WHITE PAPER

GROUND LABS

Do You Know

Where Your PII Data Is?

It might sound like a simple question, but the implications if you don't may lead to massive reputational damage, loss of customer trust and fines running into the millions.

CONTENTS

What is PII?

Personally identifiable information (PII) is any data that can be used to identify a person, either on its own or when combined with other information. Many types of data can be considered PII, and the kinds of data that constitute PII vary from industry to industry, company to company, and individual to individual. This can range from basic identifying elements, such as name, address and email, to more sensitive information including social security numbers, passport details and medical or health insurance information.

Companies accumulate PII from many sources, such as employees, customers, clients, patients, students, partners and more, depending on the industry. PII may be necessary to carry out legitimate business activities, which could include: an ecommerce website that needs a customer's name and address to ship goods they've purchased; a hospital that needs medical and health insurance information to provide treatment to a patient; a bar that needs to see a driving license to confirm a person's age before selling alcohol to them. Individuals may consent to the use of their data for specific purposes, such as opting into marketing emails and offers.

Lawful Basis for Use of PII

The use of PII by organizations must have a lawful basis. Although the definition of this may change depending on the local jurisdiction, the requirements set out by the General Data Protection Regulation (GDPR) are broadly in line with other regulations. According to the Information Commissioner's Office (ICO), the UK's data regulator, these include:[1]

- **Consent:** the individual has given clear consent for you to process their personal data for a specific purpose.
- **Contract:** the processing is necessary for a contract you have with the individual, or because they have asked you to take specific steps before entering a contract.
- **Legal obligation:** the processing is necessary for you to comply with the law (not including contractual obligations).
- **Vital interests:** the processing is necessary to protect someone's life.
- **Public task:** the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law.
- **Legitimate interests:** the processing is necessary for your legitimate interests or the legitimate interests of a third party, unless there is a good reason to protect the individual's personal data which overrides those legitimate interests.

There are lots of other laws and regulations that have similar stipulations around when it is permissible to gather, process and store PII. A few from the US market — the nation with the highest number of privacy and data protection statutes — are listed below:

- **Gramm-Leach-Bliley Act (GLBA)** — requiring financial institutions to provide a privacy notice to customers explaining how their data will be used.
- **Fair Credit Reporting Act (FCRA)** — regulating how consumer reporting agencies use credit information.
- **Health Insurance Portability and Accountability Act (HIPAA)** and **Health Information Technology for Economic and Clinical Health Act (HITECH)** — for healthcare related information to aid patient treatment.
- **The Family Educational Rights and Privacy Act (FERPA)** — relating to PII protections for student educational records.
- **The Children's Online Privacy Protection Act (COPPA)** — relating to the privacy of children under the age of 13.
- **The Privacy Act of 1974** — requiring fair information practices by US federal agencies.

Collecting and Using PII

An organization should not have more PII than is needed to achieve the stated purpose, nor should the data include out-of-date or irrelevant details.

The examples below from the ICO highlight situations where PII can be collected and highlights how inappropriate use may breach GDPR.

> **A debt collection agency** is engaged to find a particular debtor. It collects information on several people with a similar name to the debtor. During the enquiry some of these people are discounted. The agency should delete most of their personal data, keeping only the minimum data needed to form a basic record of a person they have removed from their search. It is appropriate to keep this small amount of information so that these people are not contacted again about debts which do not belong to them.

If you need to process information about certain individuals only, you should collect data only related to that individual — the information is likely to be excessive or irrelevant in relation to other people.

> **A recruitment agency** places workers in a variety of jobs. It sends applicants a general questionnaire, which includes specific questions about health conditions that are only relevant to manual occupations. It would be irrelevant and excessive to obtain such information from an individual who was applying for an office job.

You must not collect personal data on the off chance that it might be useful in the future. However, you may be able to hold information for a possible event that never occurs if you can justify it.

> **An employer** holds details of the blood groups of some of its employees. These employees do hazardous work, and the information is needed in case of an accident. The employer has safety procedures in place to help prevent accidents so it may be that this data is never needed, but they are permitted to hold this information in case of emergency.

If the employer holds the blood groups of the rest of the workforce though, such information is likely to be irrelevant and excessive as they do not engage in the same hazardous work.

If you are holding more data than is necessary for your business purposes, this is likely to be unlawful, as well as a breach of the data minimization principle. Individuals will also have the right to erasure.

If you are holding more data than is necessary, this is likely to be unlawful

When Does PII Become a Problem?

In the examples above, PII may well have been collected lawfully and potentially stored within one or more IT systems. Even if PII has been collected and processed legitimately, if it is used for purposes other than those stated, it may breach regulations such as GDPR. If an organization isn't aware of where it's storing PII, or if it isn't storing PII securely, it could also be putting itself and its customers at significant risk of regulatory failure and data breach respectively.

Cybersecurity

Many types of bad actors might be motivated to steal PII, including cybercriminals and disgruntled employees. Once an unauthorized party gains access to PII, there's a wide range of nefarious things they can do with it. This may include selling the data on the black market, making purchases with stolen credit card numbers, impersonating an individual to commit fraud or filing stolen tax returns.

In January 2020, for example, hackers stole payment data from the convenience store chain Wawa and sold millions of Wawa customers' credit and debit card numbers on the dark web.[2]

Another example was the Capital One data breach in July 2019. In this incident, a former Amazon Web Services employee with insider knowledge of Capital One's data storage infrastructure stole personal data of more than 100 million Capital One customers.[3]

In 2019 a former Amazon Web Services employee, stole personal data of more than 100 million Capital One customers.[3]

Regulatory Issues

The main penalty for poor PII practices is regulatory action. The EU GDPR is among the world's toughest data protection regulations. Under the GDPR, the EU's data protection authorities can impose fines on organizations of up to €20 million (roughly $20,372,000) or 4% of their worldwide turnover for the preceding financial year, whichever is higher. Since the GDPR took effect in May 2018, over 900 fines have been issued across the European Economic Area. Although Amazon, Facebook and Google all made headlines for fines that totaled over a billion dollars — for a range of issues stemming from the way the trio way and collects and shares personal data via cookies — there are other data misuse examples that resulted in fines.

H&M, a clothing retailer, was fined $41 million by the German federal data protection authority, BfDI, for GDPR violations involving the "monitoring of several hundred employees." This allowed senior H&M staff to gain "a broad knowledge of their employees' private lives ... ranging from rather harmless details to family issues and religious beliefs." This was then used to evaluate employees' performance and make decisions about their employment.[4]

In another recent case from 2022, Enel Energia, an Italian energy supplier, was fined $29.3 million by the Italian data protection authority, GPDP, for a range of GDPR breaches including failing to gain consent or inform customers before using PII for telemarketing calls.[5]

Using a Data Discovery Solution

Data discovery solutions are software tools that are used by organizations to find and remediate PII and sensitive information across the broadest range of structured and unstructured data — whether it's stored across servers, on desktops, in emails and databases, on-premises or in the cloud.

Data Discovery

Data discovery is the first step for organizations looking to comply with data protection regulations worldwide. This stage helps organizations to understand where their data lies and what types of data they possess. Discovery starts with a purpose: that may be compliance but could also cover a variety of resilience and security purposes.

Organizations need to understand where and how their data is stored — on-premises, third-party servers or in the cloud. While they might already know the location of structured data such as a primary customer database store, unstructured data — such as that found in stray files and emails — will have to be hunted down too.

Data discovery offers new insights as to the nature of that data and its location, and critically, will help you build its risk profile.

Data Classification

Once an organization's data has been discovered it can be categorized across a variety of metrics — according to the sensitivity of the data or ease of identification of individuals from the data, for example — and classified depending on its relative risk. Many data protection regulations require organizations to classify data according to its sensitivity and complying organizations should keep this obligation at the center of their focus.

PII data will typically be high risk and classified as Confidential (for example). This indicates that it must be protected by adequate security and privacy controls, such as anonymization and encryption. If it is breached, it will cause harm to the organization and may damage the privacy of those whose data is compromised.

Data classification enables organizations to allocate resources to protect sensitive information wherever it is stored and mitigate privacy and compliance-endangering risks. Data discovery ensures that all your PII data is identified, helping to verify whether your organization is storing its data in a secure and compliant manner.

Benefits of Ground Labs ENTERPRISE RECON

Enterprise Recon is a software solution that enables sensitive data discovery across a wide variety of targets including workstations, servers, database systems, big-data platforms, email services and a range of cloud storage providers.

Enterprise Recon also includes a variety of marking and remediation options, depending on the platform where data is found, to help categorize findings and perform affirmative action on sensitive data file locations.

With over 300 built-in data types spanning PII data as it is defined by over 50+ countries, and a flexible custom data type creation module for any special or unique requirements, Enterprise Recon helps organizations identify a broad variety of critical data types that require higher levels of security to comply with regulations, standards and requirements such as PCI DSS ®, GDPR, HIPAA, CCPA and more. Enterprise Recon has several crucial advantages over many data discovery tools including:

Flexibility

Enterprise Recon can scan all file types. If your organization is storing compressed files, Enterprise Recon un-compresses them so it can scan the raw data. Enterprise Recon also includes an Optical Character Recognition (OCR) engine to extract text from images before scanning to allow discovery of data in image files.

Effectiveness

Enterprise Recon scans all files in their entirety. The solution uses a forensic-level approach to search within hidden locations such as unallocated sectors, shadow volumes, and memory. Many data discovery solutions will only scan a few bytes at the start of each file or randomly scan sections of targets to reduce scanning time. However, this method often misses sensitive data spread throughout the target systems. Missing data and false negatives can result in non-compliance with policies and regulations.

Compact and secure

Enterprise Recon is a solution that is designed to be incredibly lightweight, executing with minimal impact on target systems. A large telecommunications company using the product for over five years commented on this and stated they have never had a department complain about any slow-downs or impact while scans are running. Enterprise Recon scans data in place; data is not copied prior to scanning. This is critical for security driven requirements.

Actionable insights

Enterprise Recon offers a variety of remediation methods. It can delete, encrypt, quarantine, or mask information within files to secure the sensitive data found. Delete remediation is also available for Exchange Online/O365 targets.

Efficient operation

Enterprise Recon's workflow enables false positives to be discarded quickly. In addition, advanced filtering and custom data pattern definitions allow existing pattern types to be edited or combined. For example, you could create a custom pattern that allows you to find an address that also has a name on the same line or identify a pattern that contains specific data within X characters before or after a certain keyword.

Protect PII With GROUND LABS

Any organization that handles personal or sensitive information should implement comprehensive data discovery solutions to support compliance with GDPR, CCPA, HIPAA, PDPA, Australian Privacy, PIPEDA, and other data security standards. Knowing what kind of data your organization is storing and where it resides will help you comply with local and international security regulations and maintain trust with customers.

You can learn more about Ground Labs and Enterprise Recon at

www.groundlabs.com ▶

Established in 2007 and trusted by more than 4,500 companies in 85 countries, Ground Labs offers award-winning data discovery and management solutions for all industry sectors.

www.groundlabs.com

CONTACT:

US	+1 737 212 8111
UK	+44 203 137 9898
Ireland	+353 1 903 9162
Australia	+612 8459 7092
Asia	+65 3133 3133
Email	info@groundlabs.com

Endnotes

1 https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/

2 https://krebsonsecurity.com/2020/01/wawa-breach-may-have-compromised-more-than-30-million-payment-cards/

3 https://www.capitalone.com/digital/facts2019/

4 https://www.businessinsider.com/hm-fined-41-million-for-staff-privacy-breaches-in-germany-2020-10?r=US&IR=T

5 https://regtechtimes.com/enel-energia-imposed-fine-for-26-5-million/

DOCUMENT LAST UPDATED:
DECEMBER 2022