PUBLICATION

DOJ Final Rule Applies to Anonymized, Pseudonymized, and De-Identified Data: What Data Licensors Need to Know

Authors: Julie A. Kilgore, Alexandra P. Moylan, Alisa L. Chestler, Dandridge S. Parks
April 29, 2025

What's Changed?

The U.S. Department of Justice (DOJ) published a Data Security Program (DSP), pursuant to a final rule (Final Rule), which became effective on April 8, 2025. The DSP identifies prohibited and restricted transactions involving U.S. data access by countries of concern or by classes of covered persons. Unlike most privacy and data broker laws, the DSP does **not** exclude anonymized, pseudonymized, or de-identified data but rather expressly includes the foregoing within the definition of certain covered data. A previous alert provides additional detail on the classes of transactions, countries, persons, and covered data. This alert focuses on the inclusion of anonymized, pseudonymized, and de-identified data within the scope of covered data, the broad applicability of the DSP, and the potential impacts on such data moving forward.

Who's Feeling the Impact?

Data licensors and other entities selling, licensing, or otherwise providing access to anonymized, pseudonymized, or de-identified data and entities using such data to develop or train artificial intelligence tools. The inclusion of data that is anonymized, pseudonymized, or de-identified expands the applicability and impact of the DSP to entities who may generally be exempted from complying with obligations under other laws with respect to these categories of data.

Why Should You Care?

Violations of the DSP include both civil and criminal penalties, and the U.S. Attorney General has determined that the prohibited and restricted transactions, including those merely involving anonymized, pseudonymized, or de-identified data, pose unacceptable risks to the national security of the United States. A U.S. Federal Trade Commissioner has also recently stated that a priority of the FTC will be to work closely with the DOJ to enforce the DSP, so monitoring and investigation are likely in this area.

What's Your Next Move?

Assess data license agreements and other agreements to determine whether data covered by the DSP, including any anonymized, pseudonymized, or de-identified derivatives of that data (or artificial intelligence tools trained with such data) are implicated. Next, assess whether the impacted agreements constitute a prohibited or restricted transaction. Finally, assess whether there is either access by, or any restrictions within the agreements to limit access by, a country of concern, a covered person, or any foreign person. See our prior alerts related to next steps and guidance for additional compliance considerations.

Sensitive Data Now Includes Anonymized, Pseudonymized, and De-Identified Data:

Unlike most privacy and data broker laws to date, the DSP is not solely or primarily concerned about the identifiability of the covered data at the point of access. The DSP defines one category of covered data (*i.e.*, bulk U.S. sensitive personal data) to mean a "collection or set of sensitive personal data relating to U.S. persons, in any format, **regardless of whether the data is anonymized, pseudonymized, de-identified**, or encrypted, where such data meets or exceeds the applicable threshold set forth [within the defined term 'bulk']." In contrast, for example, under most state privacy and data broker laws, personal data or personal information that meets the requisite standard of anonymization, pseudonymization, or de-identification is also

exempt from the general requirements under the applicable law, except in limited circumstances (*e.g.*, basic contractual requirements). Additionally, under the Health Insurance Portability and Accountability Act and regulations promulgated thereunder (HIPAA), once covered data (PHI) is de-identified in accordance with the de-identification requirements set forth under HIPAA, the resulting data is no longer considered PHI and is exempt from most of HIPAA's requirements.

Broad Applicability:

The DSP can apply to agreements beyond data licensing or similar data sharing agreements. For example, for covered prohibited transactions, the DSP prohibits the provision of access to both the data and an artificial intelligence tool that is merely capable of providing access to anonymized, pseudonymized, or de-identified data, even if access to such data is not explicitly provided. The DSP contains the following example to demonstrate this prohibition's intended applicability (**emphasis** added):

A U.S. subsidiary of a company headquartered in a country of concern develops an artificial intelligence chatbot in the United States that is **trained** on the bulk U.S. sensitive personal data [including anonymized, pseudonymized, or de-identified data] of U.S. persons. While not its primary commercial use, the chatbot is **capable** of reproducing or otherwise disclosing the bulk U.S. sensitive personal health data that was used to train the chatbot when responding to queries. The U.S. subsidiary **knowingly** licenses subscription-based access to that chatbot worldwide, including to covered persons such as its parent entity. Although licensing use of the chatbot itself may not necessarily "**involve access**" to bulk U.S. sensitive personal data, the U.S. subsidiary knows or should know that the license **can be used to obtain access** to the U.S. persons' bulk sensitive personal training data if prompted. **The licensing of access to this bulk U.S. sensitive personal data is data brokerage** because it involves the transfer of data from the U.S. company (i.e., the provider) to licensees (i.e., the recipients), where the recipients did not collect or process the data directly from the individuals linked or linkable to the collected or processed data. **Even though the license did not explicitly provide access to the data, this is a prohibited transaction because the U.S. company knew or should have known that the use of the chatbot pursuant to the license could be used to obtain access to the training data, and because the U.S. company licensed the product to covered persons**.

Due to restrictions on using identifiable information to train artificial intelligence tools, many entities have turned to using anonymized, pseudonymized, and/or de-identified data for such purposes. Therefore, licensing of such tools should also be carefully evaluated for potential implication under the DSP.

Future Impact:

Because the DSP significantly deviates from the traditional approach with respect to anonymized, pseudonymized, and de-identified data, a key question that arises is whether this approach may be relied upon in future regulations or whether sensitive personal data will only be defined so broadly under laws addressing national security risks as opposed to those protecting against privacy risks. The commentary regarding the Final Rule contained a few nuggets that *may* preview changes to come, so entities regularly working with such data should continue to stay alert for future changes.

Intentional Inclusion:

The lack of an exclusion for anonymized, pseudonymized, and de-identified data was not an oversight but an intentional deviation from the approach other laws have taken previously. Within the commentary of the Final Rule, the identifiability of data was flagged as only one of many concerns the DSP aims to address. Specifically, the commentary stated, "anonymized data is rarely, if ever, truly anonymous, especially when anonymized data in one dataset can become identifiable when cross-referenced and layered on top of another anonymized dataset." Additionally, "[a]nonymized data itself can present a national security risk, as can pattern-of-life data and other insights that harm national security from anonymized data itself." Finally, "advances in technology, combined with access by countries of concern to large datasets, increasingly enable

countries of concern that access this data to re-identify or de-anonymize data, allowing them to reveal exploitable sensitive personal information on U.S. persons." Thus, the potential for re-identification highlights a key reason the DSP explicitly applies to anonymized, pseudonymized, and de-identified data, not only identifiable data. However, the potential for re-identification is not new or unique to national security risks when transacting with these categories of data. The risk of re-identification is also present in many de-identified data transactions, often with contractual requirements being one of the few mechanisms, if not the only mechanism, for preventing re-identification of the data. Only time will tell whether re-identification risks that potentially impact national security will amount to more strenuous protection for individuals or if similar restrictions will be input within other privacy and data broker laws moving forward.

De-Identification Standards:

The DSP includes restricted transactions that are permitted if entities comply with specified security requirements established by the Cybersecurity and Infrastructure Security Agency (CISA). The CISA data-level specifications generally require the implementation of mitigation techniques "sufficient to fully and effectively prevent access to covered data that is linkable, identifiable, unencrypted, or decryptable using commonly available technology by covered persons and/or countries of concern." While anonymization, pseudonymization, or de-identification of the data may be utilized in combination with other security requirements to achieve these specifications, all methods of such techniques permitted under other laws *may* not meet the CISA security requirements.

Commentary to the Final Rule suggests that the DOJ does not view all de-identification techniques as equal. The DOJ expressly confirmed its agreement with a commenter's recommendation to include de-identified PHI within covered data because "the HIPAA de-identification standards are out of date, and do not protect individuals in today's data-rich and computational-rich environment[...]" and the DSP should address "the ever-increasing ability to re-identify supposedly de-identified data." As a result, the DSP aims to strike a balance to allow restricted transactions that use robust anonymization, pseudonymization, or de-identification as specified by CISA's security requirements but prohibit the use of techniques that do not meet those standards. Therefore, for each restricted transaction, de-identification or similar techniques must be evaluated to determine if they meet CISA's data-level requirements. Again, it is unknown at this time whether the CISA standards or the criticisms of traditional techniques will be leveraged more broadly in the future.

For more information or assistance on this topic, please contact Julie A. Kilgore, Alexandra P. Moylan, CIPP/US, AIGP, Alisa L. Chestler, CIPP/US, QTE, Dan S. Parks, or another member of Baker Donelson's Data Protection, Privacy and Cybersecurity Team.