatile memory, I probably have bigger problems than protecting this little bit of data, right?" Not really.

Various independent researchers have demonstrated effective side-channel attacks against memory shared by multiple processes. A *side-channel attack* exploits information that is being leaked by a cryptosystem. As we will see in our discussion of cryptology in Chapter 8, a cryptosystem can be thought of as connecting two channels: a plaintext channel and an encrypted one. A *side channel* is any information flow that is the electronic by-product of this process. As an illustration of this, imagine yourself being transported in the windowless back of a van. You have no way of knowing where you are going, but you can infer some aspects of the route by feeling the centrifugal force when the van makes a turn or follows a curve. You could also pay attention to the engine noise or the pressure in your ears as you climb or descend hills. These are all side channels. Similarly, if you are trying to recover the secret keys used to encrypt data, you could pay attention to how much power is being consumed by the CPU or how long it takes for other processes to read and write from memory. Researchers have been able to recover 2,048-bit keys from shared systems in this manner.

But the threats are not limited to cryptosystems alone. The infamous Heartbleed security bug of 2014 demonstrated how failing to check the boundaries of requests to read from memory could expose information from one process to others running on the same system. In that bug, the main issue was that anyone communicating with the server could request an arbitrarily long "heartbeat" message from it. Heartbeat messages are typically short strings that let the other end know that an endpoint is still there and wanting to communicate. The developers of the library being used for this never imagined that someone would ask for a string that was hundreds of characters in length. The attackers, however, did think of this and in fact were able to access crypto keys and other sensitive data belonging to other users.

More recently, the Meltdown, Spectre, and BranchScope attacks that came to light in 2018 show how a clever attacker can exploit hardware features in most modern CPUs. Meltdown, which affects Intel and ARM microprocessors, works by exploiting the manner in which memory mapping occurs. Since cache memory is a lot faster than main memory, most modern CPUs include ways to keep frequently used data in the faster cache. Spectre and BranchScope, on the other hand, take advantage of a feature called speculative execution, which is meant to improve the performance of a process by guessing what future instructions will be based on data available in the present. All three implement side-channel attacks to go after data in use.

So, how do we protect our data in use? The short answer is, we can't, at least for now. We can get close, however, by ensuring that our systems decrypt data at the very last possible moment, ideally as it gets loaded into the CPU registers, and encrypt it as it leaves those registers. This approach means that the data is encrypted even in memory, but it is an expensive approach that requires a cryptographic co-processor. You may encounter it if you work with systems that require extremely high security but are in places where adversaries can put their hands on them, such as automated teller machines (ATMs) and military weapon systems.

A promising approach, which is not quite ready for prime time, is called *homomorphic encryption*. This is a family of encryption algorithms that allows certain operations on the encrypted data. Imagine that you have a set of numbers that you protect with homomorphic encryption and give that set to me for processing. I could then perform certain operations on the numbers, such as common arithmetic ones like addition and multiplication, without decrypting them. I add the encrypted numbers together and send the sum back to you. When you decrypt them, you get a number that is the sum of the original set before encryption. If this is making your head hurt a little bit, don't worry. We're still a long ways from making this technology practical.

Standards

As we discussed in Chapter 1, *standards* are mandatory activities, actions, or rules that are formally documented and enforced within an organization. Asset security standards can be expensive in terms of both financial and opportunity costs, so we must select them carefully. This is where classification and controls come together. Since we already know the relative value of our data and other information assets and we understand many of the security controls we can apply to them, we can make cost-effective decisions about how to protect them. These decisions get codified as information asset protection standards.

The most important concept to remember when selecting information asset protection standards is to balance the value of the information with the cost of protecting it. Asset inventories and classification standards will help you determine the right security controls.

Scoping and Tailoring

One way to go about selecting standards that make sense for your organization is to adapt an existing standard (perhaps belonging to another organization) to your specific situation. *Scoping* is the process of taking a broader standard and trimming out the irrelevant or otherwise unwanted parts. For example, suppose your company is acquired by another company and you are asked to rewrite some of your company's standards based on the ones the parent company uses. That company allows employees to bring their own devices to work, but that is not permitted in your company. You remove those sections from their standard and scope it down to your size. *Tailoring*, on the other hand, is when you make changes to specific provisions so they better address your requirements. Suppose your new parent company uses a particular solution for centralized backup management that is different from the solution your company has been using. As you modify that part of the standard to account for your platform, you are tailoring it to your needs.

Data Protection Methods

As we have seen, data can exist in many forms and places. Even data in motion and data in use can be temporarily stored or cached on devices throughout our systems. Given the abundance of data in the typical enterprise, we have to narrow the scope of our data protection to the data that truly matters. A *digital asset* is anything that exists in digital

form, has intrinsic value to the organization, and to which access should be restricted in some way. Since these assets are digital, we must also concern ourselves with the storage media on which they reside. These assets and storage media require a variety of controls to ensure data is properly preserved and that its integrity, confidentiality, and availability are not compromised. For the purposes of this discussion, "storage media" may include both electronic (disk, optical discs, tape, flash devices such as USB "thumb drives," and so on) and nonelectronic (paper) forms of information.

The operational controls that pertain to digital assets come in many flavors. The first are controls that prevent unauthorized access (protect confidentiality), which, as usual, can be physical, administrative, and technical. If the company's backup tapes are to be properly protected from unauthorized access, they must be stored in a place where only authorized people have access to them, which could be in a locked server room or an offsite facility. If storage media needs to be protected from environmental issues such as humidity, heat, cold, fire, and natural disasters (to maintain availability), the media should be kept in a fireproof safe in a regulated environment or in an offsite facility that controls the environment, so it is hospitable to data processing components.

Companies may have a digital asset library with a librarian in charge of protecting its resources. If so, most or all of the responsibilities described in this chapter for the protection of the confidentiality, integrity, and availability of media fall to the librarian. Users may be required to check out specific resources from the library, instead of having the resources readily available for anyone to access them. This is common when the library includes licensed software. It provides an accounting (audit log) of uses of assets, which can help in demonstrating due diligence in complying with license agreements and in protecting confidential information (such as PII, financial/credit card information, and PHI) in libraries containing those types of data.

Storage media should be clearly marked and logged, its integrity should be verified, and it should be properly erased of data when no longer needed. After a large investment is made to secure a network and its components, a common mistake is to replace old computers, along with their hard drives and other magnetic storage media, and ship the obsolete equipment out the back door along with all the data the company just spent so much time and money securing. This puts the information on the obsolete equipment and media at risk of disclosure and violates legal, regulatory, and ethical obligations of the company. Thus, overwriting (see Figure 6-2) and secure overwriting algorithms are required. Whenever storage media containing highly sensitive information cannot be cleared or purged, physical destruction must take place.

When storage media is erased (*cleared* of its contents), it is said to be *sanitized*. In military/government classified systems terms, this means erasing information so it is not readily retrievable using routine operating system commands or commercially available forensic/data recovery software. Clearing is acceptable when storage media will be reused in the same physical environment for the same purposes (in the same compartment of compartmentalized information security) by people with the same access levels for that compartment.

Not all clearing/purging methods are applicable to all storage media—for example, optical media is not susceptible to degaussing, and overwriting may not be effective when