# The data lifecycle

Organizations of all sizes handle a large amount of data that must be kept private. You learned that data can be vulnerable whether it is at rest, in use, or in transit. Regardless of the state it is in, information should be kept private by limiting access and authorization.

In security, data vulnerabilities are often mapped in a model known as the data lifecycle. Each stage of the data lifecycle plays an important role in the security controls that are put in place to maintain the CIA triad of information. In this reading, you will learn about the data lifecycle, the plans that determine how data is protected, and the specific types of data that require extra attention.

## The data lifecycle

The data lifecycle is an important model that security teams consider when protecting information. It influences how they set policies that align with business objectives. It also plays an important role in the technologies security teams use to make information accessible.

In general, the data lifecycle has five stages. Each describe how data flows through an organization from the moment it is created until it is no longer useful:

- Collect
- Store
- Use
- Archive
- Destroy

Protecting information at each stage of this process describes the need to keep it accessible and recoverable should something go wrong.

## Data governance

Businesses handle massive amounts of data every day. New information is constantly being collected from internal and external sources. A structured approach to managing all of this data is the best way to keep it private and secure.

*Data governance* is a set of processes that define how an organization manages information. Governance often includes policies that specify how to keep data private, accurate, available, and secure throughout its lifecycle.

Effective data governance is a collaborative activity that relies on people. Data governance policies commonly categorize individuals into a specific role:

- **Data owner:** the person that decides who can access, edit, use, or destroy their information.

- **Data custodian**: anyone or anything that's responsible for the safe handling, transport, and storage of information.
- **Data steward**: the person or group that maintains and implements data governance policies set by an organization.

Businesses store, move, and transform data using a wide range of IT systems. Data governance policies often assign accountability to data owners, custodians, and stewards.

**Note:** As a data custodian, you will primarily be  responsible for maintaining security and privacy rules for your organization.

## Protecting data at every stage

Most security plans include a specific policy that outlines how information will be managed across an organization. This is known as a data governance policy. These documents clearly define procedures that should be followed to participate in keeping data safe. They place limits on who or what can access data. Security professionals are important participants in data governance. As a data custodian, you will be responsible for ensuring that data isn’t damaged, stolen, or misused.

## Legally protected information

Data is more than just a bunch of 1s and 0s being processed by a computer. Data can represent someone's personal thoughts, actions, and choices. It can represent a purchase, a sensitive medical decision, and everything in between. For this reason, data owners should be the ones deciding whether or not to share their data. As a security professional, protecting a person's data privacy decisions must always be respected.

Securing data can be challenging. In large part, that's because data owners generate more data than they can manage. As a result, data custodians and stewards sometimes lack direct, explicit instructions on how they should handle specific types of data. Governments and other regulatory agencies have bridged this gap by creating rules that specify the types of information that organizations must protect by default:

- **PII** is any information used to infer an individual's identity. Personally identifiable information, or PII, refers to information that can be used to contact or locate someone.
- **PHI** stands for protected health information.  In the U.S., it is regulated by the Health Insurance Portability and Accountability Act (HIPAA), which defines PHI as “information that relates to the past, present, or future physical or mental health or condition of an individual.” In the EU, PHI has a similar definition but it is regulated by the General Data Protection Regulation (GDPR).
- **SPII** is a specific type of PII that falls under stricter handling guidelines. The *S* stands for sensitive, meaning this is a type of personally identifiable information that should only be accessed on a need-to-know basis, such as a bank account number or login credentials.

Overall, it's important to protect all types of personal information from unauthorized use and disclosure.

## Key takeaways

Keeping information private has never been so important. Many organizations have data governance policies that outline how they plan to protect sensitive information. As a data custodian, you will play a key role in keeping information accessible and safe throughout its lifecycle. There are various types of information and controls that you’ll encounter in the field. As you continue through this course, you’ll learn more about major security controls that keep data private.