

# **What is PII, non-PII and personal data? And how to protect each**



# What is PII, non-PII and personal data? And how to protect each

**Personally identifiable information (PII)** and **personal data** are two classifications of data that often cause confusion for organizations that collect, store and analyze such data.

PII is used in the US but no single legal document defines it. The legal system in the United States is a blend of numerous federal and state laws and sector-specific regulations. They all define and classify different pieces of information under the PII umbrella.

On the other hand, personal data has one legal meaning, which is defined by the General Data Protection regulation (GDPR), accepted as law across the European Union (EU).

Both terms cover common ground, classifying information that could reveal an individual's identity directly or indirectly.

But why is all that so important? As a website admin, app creator or product owner, you need to be aware that the traces visitors and users leave behind could be of a sensitive nature. These traces might enable you to identify individuals, so you need to handle such data with the utmost caution. From a legal standpoint, it could be a matter of breaches and violations with serious consequences. Grasping the bigger picture is crucial for your organization's security and legal compliance.

## What is personally identifiable information (PII)?

PII is often referenced by US government agencies and non-governmental organizations. Yet the US lacks one overriding law about PII, so your understanding of PII may differ depending on your particular situation.

The most common definition is provided by the [National Institute of Standards and Technology \(NIST\)](#), which says that:

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

However, the line between PII and other kinds of information is blurry. As stressed by the US [General Services Administration](#), the “definition of PII is not anchored to any single category of information or technology. Rather, it requires a case-by-case assessment of the specific risk that an individual can be identified”.

### **What pieces of information are considered PII?**

According to NIST, PII can be divided into two categories: linked and linkable information.

**Linked information** is more direct. It could include any personal detail that can be used to identify an individual, for instance:

- Full name
- Home address
- Email address
- Social security number
- Passport number
- Driver's license number
- Credit card numbers
- Date of birth
- Telephone number
- Owned properties e.g. vehicle identification number (VIN)
- Login details

- Processor or device serial number\*
- Media access control (MAC)\*
- Internet Protocol (IP) address\*
- Device IDs\*
- Cookies\*

\* **Of note!** NIST states that linked information can be “Asset information, such as Internet Protocol (IP) or Media Access Control (MAC) address or other host-specific persistent static identifier that consistently links to a particular person or small, well-defined group of people”. That means cookies and device ID fall under the definition of PII.

**Linkable information** is indirect and on its own may not be able to identify a person, but when combined with another piece of information could identify, trace or locate a person.

Here are some examples of linkable information:

- First or last name (if common)
- Country, state, city, zip code
- Gender
- Race
- Non-specific age (e.g. 30-40 instead of 30)
- Job position and workplace

## What is non-PII?

Non-personally identifiable information (non-PII) is data that cannot be used on its own to trace, or identify a person.

Examples of non-PII include, but are not limited to:

- Aggregated statistics on the use of product / service
- Partially or fully masked IP addresses

However, the classification of PII and non-PII is vague. Moreover, NIST doesn't reference cookie IDs and device IDs, so many AdTech companies, advertisers, and publishers consider them as non-PII. As we'll see, this is in contrast to the definition of personal data, which treats such digital trackers as information that could identify an individual.

## What is personal data?

Personal data is a legal term that the [GDPR](#) defines as the following:

### §

#### **Article 4(1):**

'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

This definition applies not only to a person's name and surname, but to details that could identify that person. That's the case when, for instance, you're able to identify a visitor returning to your website with the help of a cookie or login information.

Under the GDPR you can consider cookies as personal data because according to

---

---

**§**

---

---

**Recital 30:**

Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.

And the definition of personal data covers various pieces of information such as:

- transaction history
- IP addresses
- [browser history](#)
- posts on social media

Basically, it's any information relating to an individual or identifiable person, directly or indirectly.

# What is non-personal data?

Following the GDPR provisions, non-personal data is data that won't let you identify an individual. The best example is anonymous data. According to

## §

### Recital 26:

The principles of data protection should therefore not apply to **anonymous information**, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.

Other examples of non-personal data include, but are not limited to:

- Generalized data, e.i. age range e.g. 20-40
- Information gathered by government bodies or municipalities such as census data or tax receipts collected for publicly funded works
- Aggregated statistics on the use of a product or service
- Partially or fully masked IP addresses

**To learn more about data anonymization, read our other blog posts:**

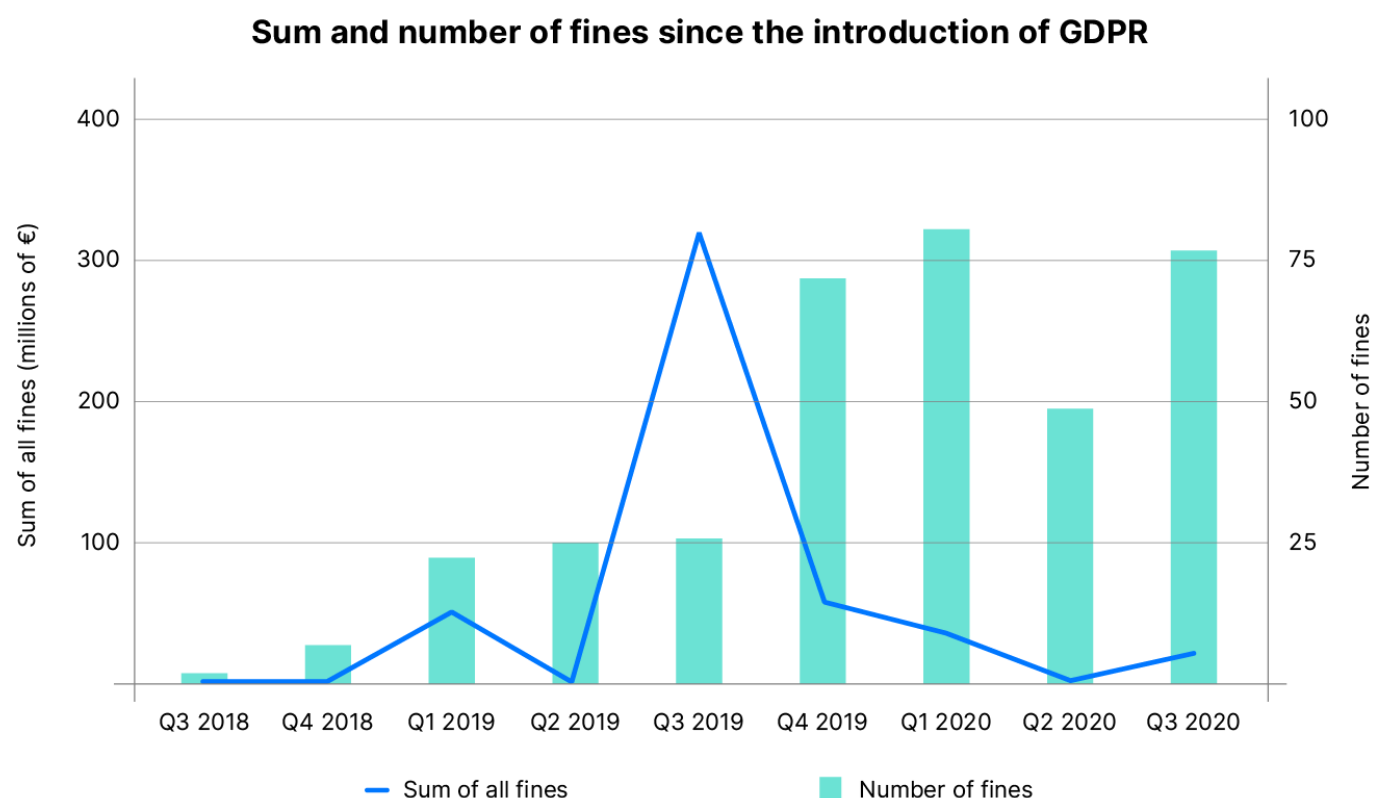
- [The ultimate guide to data anonymization in analytics \[updated\]](#)
- [Anonymous tracking: how to do useful analytics without personal data](#)

# How PII differs from personal data

As we've already mentioned, in certain contexts the differences between these two types of data seem quite vague. If we need to draw a clear line here, then we would apply the legal framework and whom this data applies to.

## Legal framework

All rules and responsibilities regarding personal data are set out by [the GDPR](#), which aims to strengthen and unify data collection from EU residents. This also means that there is a more unified approach to enforcement, which has been steadily increasing since May 2018, when GDPR entered into force.



It's much harder to define a single piece of legislation that controls PII because of the lack of a single federal law governing its use. However, among the various laws that do govern the collection and usage of PII, the most prominent are:

- The [U.S. Privacy Act](#), which governs how to collect, maintain, use and disseminate PII
- [The Health Insurance and Portability Act \(HIPAA\)](#) governing patient privacy
- The [Children's Online Privacy Protection Act \(COPPA\)](#), designed to protect the personal information of children under the age of 13



Furthermore, both governmental and non-governmental organizations regulate the proper use of PII, including:

- [The Federal Trade Commission \(FTC\)](#) and its Department of Consumer Protection
- Local Departments of Consumer Affairs
- The Federal Communications Commission (FCC)
- The National Institute of Standards and Technology (NIST)
- The Network Advertising Initiative (NAI), a self-regulatory organization

### **Where rules on PII and personal data apply**

Since personal data is strictly connected to the GDPR, it concerns all residents and citizens of the member states of the European Economic Area – the 28 Member States of the EU plus Iceland, Liechtenstein, and Norway. We'll refer to this group as EU residents, for short.

Still, the scope of the GDPR is not really limited to the EU. It impacts not only EU-based entities, but virtually every business dealing with the data of EU residents.



By contrast, it's much more difficult to determine the jurisdictions where PII is applicable.



Even in the US, where PII is certainly applicable, how it's applied varies both from state to state and from sector to sector. Several legal documents and industry standards have their own opinion about what PII is.



As a result, determining who PII applies to and how is quite difficult.

# PII vs personal data comparison table

PII, personal data or both		
Personally identifiable information (PII) – a term regularly used in AdTech and MarTech as well as US government agencies such as the National Institute of Standards and Technology (NIST).	vs	Personal data – a legal term defined by the General Data Protection Regulation (GDPR).
We've split the data into three categories: direct, less direct and modern digital identifiers. These aren't official categories, they are just there to make the table more readable.		

Direct data that could identify and individual in very few steps		
	PII 	PERSONAL DATA 
Name: full name, maiden name, mother's maiden name or alias	✓	✓
Home address	✓	✓
Email address	✓	✓
Date of birth	✓	✓
Telephone number	✓	✓
Personal identification numbers: social security number (SSN), passport number, driver's license number, taxpayer identification number, etc.	✓	✓
Personal characteristics: photographic images, fingerprints, or handwriting	✓	✓
Biometric data: retina scans, voice signatures, or facial geometry	✓	✓
Information identifying personally owned property: VIN number or title number	✓	✓
Login details	✓	✓

<b>Less direct</b> data that reduces the number of possible individuals, but just one or two of these probably wouldn't be sufficient to identity someone		
	<b>PII</b> 	<b>PERSONAL DATA</b> 
Country, state, city, postcode	✓	✓
Place of birth	✓	✓
Gender	✓	✓
Race	✓	✓
Religion	✓	✓
Non-specific age (e.g. 30-40 instead of 30)	✓	✓
Employment information	✓	✓
Office telephone number	✓	✓
Office/work address	✓	✓

<b>Modern digital identifiers</b> Data from the internet age which we're learning can more and more easily identify individuals, especially in combination with other pieces of personal data		
	<b>PII</b> 	<b>PERSONAL DATA</b> 
Device IDs	✗*	✓**
IP addresses	✗*	✓**
Cookies	✗*	✓**
Browser type	✗*	✓**
Device type	✗*	✓**
Plug-in details	✗*	✓**

Language preference	X*	✓**
Time zones	X*	✓**
Screen size, screen color depth, and system fonts	X*	✓**
Browsing history	X*	✓**
<p>* it's still unclear for PII since different laws govern the use of PII and are not always consistent in the classification: for now each of these is considered on a case by case basis</p> <p>** GDPR states that even those kinds of information can be considered personal data if there is any potential to use them to single out or identify an individual. This is detailed in Recital 30 of the new law: "Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them."</p>		

## Protection of personal data

Guidelines for protecting personal data are easy to come by because the GDPR clearly lays them out. Here are the most important ones to keep in mind.

### The principle of lawfulness

Most importantly, the GDPR requires a clear objective and valid grounds to collect and use personal data. That means having a “lawful basis”, which is the starting point for the principle of lawfulness.

The GDPR dictates that the reason for processing data, the lawful basis, needs to be based on “necessity”. It could be, for instance, fulfilling a contract requirement or providing a service. That necessity is from the point of view of the data owner, the individual, or data subject in GDPR jargon. Even if a business believes it needs the data to make a profit, this is not an appropriate lawful basis.

## The principle of integrity and confidentiality

One of the key ways to protect data is to ensure its security. The GDPR addresses this matter in its principle of integrity and confidentiality. Article 5(1) states that personal data shall be:

### §

#### Article 5(1):

processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality')

What's more, Article 32 underlines how important is for businesses to take an appropriate approach to secure personal information:

### §

#### Article 32:

Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk.

However, the GDPR doesn't exactly say what this security should look like in practice. That's because a lot depends on the organization in question. For instance, a hospital with sensitive information about its patients will take different steps than a blogger with a newsletter.



If you'd like to dive deeper into those principles, read the [ICO's dedicated guide](#).

## Data protection by design

Protection of personal data is the key objective of the GDPR and in Article 25 you'll find the principle of data protection by design and default. It has its roots in the already existing concept of "[privacy by design](#)."

Data protection by design means adopting technical and organizational measures in the initial design phases of processing operations. In this way, your organization ensures that privacy and security mechanisms are in place from the beginning.

Some practical examples are:

- Pseudonymization – replacing or removing information within a data set that enables identification of a specific person, using methods such as encryption, scrambling or masking
- Anonymization – removing personal information from a data set to make it impossible to identify a particular person
- Monitoring of data processing
- Adding new privacy features and improving existing ones

Keep in mind that under [GDPR](#), pseudonymization techniques are not enough to provide full data anonymity. Recital 26 says:

*Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person[...]*

To learn more about all the details of this issue, read our post:

[Data pseudonymization in web analytics: the ultimate guide](#)

## Data protection by default

Data protection by default is based on two key principles: data minimization and purpose limitation.

Data minimization lays down the idea of only processing data that is indispensable for a particular purpose. According to the GDPR, the data should be “adequate, relevant and limited to what is necessary,” enabling you to fulfill the specific purpose of processing that piece of data.

Purpose limitation means you specify your processing purpose, document it and inform individuals about this purpose before any processing starts.

Organizations that act in line with those principles will collect only the minimum amount of data possible and keep it for only as long as is necessary to fulfill the purpose for which it was collected.

From a practical point of view, this means starting any project by asking questions.

Questions such as:

- What kind of data do I need?
- Why do I need this data?
- How much data is necessary?

It's best to then establish a review process to regularly verify data you hold and delete what you no longer need.



If you want to learn more about data protection by design and default, read our blog post: [Privacy by design under the GDPR](#)

## Data protection impact assessment

The GDPR offers recommendations in some critical scenarios. For example, it recommends performing a data protection impact assessment (DPIA) when it's likely that the processing might cause high risk to individuals. [DPIAs](#) help organizations lower that risk by recognizing and mitigating possible threats.

According to Article 35, the regulation says:

### §

#### Article 35:

Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data.

Consider running a DPIA when you are:

- Using a new technology
- Processing sensitive data on a large scale, e.g. data related to racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership
- Doing systematic monitoring of public areas

## Protection of personally identifiable information (PII)

Unlike for personal data, there is no one overriding law or document that governs and controls how PII is defined and treated.

The most widely referenced document about PII is [NIST's Guide to Protecting the Confidentiality of Personally Identifiable Information](#). Although created in 2010, organizations across different sectors still use it as a foundation for data protection protocols.



The document provides similar recommendations to those the GDPR provides for personal data. For instance, NIST explains that “organizations should minimize the use, collection, and retention of PII to what is strictly necessary to accomplish their business purpose and mission.” This echoes GDPR’s principles of lawfulness and minimization.

Other PII protection methods will also sound familiar:

- Privacy Impact Assessment
- Data encryption
- Data anonymization

### **How to choose appropriate safeguards for PII**

The NIST guide makes the point that not all PII needs to be protected equally and proposes [six key factors](#) to consider when deciding what safeguards to apply:

1. How easily the PII can be tied to specific individuals
2. The number of individuals whose PII is stored in the system
3. The sensitivity of the data
4. The context of how the data will be used, stored, collected, or disclosed
5. Legal obligations to protect the data
6. The location of the data, and level of authorized access to the data

### **Foundations of PII protection**

NIST explains further that the protection of PII requires a combination of measures such as “operational safeguards, security controls and safeguards related to privacy.” In its *Guide to Protecting the Confidentiality of Personally Identifiable Information*, you’ll find a comprehensive set of rules in that field. Here are the key ones:

#### **Operational safeguards**

The protection of PII starts at the businesses’ operational level. It involves creating and establishing detailed policies and procedures for managing PII. The safeguards include making the staff aware of the importance of this issue, informing them about possible risks such as the latest phishing scams.

Finally, it means providing employees with training about the best practices for handling and protecting PII.

### Privacy-specific safeguards

The protection of PII confidentiality requires different mechanisms which are not usually necessary with other types of information. That could be, for instance, data anonymization and de-identifying information, also known as encryption.

Privacy-specific safeguards help businesses follow the data minimization principle. They also allow organizations to use and maintain data without risking its confidentiality.

### Security controls

NIST offers recommendations regarding security controls for protecting PII, but organizations are in the best position to decide on what's best for them. NIST suggests:

- **Access enforcement** – control over the access to PII using control policies and access enforcement mechanisms, e.g. granting access to the data based on the user's role
- **Remote access** – prohibiting or restricting access to PII and when a user has a remote access and ensuring communication is encrypted
- Identification and authentication of users
- **Separation of duties** – e.g. users who handle de-identified PII wouldn't also be in roles that grant them access to the information needed to re-identify the records
- Information system monitoring
- **Least privilege** – making sure that users have access to only the data they need
- **Audit review, analysis and reporting** – conducting a regular review and analysis of records after the information system audit to spot any unusual activities affecting PII

## Staying up to date on data privacy regulations

The broad definitions of PII and personal data are evolving to cover more and more kinds of data. The differences between the two are also becoming less distinct. The legal requirements are getting stricter on both sides of the Atlantic.

Those changes will bring new challenges. For organizations of all kinds, this means taking a closer look at the data they collect and keeping up with the changing legal landscape to stay compliant.

### **Responsibly collect, store and use PII and personal data**

Learn more about an analytics platform that gives you a complete set of data  
privacy and security features

**Request a personalized demo**

# About Piwik PRO

Piwik PRO is the first privacy-oriented alternative to Google Analytics. Created in 2013, the Piwik PRO Analytics Suite is chosen by governments, public institutions, private companies and enterprises. They turn to our stack to analyze and optimize the full customer journey while adhering to the strictest security policies and privacy regulations in the world. The platform comes together with high-touch customer care, training and professional guidance in obtaining the most valuable information for your business.

## Contact

### EMEA

+48 71 716 69 50

### DACH

+49 2203 989 620

### BENELUX

+31 858 881 458

### NORTH AMERICA

+1 (888) 444 0049

<https://piwik.pro>

[sales@piwik.pro](mailto:sales@piwik.pro)

