De-identification v. Anonymization

De-identification and anonymization are distinct data protection techniques. De-identification removes, deletes, or masks key personal identifiers or other unique attributes, such as client names or other fields, to reduce the risk of identifying individuals, organizations, or strategies. With de-identification, the data may still be discernable through other means, such as combining larger data sets or using codebooks, thereby retaining some re-identification risk.

Anonymization goes further. By permanently breaking the original dataset links, anonymization irreversibly alters data so identities and other key attributes cannot be ascertained, significantly reducing re-identification risk. Anonymized data is less granular than de-identified data, but data controllers and vendors can use it with fewer regulatory and legal risks because it can't be reassembled into personal data or other protected information.

Why it Matters. In AI service agreements and data protection addenda, de-identification prioritizes data integrity, while anonymization grants the vendor a broader range of uses for training and development. Organizations should evaluate whether data should be de-identified or anonymized and should always run tests to confirm threat actors cannot exploit it.

Amanda Cline
Senior Counsel