research

Managing the Risks of Data Production in the Wake of the GDPR

Michael Becker & Xavier Diokno

Managing the Risks of Data Production in the Wake of the GDPR

In the United States, preserving potentially relevant information for litigation is the law of the land. With this focus on eDiscovery obligations, courts are typically less than forgiving when parties and their counsel withhold documents on the basis of data protection. Relying on foreign statutes to excuse the failure to produce data has often raised not only eyebrows but also suspicions about the withholding party's underlying motives.

But the recent, well-publicized implementation of the General Data Protection Regulation (GDPR) gives European data subjects far-reaching control over their personal data—a broad category that encompasses not just a person's name and birthdate but also demographic background, biometric data, health information, computer IP address and much more. Now, anticipating some pushback from U.S. judges, parties litigating cross-border matters must take steps to remediate the privacy issue before they appear in court.

Given the stakes involved in large, cross-border matters—U.S. discovery sanctions versus GDPR violations that can reach up to 4 percent of a company's global annual turnover or 20 million Euros, whichever is greater—companies must reconcile the requirements of the changing data privacy landscape with their eDiscovery obligations.

Fortunately, by using technology to reduce the amount of personal data retained and outsourcing to shepherd that data through the eDiscovery process, companies can affordably achieve workable solutions that serve both masters.

Using Technology to Minimize Data and Therefore Risks

Data minimization, one of the key tenets of the GDPR, requires organizations to collect and keep only as much data as is required to successfully accomplish a task. This means that organizations should take all necessary steps to limit the amount of personal data that they retain. This automatically reduces the amount of personal data that could be processed in the event of litigation. Fortunately, this reasonable step aligns with the need to limit discovery to relevant and responsive data.

Data-culling technology can help businesses satisfy their data-protection requirements while mitigating their production risk. There are two primary data-culling technologies at organizations' disposal. The first focuses on metadata, while the second uses artificial intelligence, specifically machine learning, techniques.

The first of these is nothing new: historically, organizations have used simple metadata culling techniques, such as filtering by custodians, date ranges, file types, file extensions and the like, to slice and dice their data. They then can apply search terms to documents and their metadata to further refine their results. This minimizes the occurrence of personal data such as names and email addresses within an eDiscovery collection.

In addition to these traditional tools, more advanced tools allow users to further reduce their stores of personal data. For example, with email threading, the application considers the metadata of both emails and their attachments, identifying messages with unique content. For instance, if two users exchanged 10 emails, the email threading application would group the emails, analyze the content of each email, and identify those emails that contain unique content, often times the last one in the series. In this

example, by limiting review to the last email, the content of all 10 emails can be reviewed in one email. These techniques can significantly reduce a data pool, and the less data you have, the lower the risk of inadvertently revealing private information.

Even more sophisticated review applications use machine learning techniques to analyze the actual text and substance of documents. These algorithms can organize documents in a given corpus according to not just keywords but also their actual conceptual content. For example, with predictive coding, a computer algorithm evaluates documents according to their core concepts, such as potential liability or contractual terms. To train the algorithm, senior document reviewers identify and submit exemplars of relevant documents. The algorithm then creates a ranked hierarchy of conceptually related documents based on their similarity to the manually reviewed exemplars, expediting review. By leveraging machine learning, the number of documents is reduced to those that are subject to the request, and in turn, reduces the risk of providing personal information.

Technology can also protect any lingering personal data in the documents that remain after culling. Instead of spending hours manually redacting employee information from documents, you can choose a review tool that searches for and automatically redacts certain programmable textual combinations, such as email addresses and employee identification or Social Security numbers. Anonymization techniques can permanently remove all personal data from a document: for example, you can prevent opposing parties from being able to identify individuals by anonymizing their phone numbers into a single business phone number. Finally, pseudonymization takes this one step further, removing all identifying data but maintaining the links between records that relate to the same individual.

Outsourcing eDiscovery to Further Mitigate Risks

When a pipe bursts in your home, you assess your home-repair abilities, the potential cost of making the repair and the risk of not fixing the source of the problem. If you lack the requisite expertise, your next steps are to turn off the water supply and Google a plumber.

The process should be much the same when it comes to handling eDiscovery: consider your expertise, estimate the cost of the project, and calculate the risk of missing data or performing a step in the EDRM incorrectly. Perhaps you've hired an IT person who has experience in managing discovery projects, and thus it makes more sense to keep some matters in house. But in most cases, IT staff lack the expertise required to assess and plan for the full panoply of legal risks, and organizations lack the manpower necessary to grapple with eDiscovery demands.

Either way, when you factor in the costs of potential violations—both of U.S. discovery obligations and of cross-border regulations like the GDPR—handling eDiscovery with limited internal resources can become incredibly expensive. With the leadership in most companies laser-focused on reducing costs, it's important to realize that outsourcing eDiscovery to an experienced provider can lead to significant cost savings—as well as fewer headaches and reduced risks.

First, from a cost-control perspective, it probably does make sense for organizations to take responsibility for the simpler tasks of eDiscovery: information management, data preservation and case management, for example. Low-level data filtering, such as the metadata filtering described above, can also be safely and economically performed in house.

But most businesses lack the dedicated discovery project managers, counsel and staff essential to performing more complex tasks—or at least to performing them in the most cost-effective, defensible ways. By outsourcing these higher-level tasks, the legal team—whether in house or external—can focus on strategic work like GDPR compliance without frittering away their bandwidth on work that can be completed using lower-cost resources.

A second consideration is scalability. Most organizations lack the internal resources required to capably handle a large eDiscovery project. Choosing a managed services provider enables you to have appropriate, trained resources at the ready whenever the need arises, without spending the money to maintain the full workforce necessary for substantial eDiscovery at all times.

Finally, think about the speed of change when it comes to eDiscovery technology. Bringing technology in house can seem like an excellent cost-cutting measure—that is, until the next iteration of your new software arrives and you have to upgrade your software and retrain your personnel on how to use it. Outdated technology is about as useful as no technology at all. Established eDiscovery providers are constantly evaluating new technology and adding it to their toolset. By maintaining the latest and greatest in review technology, they spare you the time-consuming learning curve for new technology or new iterations of existing technology. Moreover, they have the expertise to recommend the appropriate workflows that can end up saving you considerable effort—not to mention even more time and money.

Conclusion

The arrival of the GDPR makes the already expensive prospect of eDiscovery even more daunting. With the GDPR, your goal is to limit the amount of information that crosses borders. That means being as precise as possible with the information that you manage in eDiscovery to avoid realizing your worst fear: producing a document that falls outside the scope of the request and that contains protected personal data. Data-culling techniques, combined with the expertise of a seasoned eDiscovery provider, can help organizations balance their competing goals of satisfying U.S. discovery mandates and EU privacy demands.

info@consilio.com
www.consilio.com

Consilio + ADVANCED DISCOVERY®