Coalition of Technology Resources for Lawyers (CTRL)

WHITE PAPER

**White Paper** written by Osterman Research
Published **September 2019**
Sponsored by the **Coalition of Technology Resources for Lawyers (CTRL)**

# Big Data is Dead! Yet "Small" Data Isn't Ready for Primetime

Relativity®

***While taking some steps toward compliance, companies generally lack organizational and technological measures to satisfy new regulatory norms mandating minimization of personal data.***

# Executive Summary

Personal data, much of it unstructured in various data repositories on-premises and in the cloud, represents a major risk factor for organizations of all sizes. With increasingly rigorous privacy regulations like the European Union's General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA), and (effective January 1, 2020) the California Consumer Privacy Act (CCPA), organizations must properly manage their personal data stores that contain sensitive and confidential information. A key element of this management process is **data minimization** – retaining only data that is essential for a business purpose and in compliance with various regulations, legal directives, and best practices.

To better understand how organizations are minimizing their retention of personal data (or not), the **Coalition of Technology Resources for Lawyers (CTRL)** in partnership with Osterman Research sponsored an in-depth survey, which forms the basis of the results discussed herein.

***Decision makers lack insight into their data (particularly unstructured information).***

## KEY TAKEAWAYS

The following are the five key takeaways from the research that we conducted with 119 individuals knowledgeable about their organization's practices around the collection, storage and governance of personal data:

- **Managing personal data is key**
  Managing the enormous and growing volumes of sensitive and confidential personal data is essential for most organizations. Most decision makers recognize the risks that they face, at least conceptually, if not through practical action.

- **Organizations lack insight into their data**
  Decision makers lack insight into their data (particularly unstructured information), which means they are not taking the appropriate steps to protect their data or remediate problems as they should. In short, decision makers are not practically addressing the significant risks that they face.

- **Data hygiene is often reactive, not proactive**
  Data hygiene, on those occasions when it is performed, tends to be more reactive and episodic than proactive and continuous. The result is that information managers and others are taking a "band-aid" approach to serious problems that should be getting much more forward-thinking attention.

- **Regulatory and privacy obligations are surprisingly not key drivers**
  Few organizations are addressing their management of personal data in a proactive manner in compliance with the various requirements to which these organizations are subject. For example, ephemeral messaging technologies could advance data minimization objectives while safeguarding personal data. Nevertheless, few organizations seem to have implemented or even understand the capabilities of these technologies.

- **There is a fear of spoliation**
  One of the reasons that personal data is not remediated more aggressively appears to be fear of data spoliation. However, it's entirely possible that many decision makers are not willing to tackle the more difficult problems associated with remediating personal data and are using spoliation as a pretext to defer taking actionable steps to minimize personal data.

# Dealing with Personal Data

## A MANDATE TO MINIMIZE RETENTION OF PERSONAL DATA

We discovered that nearly three in five organizations (58 percent) have a corporate mandate to minimize the retention of personal data. Moreover, we also found that the vast majority (91 percent) of organizations limit their collection of personal data for a variety of reasons as discussed later in this report. As a result, most organizations *claim* to have implemented some form of data minimization in the context of how they collect and store personal data.

## WHAT DOES "DATA MINIMIZATION" REALLY MEAN?

The term "data minimization" means different things to different decision makers and it can have more than one meaning. For example, as shown in Figure 1, 70 percent of those surveyed consider that data minimization means to prevent the collection of personal data that is not necessary in the context of their specific business objectives. Similarly, 64 percent view data minimization as preventing the collection of personal data for which the business has no legitimate business purpose in its retention. However, slightly more than one-half of decision makers view data minimization as the process of deleting unnecessary information – something more akin to defensible deletion after the fact - than a proactive prevention of collecting personal data in the first instance.

**Figure 1**
**"When you hear the term 'data minimization', what does that mean to your organization?"**
Percentage of Organizations

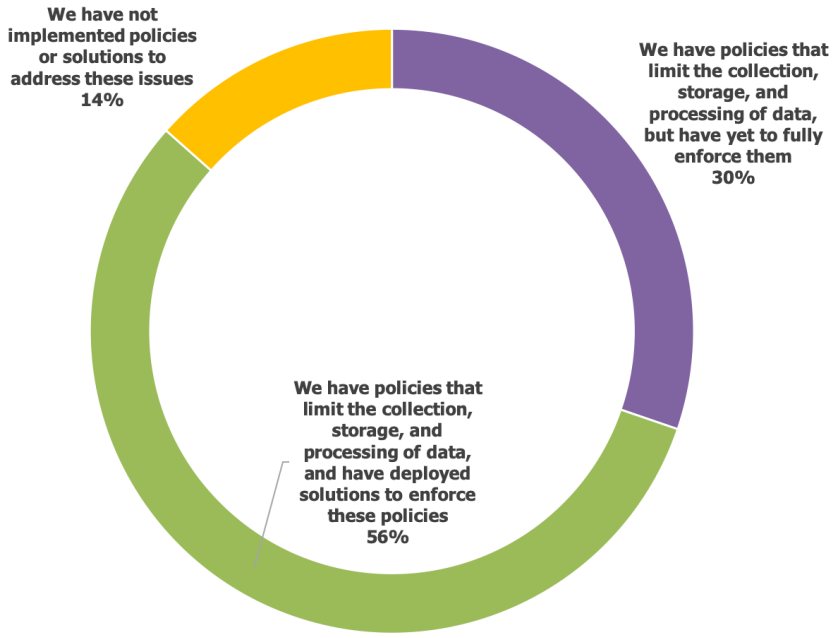

***The term "data minimization" means different things to different decision makers and it can have more than one meaning.***

Among organizations that do not have a corporate mandate to minimize the retention of personal data, two-thirds of those entities lack regulatory or business drivers that require them to do so. Some claim to not have a mandate because they may want or need to use personal data sometime in the future.

## MOST LIMIT THEIR COLLECTION OF PERSONAL DATA

But what does "data minimization" really mean as a practical matter? We asked organizations to describe how they collect, store and process personal data. As shown in Figure 2, what we found is that the majority (56 percent) have policies that limit the collection, storage, and processing of data, and they have deployed solutions to enforce these policies. However, we also found that 30 percent of those organizations do not yet fully enforce the policies they have in place. Another 14 percent of organizations have no policies or solutions to address their obligations around personal data.

**Figure 2**
**"Which of the following best describes your organization's collection, storage and processing of personal data?"**



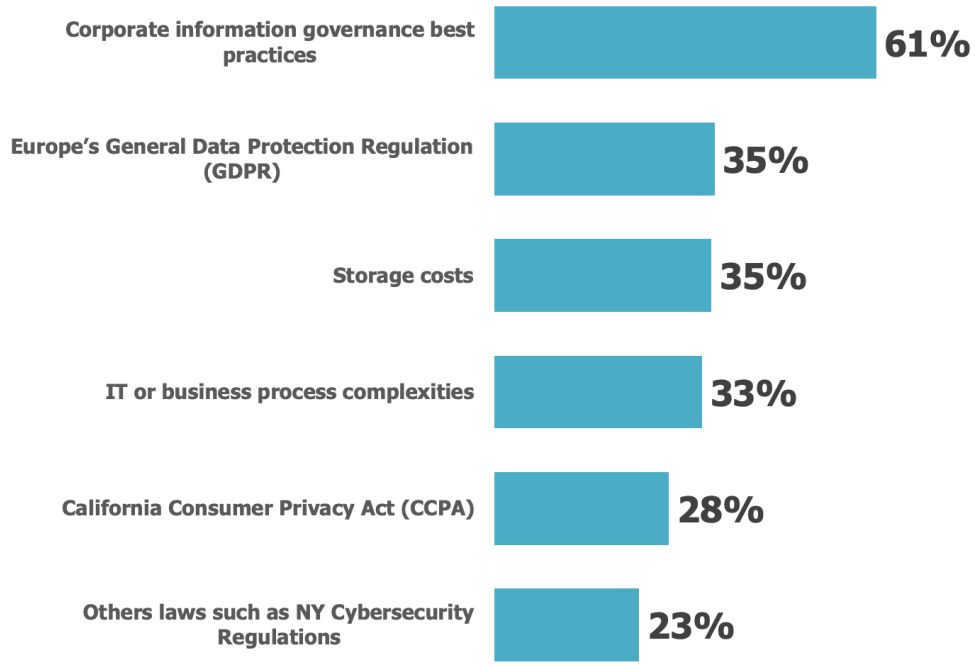

***Not all organizations have policies or solutions in place to address personal data obligations.***

## THE DRIVERS FOR DATA MINIMIZATION

The drivers for data minimization will vary widely depending on a number of factors, including the jurisdictions in which an organization operates, the specific regulations that it must satisfy, the types of data that it possesses and controls, the risk tolerance of its senior management and legal counsel, and so forth.

Our research, as shown in Figure 3, shows that corporate information governance best practices clearly dominate the drivers for minimizing the retention of personal data, with 61 percent of survey respondents citing this as an "important" or "major" factor in their decision process. Storage costs, IT needs, and business process complexities seem to be more significant drivers of data minimization than regulations like the GDPR or the CCPA.

**Figure 3**
**Drivers for Minimizing the Retention of Personal Data**
Percentage Responding an "Important" or "Major" Factor



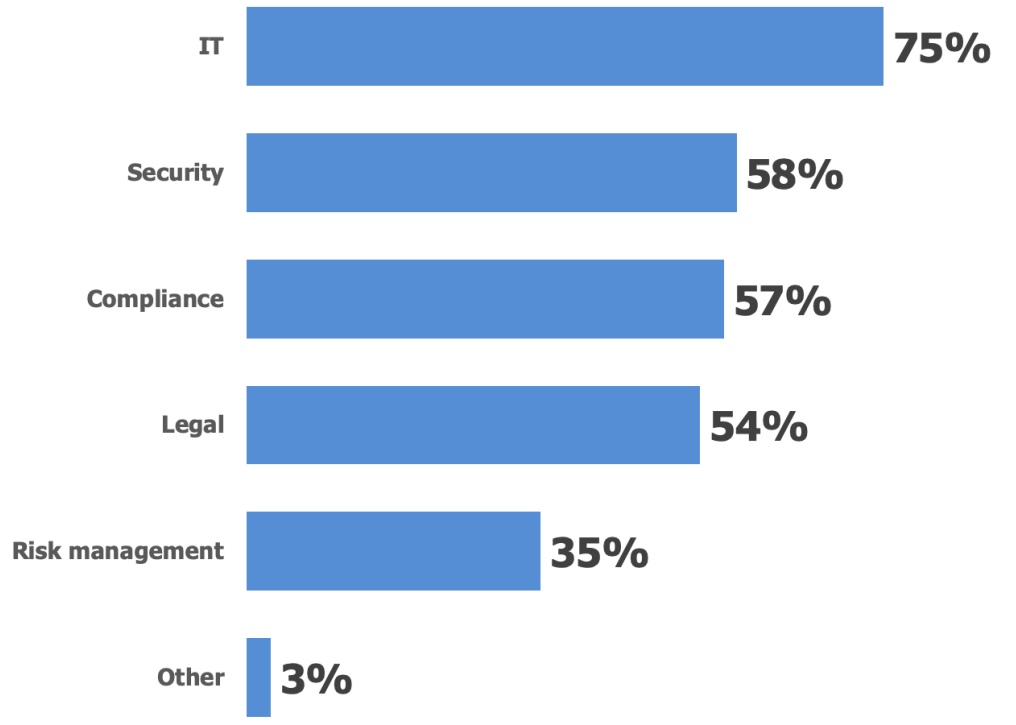

***The drivers for data minimization will vary widely depending on a number of factors.***

## IT IS LEADING THE MINIMIZATION MANDATE

As shown in Figure 4, IT is leading the charge for their organizations' data minimization mandate, with 75 percent of organizations citing IT as responsible for issuing (and not just implementing) the mandate. Other groups involved include security, compliance, legal and risk management.

The fundamental problem with IT issuing any sort of data minimization mandate is that minimizing the retention of data is not a primary focus of IT. The principal drivers for data minimization are focused on reducing the risks associated with retaining personal data or other sensitive information in violation of privacy regulations like the GDPR or other legal considerations. Obviously, it's important that IT be involved in the data minimization process from a technical standpoint, but the stakeholders who own compliance, legal or line-of-business management should in most cases be primarily responsible for *issuing* such a mandate.

**Figure 4**
**Groups Responsible for Issuing the Data Minimization Mandate**
Percentage of Organizations

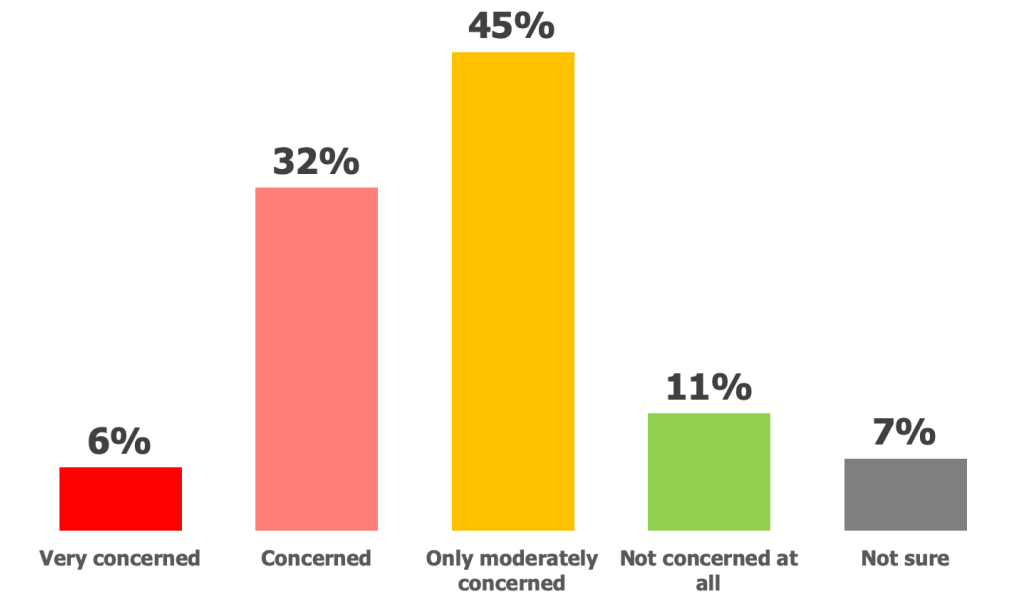



***The fundamental problem with IT issuing any sort of data minimization mandate is that minimizing the retention of data is primarily not an IT issue.***

## WHAT ARE THE DANGERS OF DATA MINIMIZATION?

Are there any dangers associated with data minimization? Yes, the primary one being deleting data that might later be required for a regulatory, legal or best practice purposes. When survey respondents were asked to what extent they are concerned that data minimization might result in spoliation of data, over eighty percent expressed concern over this possibility. In addition, well over a third of respondents confirmed they were either "concerned" or "very concerned" about this possibility, as shown in Figure 5.

**Figure 5**
**"To what extent are you concerned that by minimizing your data, your company may be more exposed to spoliation risks?"**



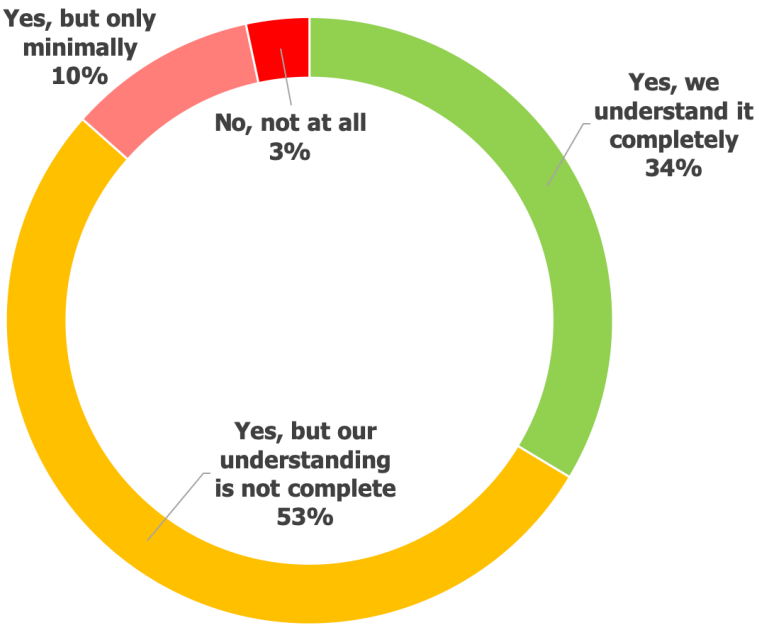

***Are there any dangers associated with data minimization?***

# Understanding Corporate Data

## MOST DECISION MAKERS DON'T FULLY UNDERSTAND THEIR DATA

Only one-third of corporate decision makers understand their corporate data in the context of any risky, sensitive or personal data that it might contain, as shown in Figure 6. Our research disturbingly found that more than one-half of decision makers do not have a full understanding of their corporate data, while about one in eight have only a minimal or no understanding of their data. This widespread lack of understanding of stored corporate data can lead to several negative consequences. Among them are poor information governance and an increased risk of data breaches given the organization's inability to understand how to protect key data assets.

**Figure 6**
**"Are you able to understand your data universe to determine if it contains any risky, sensitive or personal data?"**




***Only one-third of corporate decision makers understand their corporate data in the context of any risky, sensitive or personal data that it might contain.***

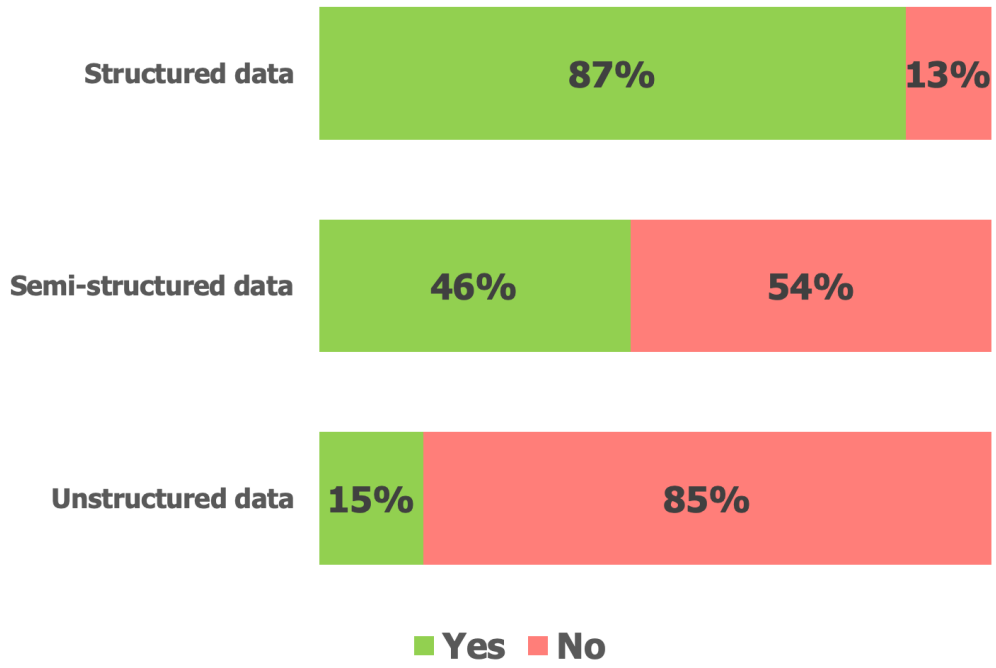
## DATA-MAPPING IS NOT COMMON

Most organizations have not undertaken a data-mapping project to facilitate compliance with any data minimization mandates that might be in place, with only 43 percent of the organizations surveyed having completed any sort of data-mapping initiative.

### MOST ORGANIZATIONS DON'T HAVE AN INVENTORY OF THEIR UNSTRUCTURED DATA

Unstructured data is truly "dark" for most organizations. As shown in Figure 7, our research revealed that the vast majority of organizations have a comprehensive data inventory for their structured data – data contained in corporate databases and the like. Yet, significantly fewer (only one in seven) have a comprehensive data inventory for unstructured data.

**Figure 7**
**"For which of the following do you have a comprehensive data inventory?"**




*Most of the "breachable" data organizations possess is unstructured.*

The fact that so few organizations have a comprehensive data inventory for their unstructured data creates an enormous risk for these organizations. This is because most of the "breachable" data organizations possess is unstructured in the form of email messages, text messages, documents, video, audio and a wide range of other data types. Conservatively, at least 80 percent of corporate data is of the unstructured variety. The fact that so few organizations have a handle on their inventory of this content makes their management of personal data very risky. For example, these organizations typically won't have a sufficient understanding of what data types they have, what their data stores contain, whether or not sensitive or confidential data exists in their data stores, whether or not data that should be encrypted is actually encrypted, or a good map of the locations in which their data is stored.

## Problems in Managing Personal Data

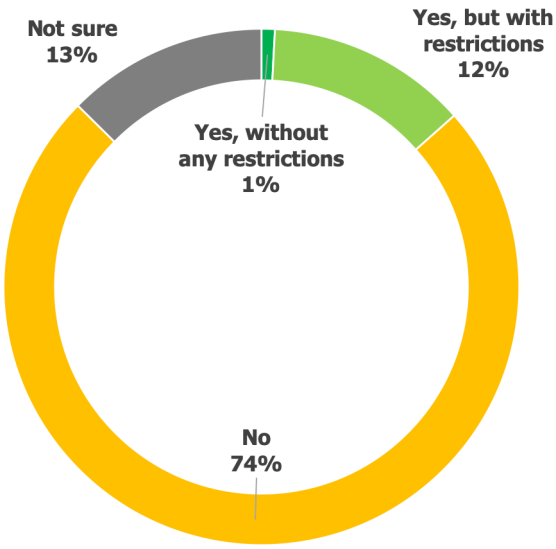
Our research found that one in 11 organizations (nine percent) are not limiting their collection of personal data – in fact, these organizations are collecting everything with the hopes of using it at some future date. That creates an untenable situation for personal data management. More troubling, it can create a storage nightmare that makes regulatory compliance and eDiscovery more difficult and expensive.

## TECHNOLOGY SOLUTIONS ARE NOT WIDELY USED

Among the 58 percent of organizations that have a corporate mandate to minimize the retention of personal data, technology solutions are not widely deployed. For example, 76 percent of organizations with a minimization mandate will conduct periodic clean-up initiatives, 58 percent employ records management systems, and 43 percent use structured databases with enabled expiration functionality. However, only six percent of organizations that have a mandate to minimize the retention of personal data have officially deployed ephemeral messaging for communications to address their personal data management mandate.

And yet, across all of the organizations surveyed, only one in eight (13 percent) are using ephemeral messaging, in most cases with restrictions, as shown in Figure 8. The vast majority of organizations are not using ephemeral messaging, and another 13 percent of those surveyed are not sure if it is used. Among the most commonly mentioned ephemeral messaging solutions in use are Gmail Confidential, Confide and Telegram.

**Figure 8**
**"Is ephemeral messaging being used in your organization?"**




***Many organizations are not taking proactive steps to address their data management risks.***

# Summary

Limiting the retention of personal data is essential if organizations are going to successfully mitigate the risks and costs associated with data breaches, regulatory violations, and legal/regulatory events. However, a substantial number of organizations are not limiting their retention of personal data, nor are they using technologies that would be useful in mitigating risks, such as ephemeral messaging. In short, many organizations are not taking proactive steps to address their data management risks, despite the critical need to do so. Given the sheer volume of regulatory initiatives that are on the horizon, this untenable situation is likely to be short-lived.

# About CTRL

CTRL is an industry forum dedicated to advancing the discussion on the use of technology and analytics in the practice of law. CTRL was born out of the idea that practical guidance and open source collaboration have been missing from the eDiscovery and Information Governance space. The time has come for "the rest of us" to discover better solutions through collaboration and active dialog. CTRL believes that through the open exchange of information and best practices, we can improve the practice of law with technology. Our resources are available for public consumption and comment, and we hope that you will react to the ideas offered by this group. CTRL is meant to provide a laboratory for practical experimentation in the hopes of discovering new and better solutions.