NAVIGANT

VedderPrice

GLOBAL LEGAL TECHNOLOGY SOLUTIONS

CREATING A DATA INVENTORY: THE FIRST STEP IN MANAGING PRIVACY AND DATA SECURITY RISK

By David Manek, Bruce A. Radke, and Michael J. Waters

As high-profile data system breaches continue to make headlines, businesses and public agencies of all types are focusing with increasing intensity on data privacy and security concerns. Upgrading information security systems, developing incident response plans, conducting penetration testing, and carrying out various other information governance, privacy, and data management activities can be both time-consuming and expensive. They also can put significant strains on limited IT resources. Yet they are essential in today's information-driven organizations, where the theft, loss, or unauthorized exposure of sensitive information can inflict devastating consequences in terms of reputational risk, increased compliance and regulatory costs, and direct financial losses.

Unfortunately, the effectiveness of data privacy and security initiatives is often greatly diminished if the organization fails to take a fundamental and critical first step: developing and maintaining an accurate, comprehensive, and up-to-date data map. In addition, as regulatory agencies increase their scrutiny of data security issues, the ability to identify and locate sensitive data quickly becomes even more critical – which means effective data mapping is an essential compliance and regulatory response tool.

DATA MAPPING BASICS

There are a variety of precise, technical definitions of a data map, but at the most fundamental level, a data map is a document that tells you where certain types of data are located within your organization and how that data can be accessed. Put another way, a data map is a listing of what information and data you have, where it is located, and who is responsible for managing it.

A data map typically provides both a data inventory and data flow diagrams that depict how data moves through the organization. These data flow diagrams can take various forms such as Visio diagrams or other graphic representations, as well as spreadsheets that list all the various types of data included in the map, and then map and cross-reference the various types to depict how and by whom each data type can be accessed. Graphic representations are useful for helping decision-makers visualize data flows and connections, but spreadsheet-based or matrix-based data inventories are often more information-rich and practical to end users. The terms data map and data inventory are used interchangeably throughout the article; however, one should envision a matrix-based data inventory as the ultimate goal deliverable.

DATA INVENTORY DRIVERS – WHY IT MATTERS

The specific features and characteristics of the data inventory and the particular types of data to be mapped must be aligned to each organization's business needs and customized to its specific legal and regulatory environment. There are typically four business drivers that prompt an organization to create a data map:

1. **Legal, regulatory and privacy compliance** – A data map is used to identify specific types of information that must be tracked and reported under the requirements of various laws, regulations, or professional standards. Examples include the Gramm-Leach-Bliley Act , the Health Insurance Portability and Accountability Act (HIPAA), and the Health Information Technology for Economic and Clinical Health Act, as well as data requirements imposed by organizations such as the Financial Industry Regulatory Authority or the International Organization for Standardization. In addition, several state governments, such as Oregon and Massachusetts, require businesses that retain personal information of state residents to develop and implement a written information security program, which includes certain minimum administrative technical and physical safeguards. Up-to-date data mapping is an essential component of such programs.

   Rapidly approaching is the General Data Protection Regulation (GDPR) where the European Union (EU) intends to strengthen data protection of data collected from EU data subjects. GDPR will be enforced starting in May of 2018 and extends to organizations collecting personal information from EU data subjects regardless of where the organization is headquartered or hosting the data. Under GDPR, U.S.-based companies collecting personal information from EU citizens may be forced to demonstrate compliance with GDPR. According Article 83 of GDPR, penalties for noncompliance of GDPR can reach 20 million euros or 4 percent of total revenue, whichever is higher. To demonstrate compliance with GDPR, organizations must track a large number of data points such as which systems in the organization contain personal data collected from EU citizens, the purpose of the data, if consent was given by the data subject, and what security protocols are in place to protect the personal data. Building and maintaining a data inventory is a fundamental building block to demonstrating compliance with GDPR.

2. **Intellectual Property (IP) Risk** – A data inventory is essential for identifying and tracking an organization's financial data, intellectual property and trade secrets as well as other sensitive data such as personally identifiable information or protected health information. In addition to helping organizations apply required access controls and security safeguards, a data map can also be helpful in demonstrating compliance with those controls, and showing that it has adequately implemented necessary safeguards to protect sensitive data from potential risk in the various repositories where that data resides.

3. **E-discovery** – A data map can greatly simplify and streamline responses to e-discovery requests in association with legal actions, as well as responses to other discovery motions and subpoenas. Corporate legal departments can leverage the components of a data inventory to identify which systems could be relevant to a new case. Legal departments can use the data inventory as a menu to select which systems they want collected and preserved in response to pending litigation. In public sector organizations, a data map is also extremely helpful in replying to Freedom of Information Act requests, first by streamlining the process of determining whether the requested information exists, and second by pinpointing where and in what format it is stored. Knowing where electronically stored information resides in an organization's systems can not only accelerate an organization's response to such requests, it can greatly reduce the time and resources that are required to comply.

4. **Data Management, Retention and Disposition Policies** – A data map is an essential first step in establishing effective data management policies that prioritize what data an organization must or should retain and the resources that will be needed to safeguard that information. You need to know what information you have and where that information resides in order to dispose of records when their retention periods expire. The process of developing the data map can be helpful itself – in our experience, management is almost always surprised to discover just how much information the organization is retaining.

BREACH RESPONSE – WHEN A DATA MAP IS VITAL

While the various compliance and business drivers discussed thus far clearly demonstrate the benefits of maintaining up-to-date data maps, the most crucial and urgent need for data mapping relates to a situation that has become all too familiar to many organizations: a successful breach that leads to the exposure of sensitive information, outright theft of proprietary data, or a ransomware attack in which data is held hostage, threatening to shut down mission-critical operations. When such a breach occurs, having an accurate and up-to-date data map is vital.

Consider, for example, how your organization would respond if you discover that confidential company documents have been posted online. The documents might contain personally identifiable information (PII), protected health information (PHI), or other types of nonpublic personal information (NPI). In such a scenario – which is unfortunately becoming commonplace – the first step is to try to get the information removed from the online source. But that is only a Band-Aid – there is no way to determine how far and how quickly the leaked information might have been spread already.

The second – and equally important – step is to determine where the information came from. Was it accessed by an external attacker or was it leaked from an internal source? Or did the leak occur from a third party that had authorized access to the data but failed to protect it?

To determine the source of the breach, you must first determine where the data resides. Is it stored locally on internal systems only, or is it made available to others? Who has access to the information? What safeguards were in place to protect the data, and why did they not work? What steps must be taken immediately to contain the damage and prevent the exposure of additional information?

All of these questions must be answered quickly, not only to minimize the damage but also to comply with various notification requirements – both to regulators and to potentially affected individuals. Many state and federal data breach notification laws require notice of the breach to affected individuals and regulators within a short amount of time. Without an accurate and up-to-date data map, your organization must resort to one of three costly and complex options to comply with these requirements:

- **Option 1:** Install and implement an enterprise software solution that allows you to search your infrastructure by document name or other characteristics. Unfortunately, there is no guarantee such a solution will identify all relevant documents. These programs' effectiveness depends on whether they can index the far corners of your infrastructure. Moreover, they cannot index loose devices such as mobile devices, or flash drives, nor can they index data that is stored by third parties. The time required to identify, select, and install such software is another significant problem, especially since timely response is critical.
- **Option 2:** If you have some idea what servers or devices were compromised, you can hire a consulting firm to create an image of the device, process it, and scan the information, looking for the specific document or other nonpublic information. While this approach is faster than the first option, it nevertheless is time-consuming and, like enterprise software solutions, it still can fail to address all sources or devices.
- **Option 3:** If the first two options fail, you must resort to the third option – a manual review or "treasure hunt." This involves laboriously combing through the entire organization trying to identify who had access to the exposed documents and where the PII was stored. When deadlines are tight and potential penalties are mounting, such a process is not only laborious and costly, it also is prone to potential errors and gaps, and consumes an inordinate amount of resources that would be better devoted to mitigating future risk.

Obviously, knowing what systems contain PII, PHI, NPI or other sensitive data can dramatically accelerate incident investigation and containment. A comprehensive and accurate data map makes it possible to you respond to a breach quickly, efficiently and systematically.

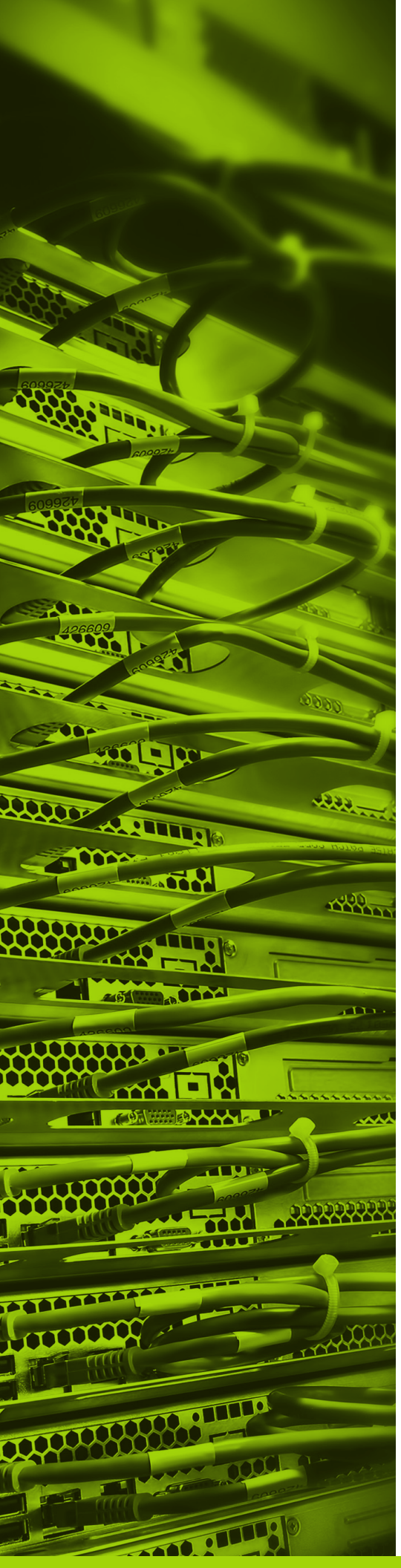
In addition, the process of developing the data map can help minimize the amount of sensitive information that is stored, reducing both the risk of a breach occurring and the extent of the damage if one does occur. Data mapping also helps support the important practice of deleting data when it is no longer needed for business purposes, thus helping to reduce potential regulator concerns over the unnecessary retention and storage of sensitive information.

BUILDING A DATA INVENTORY STEP-BY-STEP

Before examining current data mapping best practices, it can be helpful to step back and take a high-level view of the process of building a data map. Broadly speaking, the process encompasses four general steps:

- **Step 1: Define the scope and other business requirements.** Is the map going to be enterprise-wide, or will it focus only on certain departments or functions? What data repositories will be represented in the data map, and what data elements within those repositories will be collected and maintained?

  In addition to defining the overall scope of the data map, it is also important to reach consensus on other related parameters, such as the taxonomy that will be used to organize and categorize the repositories, and the tools, formats, and capabilities that will be used to collect and maintain the data map.
- **Step 2: Develop the data map design and framework.** Begin by identifying who will be the users of the data map, both today and in the future, and get their input and buy-in. Key stakeholders will likely include the legal, compliance, and risk management functions, as well as IT. Their input is critical in order to define the acceptable level of detail and accuracy. Too many fields and discrete data points will only clutter the process and make it difficult to maintain the map. But too few data elements can cause the map to be inadequate.
- **Step 3: Populate the data map.** Identify what information is stored and where that information resides. This is the heart of the process, and generally is accomplished using in-person interviews, online surveys, or a combination of the two. Online survey tools can be very effective when they are designed in a way that enables business owners to respond to very specific questions about the information they require and retain.



- **Step 4: Develop a process for maintaining the data map.** Most organizations develop new systems and new business processes on a regular basis, which means they also are regularly acquiring new types of information. This makes it essential that the data map be regularly updated and improved as well. Quarterly maintenance of the data map is a good minimum requirement, along with a full-scale audit and update on an annual basis. In addition to developing a systematic process for these updates, it also will be necessary to identify appropriate contacts, tools and data sources, and train responsible staff on maintenance and upgrade procedures.

GETTING STARTED WITH DATA MAPPING

The importance of achieving clarity on scope and other requirements at the outset cannot be overstated. Unfortunately, this step can be easily overlooked – and often is. Many data mapping projects have failed because the organization rushed into the effort without taking the time to first reach a consensus on critical questions of definition and scope.

For organizations that are developing their first data inventory or that are relatively new to the data inventory process, it often is helpful to begin with an initial pilot that focuses first on high-risk records or specific types of information such as PII, PHI, or other NPI. Business systems that are known to contain this type of information would rank higher in priority than a general ledger accounting system, for example. Later phases can then expand the scope to encompass the organization's financial data, intellectual property and trade secrets, as well. An alternative approach is to carry out the pilot program in a single business unit only, and then expand to other segments in later phases.

Starting the data inventory process with an initial pilot that is limited in scope is also a good practice for smaller organizations with more limited resources, where it is important to identify those areas where they will achieve the best return on their initial investment. A pilot program also allows you to work out problems in the data collection process with less impact on resources. In addition, it can help you build momentum and achieve some quick wins, which in turn will generate greater buy-in and support from the affected departments or individuals.

In addition to defining the overall scope of the data map, it is also important to carefully consider what data repositories will be covered. For example, are you going to cover servers, work stations, and databases only? Or will the data map also encompass online sharing sites, laptops, smartphones and other mobile devices? Will the focus be limited to electronic records and data only, or will it cover both electronic and hard copy records? Often the initial assumption is that the data map will cover electronic data only, but entities that are covered by HIPAA generally need to conduct a comprehensive risk analysis of all sources of PHI, including paper.

Initial scope questions such as these are areas where organizations are most likely to have missteps. Coming to an appropriate answer is a matter of balance. For example, many times organizations decide to cover every possible data repository and end up never completing the data map at all. On the other hand, focusing too narrowly can mean you miss critical repositories and end up with a map that is inadequate. Again, often the most effective approach is to focus initially on those repositories that contain critical information, and then expand to other repositories that contain less critical information as a follow-up phase.

BEST PRACTICES AND NEXT STEPS

Another area of concern involves understanding the specific reporting and formatting requirements of the regulatory agencies to which your organization is subject. For example, when the Consumer Financial Protection Bureau issues a civil investigative demand letter, it typically requires production of transactional data within a very short timeframe, and generally accepts transaction data production in certain formats, including SQL backup files; Microsoft Access files; XML, CSV and TSV files; and sometimes Microsoft Excel files. Most modern business systems can create SQL backup files in a very short timeframe, but with some older mainframe systems the time required to convert data into an acceptable format can be substantial.

In terms of overall best practices, that National Institute of Standards and Technology (NIST) has published useful guidance in this area. One NIST publication that is particularly on point regarding data mapping is, "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII): Recommendations of the National Institute of Standards and Technology."[1] Although it was published in 2010, its guidance is still very relevant to today's users.

Regarding the actual content of the data map, the amount of information it contains can range from the bare minimum – such as designating a subject matter expert (SME) for each affected system and a description of the nature of the data it contains – to very mature data maps that address a wide range of parameters, including:

- SME for each system
- Nature of the data
- Format of the system
- Retention policy
- Backup procedures and frequency
- Identification of all PII residing in the system
- Confirmation that the use, collection and retention of PII is limited to that which is strictly necessary
- Categorization of all PII by its confidentiality impact and its value to the organization
- Application of appropriate safeguards for PII
- Annual review with appropriate executives, security officers and legal counsel

As noted earlier, the information in a data map can be presented in various ways. Most people picture a data flow diagram or entity relationship diagram. Those are good exhibits for executive use when an event occurs, because they present a graphic roadmap for getting the incident response process started. On the other hand, the level of management that normally would be reviewing a data map for compliance or regulatory response purposes might find a familiar column-based spreadsheet format to be more practical.

Regardless of the format and structure that is eventually chosen for the data map, the most important point is to get started on the process of scoping and developing the data map. A good place to start is by having a conversation with your IT team to identify what types of data inventory might already be in place.

If your organization has not yet begun this process, the time to begin is now. If you already have data mapping in place or under development, it is important to regularly review and update your data map to ensure it remains current with your entity's ever-changing data environment. With high-profile data system breaches continuing to make headlines, and with regulatory agencies focusing even more intently on privacy and data security issues, a comprehensive and accurate data map is an indispensable tool in helping to identify, locate, and manage sensitive personal and organizational data.

David Manek, CPA, CFF, CFE, CAMS, CBIP, is a director and national leader of Navigant's Technology Solutions Structured Data practice. Dave's team focuses on the collection, analysis and production of large enterprise data systems used to support litigation and regulatory investigations. Dave and the rest of Navigant's Structured Data practice are regularly building and maintaining information-rich data inventories for their clients. He can be reached at 312.583.6841, or at dmanek@navigant.com.

Bruce A. Radke, J.D., and Michael J. Waters, J.D., are shareholders in the Chicago office of Vedder Price and co-chair of the firm's Privacy, CyberSecurity & Media practice group. Bruce and Mike regularly counsel private- and public-sector clients on various privacy and data security issues and have, along with the other Vedder Price attorneys, assisted clients in responding to numerous significant data security incidents. Bruce can be reached at 312.609.7689, or at bradke@vedderprice.com. Mike can be reached at 312.609.7726, or at mwaters@vedderprice.com.

1. http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf

CONTACTS

DAVID MANEK

Director
312.583.6841
dmanek@navigant.com

BRUCE A. RADKE

Shareholder
312.609.7689
bradke@vedderprice.com

MICHAEL J. WATERS

Shareholder
312.609.7726
mwaters@vedderprice.com

navigant.com

About Navigant

Navigant Consulting, Inc. (NYSE: NCI) is a specialized, global professional services firm that helps clients take control of their future. Navigant's professionals apply deep industry knowledge, substantive technical expertise, and an enterprising approach to help clients build, manage and/or protect their business interests. With a focus on markets and clients facing transformational change and significant regulatory or legal pressures, the Firm primarily serves clients in the healthcare, energy and financial services industries. Across a range of advisory, consulting, outsourcing, and technology/analytics services, Navigant's practitioners bring sharp insight that pinpoints opportunities and delivers powerful results. More information about Navigant can be found at navigant.com.

linkedin.com/company/navigant

twitter.com/navigant