Published on *The National Law Review* https://natlawreview.com

Federal Trade Commission Hashes Out Aggressive Interpretation of Data Anonymization: What You Need to Know

Article By:

Benjamin W. Perry

Lauren N. Watson

The Federal Trade Commission (FTC) has a long-standing habit of creating legal obligations through blog posts. Recent communications from the FTC by way of its Office of Technology Blog evidence an aggressive expectation regarding anonymization standards and whether hashing of information satisfies that standard. While the FTC's position in this regard is not new per se, it reflects an expectation of, and renewed focus on, compliance with a nearly impossible standard and presents challenges for businesses seeking to responsibly use personal data for research, marketing, and innovation purposes.

Quick Hits

- The FTC recently reaffirmed guidance issued in 2012 that takes the position that hashing, which is a process to convert data (such as your name or a password) into a string of characters and numbers to mask the original data, does not constitute "anonymization" of that data.
- To support that conclusion, the FTC also relies upon a standard for "anonymization" whereby "data is only anonymous when it can never be associated back to a person"—a potentially impossible result.

Background on Anonymization and Pseudonymization

The process of "de-identification" is an umbrella term involving the removal of personal identifiers from data to protect individuals' privacy. "Pseudonymization" or "de-identification" involves removing personal identifiers such that the data can no longer be attributed to a specific person without additional information. Pseudonymized data remains personal data because it can potentially be re-identified if additional information is available. There is no universal framework or singularly accepted definition for what constitutes "anonymization," which is a type of data de-identification, but anonymization is generally considered the most stringent and (theoretically) irreversible form of de-identification. This process aims to make re-identification impossible, even when the anonymized data is combined with other datasets.

Some scholars have posited that true “anonymization”—where the risk of re-identification is reduced to zero—is impossible in most or all contexts, especially with the advent of artificial intelligence tools that can quickly scan large datasets collected in different contexts to identify patterns. One famous historical example includes a search engine releasing, for research purposes, millions of ostensibly fully anonymized search queries with random numbers assigned for each user, which searches were able to be taken in context together to manually identify at least one of the specific individuals whose searches were part of the disclosed information. Remarkably, the individual was identified, even without using the automated technologies that have become increasingly sophisticated and publicly available in the years since.

Because of the practical implications associated with truly anonymizing data such that it can never be re-identified, regulators have generally offered specific methods or general guidance to define and accomplish that standard. To provide some context from other regulatory approaches, we will explore a few common jurisdiction- and industry-specific definitions before delving into the FTC’s recent position statement on anonymization.

De-identification and Pseudonymization Under HIPAA and GDPR

In the United States, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) has long provided the most concrete guidelines regarding how data can be considered “de-identified” for regulatory purposes. HIPAA offers two de-identification methods: the safe harbor method and the expert determination method.

- **Safe Harbor Method.** The safe harbor method involves the removal of eighteen specific identifiers from the dataset, and the entity must not have actual knowledge that the remaining information could be used to identify an individual, in addition to complying with other requirements such as the ongoing evaluation of the risks of re-identification.
- **Expert Determination Method.** The expert determination method requires an opinion from a qualified statistical expert indicating that the risk of re-identifying an individual from the de-identified data set is very small, using statistical or scientific principles and documenting the methods and results of the analysis.

Under each of these standards, the re-identification risk is never required to be zero. This is remarkable, as health data is generally understood to be among the most sensitive types of personal information. As such, if true anonymization were possible, there would arguably be no more important industry for anonymization standards than in healthcare. Despite this, there is no definition or standard for “anonymization” under HIPAA, perhaps reflecting an understanding that anonymization is virtually impossible (at least in the healthcare context).

By contrast, Europe’s standard set forth by Recital 26 of the General Data Protection Regulation (GDPR) provides that “personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.” On its face, this appears to require zero risk of re-identification. Notably, however, the GDPR recitals are not binding law, and none of GDPR’s articles define or otherwise further discuss anonymous data. Separately, GDPR expressly defines “pseudonymisation” in a manner that recognizes re-identification is possible in some limited circumstances: “‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.”

In April 2023, the Court of Justice of the European Union (CJEU) issued a landmark decision expanding what constitutes "anonymous" data such that it is exempt from GDPR. In that case, an independent assessor was engaged to evaluate whether shareholders and creditors would have been better off under normal insolvency proceedings, and in the course of this evaluation, written observations from shareholders and creditors, which were de-identified by replacing names with alphanumeric codes, were shared with the independent assessor. The CJEU found that, when determining whether pseudonymized data constitutes personal data, it is essential to consider the recipient's ability to re-identify the data subjects. If the recipient lacks the means to re-identify the individuals, the personal data can be considered anonymized for that recipient. In other words, it is no longer "personal data"—and therefore outside of the scope of GDPR—because it does not relate to an identifiable individual (at least for the recipient's purposes).

The FTC's Perspective on Hashing

On July 24, 2024, the FTC issued an article entitled, "No, hashing still doesn't make your data anonymous," reasserting the FTC's position from 2012—also clarified in a Technology Blog article—that hashing personal information does not render it anonymous. This position has significant ramifications for corporate data management and privacy representations, especially considering the varying definitions of "anonymous" data under different data privacy laws.

Hashing is a cryptographic process that transforms input data into a fixed-size string of characters. While the resulting hash appears random, it consistently produces the same output for a given input. A prime example of this involves biometric technologies, which usually operate by scanning the biometric identifier (such as an iris, fingerprint, or palm) and creating a hashed output consisting of numbers and/or letters intended to act as a mathematical representation of the biometric identifier. Hashing is also often used with website tracking technologies to convert certain pieces of personal information, such as email addresses, phone numbers, or device IDs, into a hashed string (such as "2813448ce6316cb70b38fa2?9c8c64130"). Although the output may appear random, subsequently inputting the same information will generate an identical output when using that same hashing algorithm.

Depending upon the length and complexity of the input, it is possible in some instances to retrieve the original input of hashed data if certain precautions are not taken to protect the hashed input. However, the potential reversibility of this process was not the focus of the FTC's analysis. Instead, the FTC pointed to its prior enforcement actions against companies where hashed information, such as email addresses, were shared with third party social media companies along with other sensitive data, such as health information. In those cases, the FTC alleged that the disclosing company knew the social media companies were able to re-identify the individuals associated with the hashed email addresses, and that the hashing was done solely to protect the information in the event of a data breach. It is unclear whether the disclosing company was disclosing the hash identifiers necessary to decrypt the hashed data or whether the social media companies were able to re-identify these individuals using other identifying information, such as IP address, device ID, or contextual clues. Regardless, the FTC concluded its discussion of relevant enforcement actions with a broad and troubling conclusion: "Regardless of what they look like, all user identifiers have the powerful capability to identify and track people over time, therefore **the opacity of an identifier cannot be an excuse for improper use or disclosure.**" (Emphasis in the original.)

Key Takeaways

The FTC has seemingly adopted an aggressive stance that data cannot be disclosed to third parties,

even using pseudonyms, or unique values intended to de-identify individuals in a dataset. Its policy statement comes at a time when the scope of many federal agencies' powers are in question following the Supreme Court of the United States' decision in *Loper Bright Enterprises v. Raimondo*, which drastically alters the extent to which courts must defer to federal agencies' interpretations of ambiguous laws that they are tasked with administering. The FTC's heightened focus on anonymization increases the risk associated with making public statements about whether and when a company may disclose "anonymized" data. This is particularly troubling given that global companies may be subject to numerous data privacy regimes, which means that a statement regarding anonymization may be accurate in one jurisdiction, such as the European Economic Area, but may constitute a misleading statement (according to the FTC) in the United States if there is any re-identification risk whatsoever. As always, the devil is in the details, and companies will have to assess re-identification risks on a case-by-case basis.

The potential consequences of mishandling or misrepresenting data practices are significant, potentially resulting in regulatory investigations, enforcement actions, and reputational damage. Vigilance and informed decision-making are essential in navigating this complex privacy landscape.

National Law Review, Volume XIV, Number 229