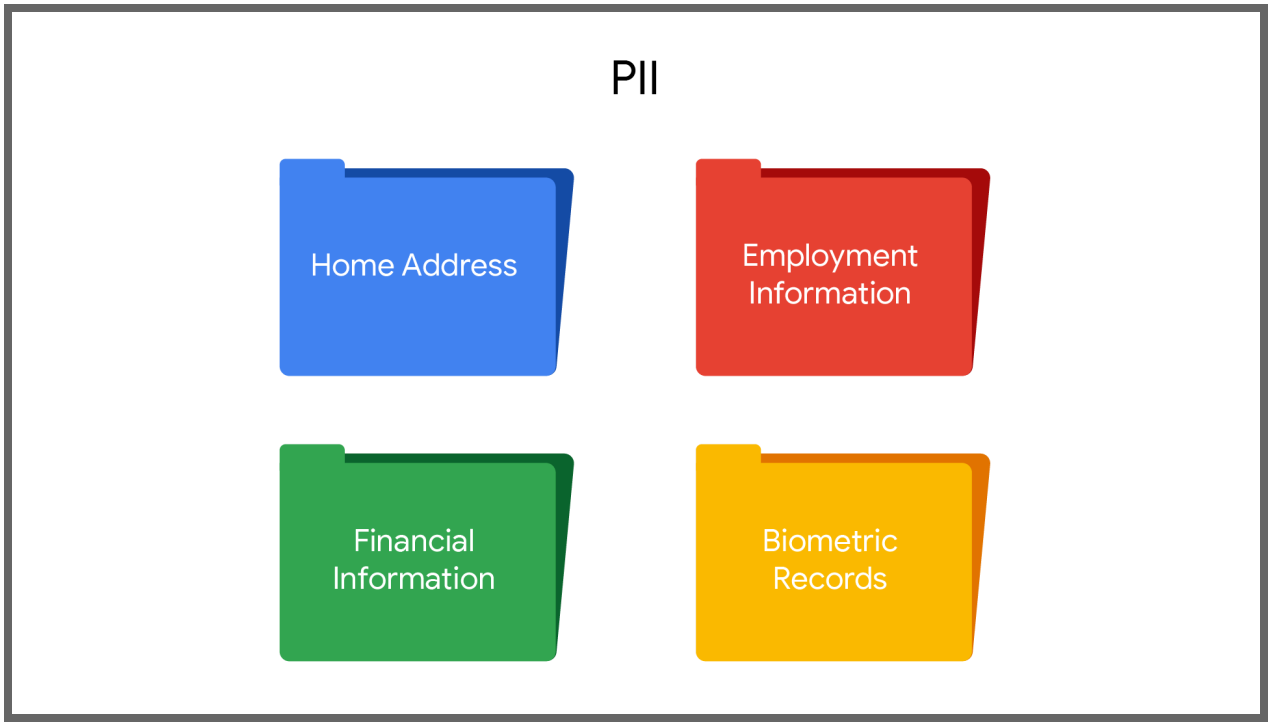
Google Cloud

# Protection of personally identifiable information (PII)

So far, you've learned that data classification is the process of analyzing structured or unstructured data to understand its sensitivity and value. You've also learned about the different levels of data sensitivity, including personal identifiable information (PII). Remember, PII is a type of highly sensitive data. In your cloud security career, you and your team will be responsible for protecting personal information from attackers. In this reading, you'll learn more about what types of information are considered PII, and ways to keep it safe. You'll also learn about data loss. Please note, the following reading should not be considered legal advice.

## What is personally identifiable information (PII)?

PII is data that can be used to identify an individual. This is usually sensitive data. Some examples include a person's name, an employee's home address, biometric records, medical data, educational data, financial data, or employment information.



## What makes data sensitive?

Sensitive data is any information that could cause embarrassment, inconvenience, or harm to an individual, or damage a company's reputation. It's important to protect sensitive data to safeguard organizations, users, and individuals.

Consider this example: A malicious actor conducts a phishing attack to gain access to user accounts. They then attempt to sell the stolen data, including social security and credit card numbers, online. This is an example of PII being accessed by an unauthorized entity and misused, causing harm to the user and the organization responsible for protecting user data.

## What is data loss?

Data loss is the unauthorized disclosure of proprietary, classified, or sensitive information. This can occur through data leakage or data theft. An example of data loss is the exposure of personal employment information to the public. As a cloud cybersecurity professional, you and your security team are in charge of protecting any and all forms of data.

### Preventing unauthorized access

Preventing unauthorized access and using de-identification are two ways you can protect against data loss. One method of preventing unauthorized access is to use identity and access management (IAM). IAM helps you secure data by helping you decide which user needs access to specific data, and assigning only the appropriate roles to that user. This is also called the Principle of Least Privilege. You also need to monitor the data usage to ensure that the users accessing it are using it only for authorized purposes.

You can decide which users need access to certain data by considering these two questions:

1. Who is accessing the data?
2. Why do they need the data?

### De-identification

Another approach to helping protect PII is through de-identification. According to the U.S. National Institute of Standards and Technology, "de-identification removes identifying information from a data set so that individual data can't be linked with specific individuals." By using de-identification, you're helping remove the parts of the data that can identify individuals. One method of de-identifying information is to redact it. To **redact** means to delete all or part of a detected sensitive value.

Another method to de-identify data is called replacement. In this approach, a cloud security professional uses a service or tool to replace a detected sensitive value with a generic value.

You can also use data masking by replacing a number of sensitive characters with a specific generic character, like a hash or an asterisk.

## Key takeaways

PII is sensitive data that can be used to identify an individual. If attackers get access to PII, they can use it to cause embarrassment, inconvenience or harm. It can also be used to harm a company's reputation. Data loss is the exposure of proprietary, classified, or sensitive data through data leakage or theft. Organizations can help protect PII and other sensitive data through the process of de-identification, or by using IAM to control data access.