

Breaking Down Big Data: De-identification Standards to Protect Personal Information

Technology, Privacy, and eCommerce





“Breaking Down Big Data” is a new column written by members of the ACC big data sub-committee. Here, they discuss how to manage the ever-changing big data issues in the legal field.

Individuals value their privacy. In contrast, businesses value the ability to leverage personal information to deliver quality products and services to meet the needs of their clients. The legal standards that regulate the protection of personal information help bridge the gap between these two opposing interests.

Generally, under these standards, businesses retain the ability to analyze large data sets and create information assets to support key business objectives. This is possible through a framework of conditions intended to protect and preserve the sensitivity of the personal information provided. This includes requirements to provide notice of the intended use of the data and offer individuals with the choice to move forward with disclosure before collection and use.

The first part of this series highlights the topic of de-identification, which is a technique required and employed by businesses to process personal information beyond typical regulatory constraints. Specifically, this article will address when de-identification may be applied, the legal standards under specific regulations for de-identifying personal information, and the effect meeting such de-identification standards has on the use of the remaining data set.

Obtaining the right to de-identify personal information

The ability to de-identify personal information is governed by statutes and contracts. For instance, under the Health Insurance Portability and Accountability Act (HIPAA), a business associate may use or disclose protected health information (PHI) as permitted by its business associate contract or as required by law. The business associate agreement must establish the permitted and required uses and disclosures of PHI, and should specify whether the business associate is permitted to de-identify PHI in accordance with 45 CFR 164.514(a)-(c).

In the context of the European Union's General Data Protection Regulation (GDPR), anonymizing data is a form of data processing. Data processing requires obtaining unambiguous consent from data subjects before proceeding (an "opt-out" notice does not qualify), unless a different legal basis exists (i.e., such processing would be required in relation to a contract entered into by the data subject). Lastly, where no statutory guidance exists, many commercial agreements will limit the processing of customer data solely to the extent necessary to deliver the services as described under the respective agreement. As a result, vendors should take precautionary measures to ensure contract language is drafted broadly enough to account for processing of customer data for purposes of current or future data analytics offerings.

Overview of de-identification standards

De-identification occurs when an individual's identity is no longer ascertainable or the risk of identifying an individual is significantly low due to the removal of direct personal identifiers (e.g., a data subject's first and last name) and indirect identifiers (phone numbers, email addresses, etc.). Regulatory requirements dictate whether a given de-identification standard has been met and the effect that meeting such standard has on the use of the remaining data set. Below is a summary of de-identification standards for two of the most prominent data protection statutes in the United States, HIPAA and GLBA, as compared against the de-identification standard under GDPR.

Regulation or Guidance	Citation	De-Identification Standard	Impact of Meeting De-Identification Standard
Health Insurance Portability and Accountability Act of 1996 (HIPAA)	45 CFR 164.514	Health information that does not identify an individual isn't within the scope of "individually identifiable health information." Use of "individually identifiable health information" first requires a determination of low risk of re-identification, which can be achieved by meeting either of two standards: 1) "Expert Determination," whereby an expert applies generally accepted statistical and scientific principles that support the low risk of re-identification; or 2) "Safe Harbor," whereby 18 identifiers outlined via statute are removed and there is no actual knowledge that residual information could identify an individual.	Exempt from HIPAA
Gramm-Leach-Bliley Act (GLBA)	16 CFR 313.3(o)(2)(ii)(B)	Personally identifiable financial information does not include information that does not identify a consumer, such as aggregate information or blind data that does not contain personal identifiers, such as account numbers, names, or addresses.	Exempt from GLBA
General Data Protection Regulation (GDPR)	Recital 26 not applicable to anonymous data	The principles of data protection do not apply to anonymous information, namely information that does not relate to an identified or identifiable natural person or to personal information rendered anonymous in such a manner that the data subject is not identifiable. The principles of data protection apply to information concerning identified or an identifiable natural person, which includes pseudonymized data. Determination of identifiability includes consideration of all means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. "Reasonably likely to be used" factors include the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing, as well as technological developments.	Anonymized data exempt from GDPR Pseudonymized data subject to relaxed GDPR requirements

Recommended best practices

To manage compliance with regulatory standards for de-identification of personal information, consider implementing the following best practices:

- Engage key stakeholders to understand the business objectives supported by collecting and processing personal information, the scope of data being collected, the intended use, and the current methods employed to de-identify that data.
- Identify regulations or other laws that govern the data processing activities, as well as any requirements for de-identification.
- Review contracts to understand rights and limitations for processing of personal information, including those related to de-identification.
- Assess currently applied practices to de-identify and use personal information against any regulatory and contractual restrictions and determine their suitability.
- Ensure the privacy program addresses when consent is needed for data processing activities and engage with technical teams to ensure opt-in/ opt-out consent mechanisms are properly built within applications, as well as tracked and managed internally.
- Identify “de-identification” as one of the intended purposes of collection in notices seeking consent.
- Work with stakeholders to build in a workflow process where individuals charged with the corporation’s compliance with privacy restrictions are notified of newly intended uses or collection of personal information (a “privacy by design” approach).

Allison Trimble



Associate Senior Counsel

DST Systems, Inc.

Soo Y. Kang



General Counsel and Director of the Consulting Division

Zasio Enterprises Inc.

Zasio Enterprises Inc. is a global leader in information governance offering technology and strategic consulting services.