Anonymization & Pseudonymization as Tools for Cross-Border Discovery Compliance

By David R. Cohen

Introduction

Companies that conduct business internationally, and their lawyers, today face a significant challenge balancing U.S. discovery demands against the requirements of Europe's General Data Protection Regulation (GDPR). Most U.S. courts and investigators expect parties that are involved in litigation or investigations to comply with requests for potentially relevant documents in their possession, custody, or control, regardless of whether the documents are located within or outside the United States. However, the data privacy and data protection rules of many countries prohibit companies from transferring to the United States (or making accessible in the United States) documents containing personal information of persons within their countries ("data subjects"). Data protection laws such as the GDPR define "personal information" broadly, including any name, email address, physical address, or other information that allows identification of any data subject. Almost all cross-border documents that might be sought in litigation will contain at least some personal information. See the GDPR, Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [2016] OJ L119/1, 33.

Historically, when parties have objected to discovery requests seeking documents from Europe or other privacy-sensitive locations, most courts have not been sympathetic to such objections, partially because there have been few documented instances where companies have actually been sanctioned for violating data or document transfer rules to comply with U.S. discovery. See, inter alia, Laydon v. Mizuho Bank, Ltd., 183 F. Supp. 3d 409 (S.D.N.Y. 2016); Brightedge Techs., Inc. v. Searchmetrics, Case No. 14-cv-01009-WHO (MEJ) (N.D. Cal. Aug. 13, 2014); In re Xarelto (Rivaroxaban) Prods. Liab. Litig., MDL No. 2592 Section L (E.D. La. Jul. 20, 2016); St. Jude Med. S.C., Inc. v. Janssen-Counotte, 104 F. Supp. 3d 1150 (D. Or. 2015). That risk of sanctions is changing, however, now that European data protection authorities have "raised the stakes" on privacy law enforcement. Under the GDPR, which became effective May 25, 2018, a single violation can result in fines up to 20,000,000 €, or 4% of the offending company's worldwide annual revenue, whichever is greater. Thus, even a single enforcement action can potentially have a ruinous impact on a company. GDPR, OJ L119 at 82-83.

The GDPR, in particular, has many companies struggling to find effective methods of protecting personal data and minimizing the amount of data that needs to be transferred or reviewed. The GDPR requires that personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed." GDPR, OJ L119 at 35 (emphasis added). This language requires companies to review their data processing policies very carefully to ensure compliance.

This turn of events has many companies scrambling to determine how to cope with the conflict between U.S. discovery and international privacy rules. Fortunately, advances in technology may provide more efficient processes to help companies comply with the GDPR. For example, mass pseudonymization or anonymization of records prior to any reviews or cross-border transfers can be a part of the solution.

Pseudonymization & Anonymization

Pseudonymization refers to the practice of altering records to replace certain personal information with alternate information to protect the identity of data subjects. For example, every instance of "Susan Brown" in a document can be changed to the pseudonym "Kathy Williams," and every instance of "Frank Jones" can be changed to the pseudonym "John Doe." Anonymization commonly refers to the practice of altering records to entirely remove or redact all of the personal information therein. Note, however, that

even the complete removal or redaction of names would be considered "pseudonymization" under GDPR Recital 26 and the definition in Article 4(5), if the organization still had the ability to link back to the pre-altered documents with the personal information intact.

Pseudonymization and anonymization are not new data protection devices. Both practices have been in use at least since the advent of discovery proceedings involving confidential personal information, though perhaps not under those terms. If sufficiently pseudonymized records are transferred to the United States without anyone there having access to the original documents or original names, no personal information will actually be transferred. Indeed, pseudonymization is recognized in three separate recitals and in five separate articles of the GDPR, as one available/appropriate safeguard to protect personal data. See GDPR Recitals 28, 29, 156 and Articles 6(4)(e), 25(1), 32(1)(a), 40(2)(d) and 89. For example, Recital 28(1) provides: "The application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations."

Pseudonymization, however, does not solve the entire problem. Under GDPR Article 4(2), the definition of "processing" is so broad that even preservation, collection, pseudonymization, or anonymization of records constitutes "processing" that implicates privacy concerns. Nevertheless, discovery in legal proceedings in the United States can implicate "legitimate interests pursued by the controller or by a third party [...]" to justify such processing—especially since the result of such processing is the removal or masking of personal information prior to review or transfer. See GDPR, OJ L119 at 36.

Employing a Compliant Workflow

Pseudonymized or anonymized versions of the original documents may not be sufficient for all purposes in U.S. discovery. However, the acceptance by U.S. courts and regulators of pseudonymized or anonymized documents in the first instance would also help to show due regard for principles of comity, when balancing legal obligations of U.S. discovery proceedings and the privacy interests of international data subjects whose personal data may be implicated in a cross-border investigation. This would also be in line with the Sedona Conference International Investigations Principles (May 2017 Public Comment Version) pp. 22-23, available at https://thesedonaconference.org/publication/International%20Investigations%20Principles.

These techniques, especially when used with other strategies, can help to satisfy the main concern of data privacy authorities—protecting the privacy of data subjects. Many privacy officials in the EU, and other jurisdictions with similar privacy protections, do not understand the breadth of U.S. discovery. There is no counterpart to broad litigation discovery in most of continental Europe. To the extent that there is any pretrial disclosure based on requests from opposing parties, the disclosure is generally limited to specific documents that the opposing party already knows about and can identify with reasonable particularity.

Contrast that with the U.S. discovery system where millions of records may be produced to an opposing party in discovery, so that the requesting party may identify those relatively few documents that actually are important for resolving disputed issues. Typically only a small fraction of the documents produced are ever used for depositions, motions, or trials. The U.S. discovery process is anathema to most European privacy officials, who think that the U.S. legal system virtually ignores the privacy rights of individuals in allowing broad disclosure of thousands or millions of documents just to help find that small fraction that may truly be necessary to resolve the dispute.

Yet production of anonymized or pseudonymized documents should be sufficient in most litigation matters and investigations for a receiving party to rule out most of the documents as unimportant, and to identify a much smaller fraction that may be necessary to resolve the dispute. Then the parties can focus on only those documents truly needed; if any of those need to be transferred in their original form, the volume has already been minimized, and further measures (like protective orders) can be used to protect the privacy interests of any individuals identified in the transferred original documents.

For example, consider the hypothetical case of Class Action Plaintiffs vs. ABC Pharmaceutical Company. Assume that the plaintiffs are alleging that a drug manufactured by ABC causes adverse effects in patients, and they're further alleging that ABC did not adequately investigate or disclose those adverse effects when originally seeking approval for the drug. Assume also that some of the original drug development and testing occurred in Europe and that discovery requests are filed seeking production of all documents relating to the development and testing of the ABC drug. There could be tens of thousands of such records in Europe.

Rather than transferring all of those records to the United States for discovery, with the attendant GDPR risks of such a large data transfer, the company could first put the potentially relevant records through pseudonymization software that filters out all names and other identifying information and replaces them with "*" or "#" signs. Perhaps a small number of names of key individuals would be identified in advance so that those names could automatically be changed to pseudonyms instead of being completely removed. At that point the pseudonymized documents could be subject to review in the United States without transferring any of the original documents or personal information.

Trained reviewers in the United States can be told what to look for to identify critical documents, or protocols can be set up with the court or adverse parties for production of the pseudonymized documents in the first instance. Based on the review of those versions, parties can identify the relatively small number of documents—maybe only dozens or hundreds out of the original tens of thousands—that may be needed in unredacted form for the litigation. There will be a much stronger argument at that point that the legitimate interests in transferring that small number of truly necessary original documents outweigh residual individual privacy interests, especially if residual concerns are addressed through confidentiality agreements, protective orders, and other measures.

Leveraging Technology

These mass anonymization and pseudonymization efforts may previously have been impractical or cost-prohibitive due to the sheer cost of anonymizing or pseudonymizing large volumes of documents. Recent technological advances, though, including the use of artificial intelligence, have the potential to alleviate these impracticalities by allowing the mass automated anonymization and pseudonymization of very large volumes of electronic text-based documents. While this still does not solve all privacy issues—for example, certain kinds of records, like image-based records, cannot yet automatically be pseudonymized or anonymized in the same way—the vast majority of documents requested in most litigation or investigation matters are amenable to this new technology.

Next steps for the legal, privacy, and records communities include:

- Spreading awareness of the new technology;
- Developing protocols and processes to use these techniques in combination with other protective measures to maximize protection of personal information;
- Seeking input and acceptance from international data protection authorities; and
- Educating parties and courts in the United States about the need for flexibility and respect for international privacy norms when addressing cross-border discovery issues.

In combination with other steps, pseudonymization and anonymization can be extremely useful tools for helping to solve the once seemingly intractable conflict between U.S. discovery requirements and international data protection laws.

About the author:
David R. Cohen is a partner at Reed Smith LLP, where he leads the firm's global Records & E-Discovery (RED) Practice Group. Mr. Cohen also co-chairs the EDRM GDPR Committee and is a member of The Sedona Conference Working Group 6, which focuses on using technology to help address litigation and e-discovery challenges. The views expressed are those of the author, but he gratefully acknowledges the contributions to this article by members of his EDRM Committee and/or Sedona Working Group 6, including Rose Jones of King & Spalding, retired Magistrate Judge James C. Francis, IV, and Taylor Hoffman of Swiss Re.