PII By Design™ | Cheat Sheet

Everything you need to know about PII By Design™ in one place. It is our unique approach to protect data and achieve compliance. Some PII, like identifiers, are more sensitive than others and require heightened protection. If you are a developer working on code that includes handling PII or personal data, you should understand the required policies from your organization.

Privacy by Design (PbD)

A method of planning and implementing a system and architecture that fully supports individual rights and protects people's data. Learn more here.

Privacy Engineering Tips

1. Avoid privacy anti-patterns:
   a. Logging PII in plaintext
   b. Passing PII in HTTP Get params in URLs as they might get logged
   c. Passing PII in Kafka and other message buses can be kept forever - tokenize it
   d. Exposing internal objects and IDs to external clients (e.g a JSON obj might contain more fields than you'd expect)
2. Tokenize PII instead of passing them directly across systems in plaintext
3. Omit identifiers in analytics pipelines as they're exposed to all employees
4. When masking - always mask on the server side
5. Login system shouldn't leak intel about existing users (e.g. logging by email is problematic)
6. Centralize and segregate identifiers and sensitive data for control
7. Encrypt, mask and tune access policies, and regularly audit PII
8. Keep an updated data inventory map of where to find customers or employees PII

Privacy Principles

Limits on collected data:

- **Transparency** – For personal data you must explain how it will be collected and processed and, if required by law, you may need to ask for consent
- **Purpose limit** – You must only use PII for specified and legitimate purposes
- **Consent** – You must honor user consent and preferences when using or sharing personal data data
- **Minimization** – You must collect and process the minimum amount of personal data required for the purpose
- **Retention** – You must define policies to delete data (automatically, or manually)

Data subject access rights:

Deletion – You must be able to delete user personal data when asked, as necessary
Access – You must provide a copy of the personal data to the user when asked
Accuracy – User data must be correct, and you need to correct it if asked

Security:

Data protection – You must keep data safe from threats.
Methods – access limitations, encryption, masking, tokenization, logging, DLP, auditing and monitoring.

Importance of De-Identification

De-identifying personal data reduces the risk to the related individuals. While truly achieving anonymized data is very hard, there's a great benefit in de-identifying or pseudonymizing data by removing identifiers. **If stolen or lost customer records do not include PII, it may not be necessary to report a breach**. In addition, cross-border transfer of pseudonymized data can be considered lower risk.
It is recommended to get rid of data when no longer needed, to anonymize it when possible, at least to de-identify it, or not collect it at all to begin with.

Key Identifiers

ANY information that helps directly identify a person. We believe it's most effective to segregate, protect, encrypt, and tokenize *at least* the Identifiers listed below as a first priority. Information about individuals is considered personal data protected by privacy laws when it is possible to identify these individuals directly or even indirectly by their attributes. Note that sometimes revealing an individual can be achieved by inferencing different data types or other related information. Generally, the larger the amount of such data available about an individual, the easier it is to identify said individual.

In Case of a Breach

If lost records contain PII, you will need to report it to the affected individuals and sometimes to regulators depending on the risks presented by the breach (usually, there are some thresholds). Note that in some cases, even if you are not sure which information was lost, you may still be obliged to report a breach.

Disclaimer

We're not lawyers, use this information on an "as-is" basis, at your own risk. Consult with a privacy professional within your organization, such as a DPO, Chief Privacy Officer, Privacy Engineer, or your legal department for understanding what's sensitive in your business case. A privacy impact assessment (PIA) may help.

Personal and Sensitive Data Attributes by Categories:

Identifiers

- **Full name**
- **Phone number**
- **Personal email**
- **Full address,** zip, country, city
- **Social security number** (SSN)
- **Tax ID**
- **National ID**
- **Passport ID**
- **Driving license number**
- **Nickname and screen handle**

Sensitive Attributes

- Biometrics
- Race
- Gender
- Sexual orientation
- Religious belief
- Political opinions
- Minors' data

Financial

- **Credit card number (PAN), debit card number, expiration date, CVV, cardholder PIN, last four digits**
- **Bank account number (BAN), routing number**
- **Financial institution account numbers**
- Account balances
- Credit score
- Payment history
- Financial assets/investments/purchases
- Tax information

Medical

- Account number
- Medical history
- Test and laboratory results
- Mental health conditions
- Treatment information

Digital Identifiers

- **Web URLs** to personal accounts
- **Username and passwords**
- **Browser cookies, fingerprinting identifiers**
- **Device information, persistent IDs, ad tracker identifiers**
- Security questions and answers
- IP address
- Browsing history, search history
- Pictures, audio, video
- Signature/avatars

Glossary

- **PII** - Personally Identifiable Information.
- **PHI** - Protected Health Information.
- **Personal Data** - information that relates to an identified or identifiable individual.
- **Deidentification** - any process of removing the association between a set of identifying data and the data subject.
- **Anonymization** - (subcategory of de-identification) whereby data subjects can *never* be re-identified.
- **Pseudonymization** - the procedure by which personal identifiers in a set of identifying data are replaced with tokens (non-sensitive identifiers).
- **GDPR** - General Data Protection Regulation.
- **CCPA** - California Consumer Privacy Act.
- **LGPD** - Brazilian General Data Protection Law.
- **PIPL** - The China Personal Information Protection Law.
- **HIPAA** - Health Insurance Portability and Accountability Act.