Parties that collect data systematically from websites by use of an automated device (“**Data Collectors**”) should be mindful of privacy laws and regulations, which have evolved significantly in recent years. Two recently-enacted privacy frameworks likely to be relevant to Data Collectors are:

(1) the European General Data Protection Regulation (“**GDPR**”), which became enforceable on May 25, 2018; and

(2) the California Consumer Privacy Act (“**CCPA**”), which became effective on January 1, 2020 (the GDPR, CCPA, and other privacy laws and regulations are referred to in this document collectively as “**Privacy Laws**”).

The GDPR and CCPA often impact Data Collectors not based in Europe or California. The CCPA applies to many entities merely “doing business” in California, and the GDPR applies to any entity globally that processes personal data relating to EU data subjects in connection with offering goods or services in the EU (or monitoring the behavior of EU data subjects).

**When is Privacy Law applicable?**

The GDPR governs “personal data,” which is defined as “any information relating to an identified or identifiable natural person” and the CCPA governs “personal information,” which is defined as “information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household” (data meeting either definition is referred to in this document as “**Personally Identifiable Information**” or “**PII**”). Data Collectors should assess whether data is PII before collecting it from a source. If the data includes PII, then collecting, storing, analyzing, or otherwise processing that data may be restricted or subject to certain requirements under Privacy Laws.

Whether data constitutes PII is often a fact-based analysis and may require advice from counsel with expertise under applicable Privacy Laws. However for a preliminary indication, below are common types of data that may be available on public-facing websites and their likely classifications under Privacy Laws.

|  |  |  |
| --- | --- | --- |
| **Often PII** | **PII unless Anonymized** | **Usually not PII** |
| Names | Limited geolocation data | Economic statistics |
| Phone numbers | Transaction data | Prices for financial instruments |
| Mailing addresses | Survey response | Production quantities |
| Email addresses | Search queries | Trading volumes |
| Social media posts and profiles |  | Aggregated data (including aggregated transaction, survey, search, or geolocation data) |
| Images of individuals |  | Corporate financial information |

**Anonymized Data**

Processing a dataset to remove identifying information is called “anonymization.” Data that is properly anonymized does not constitute PII and is therefore generally not subject to restrictions under Privacy Laws. However, efforts to anonymize data may be incomplete or imperfect, and may not be effective in excluding a dataset from the reach of Privacy Laws. For example, de-identified geolocation data may show that an individual spends most nights at a certain address, which would allow a recipient of the data to discern the individual’s home address. Data Collectors should consult with counsel and potentially data scientists to determine whether purportedly anonymized data still constitutes PII.

**Scraping and processing PII**

If a Data Collector is considering whether to collect data containing PII, the Data Collector should consult with counsel to understand applicable limitations and obligations under the GDPR and CCPA, as well as other applicable Privacy Laws. These requirements may include:

* Minimizing the scope of data collected;
* Limiting the purposes for which PII can be processed and the scope of processing activities;
* Notifying data subjects to whom the PII relates;
* Purging or providing PII relating to a data subject at the data subject’s request;
* Imposing certain contractual obligations on vendors who process PII; and
* Conducting a data protection impact assessment.

The applicability of such restrictions is further complicated in the United States, where it has been argued—but not yet held by any court—that restricting the dissemination of accurate, publicly available information is inconsistent with the First Amendment.

To the extent restrictions under Privacy Laws are valid and apply to Data Collectors, failure to abide by them can carry serious consequences. For example, in March 2019, the Polish Supervisory Authority fined a Polish company more than €200,000 for processing PII scraped from a public source without properly notifying the individual data subjects to whom the data related (as required by the GPDR).[[1]](#footnote-1)

1. [*See, e.g.*, <https://www.insideprivacy.com/data-privacy/polish-supervisory-authority-issues-gdpr-fine-for-data-scraping-without-informing-individuals/>] [↑](#footnote-ref-1)