Washington and Lee Law Review Online

Volume 73 | Issue 2 Article 11

5-4-2017

Data Flow Maps—Increasing Data Processing Transparency and Privacy Compliance in the Enterprise

Jeremy Berkowitz
Deloitte & Touche LLP

Michael Mangold
Deloitte & Touche LLP

Stephen Sharon
Deloitte & Touche LLP

Data Flow Maps—Increasing Data Processing Transparency and Privacy Compliance in the Enterprise*

Jeremy Berkowitz**

Michael Mangold***

Stephen Sharon****

Abstract

In recent years, well-known cyber breaches have placed growing pressure on organizations to implement proper privacy and data protection standards. Attacks involving the theft of

** Jeremy Berkowitz, J.D., is a Senior Consultant in Deloitte & Touche, LLP's Assurance Practice and specializes in Cybersecurity and Privacy and Data Protection. Jeremy is a member of the Maryland State Bar Association and a Certified Information Privacy Professional (CIPP).

*** Michael Mangold, J.D., is a Manager in Deloitte & Touche, LLP's Cyber Risk Services Practice and specializes in Privacy and Data Protection, providing strategic guidance on data governance and developing global compliance programs. Michael is a member of the Minnesota State Bar Association, a Certified Information Privacy Professional (CIPP/US), and Certified Information Systems Security Professional (CISSP).

**** Stephen Sharon, J.D., is a Manager in Deloitte & Touche, LLP's Cyber Risk Services Practice and specializes in Privacy and Data Protection where he balances his time between performing privacy and security risk assessments, designing global privacy programs, and providing guidance on cross-border data transfers. Stephen is a member of the bars of New York and New Jersey, holds the Certified Information Privacy Professional (CIPP/US) certification, and enjoys teaching privacy courses.

employee and customer personal information have damaged the reputations of well-known brands, resulting in significant financial costs. As a result, governments across the globe are actively examining and strengthening laws to better protect the personal data of its citizens. The General Data Protection Regulation (GDPR) updates European privacy law with an array of provisions that better protect consumers and require organizations to focus on accounting for privacy in their business processes through "privacy-by-design" and "privacy by default" principles. In the US, the National Privacy Research Strategy (NPRS), makes several recommendations that reinforce the need for organizations to better protect data.

In response to these rapid developments in privacy compliance, data flow mapping has emerged as a valuable tool. Data flow mapping depicts the flow of data through a system or process, enumerating specific data elements handled, while identifying the risks at different stages of the data lifecycle.

This Article explains the critical features of a data flow map and discusses how mapping may improve the transparency of the data lifecycle, while recognizing the limitations in building out data flow maps and the difficulties of maintaining updated maps. The Article then explores how data flow mapping may support data collection, transfer, storage, and destruction practices pursuant to various privacy regulations. Finally, a hypothetical case study is presented to show how data flow mapping was used by an organization to stay compliant with privacy rules and to improve the transparency of information flows.

Table of Contents

I. Introduction

In recent years, well-known cyber breaches have placed growing pressure on organizations to implement proper privacy and data protection standards.[1] Attacks involving the theft of employee and customer personal information have damaged the reputations of well-known brands, resulting in financial costs for incident remediation and compensation of individuals affected.[2]

---

1. *See Verizon 2015 Data Breach Investigations Report Finds Cyberthreats Are Increasing in Sophistication*, VERIZON (Apr. 15, 2015), http://www.verizon.com/about/news/2015-verizon-dbir-report-security/ (last visited Apr. 24, 2017) ("This data reaffirms the need for organizations to make security a high priority when rolling out next-generation intelligent devices.") (on file with the Washington and Lee Law Review). There were more than 2,100 confirmed breaches and 80,000 reported security incidents in 2015, involving over 700 million records. *Id.*

2. *See* PONEMON INST., 2011 COST OF DATA BREACH STUDY: UNITED STATES 15, http://www.ponemon.org/local/upload/file/2011_US_CODB_FINAL_5.pdf ("Such costs typically include help desk activities, inbound communications, special investigative activities, remediation activities, legal expenditures, product discounts, identity protection services and regulatory interventions."). The average cost to an organization of a data breach was $5.5 million, with 37% of that cost from lost business. *Id.*

Additionally, businesses are collecting and transferring more data across state and international borders than ever before, due to continuously growing business activity happening via the Internet and constant expansion into international markets. As a result, governments across the globe are actively examining and strengthening laws to better protect the personal data of their citizens, whether that data is processed by private or governmental entities, local or abroad.[3] Most prominently, the General Data Protection Regulation (GDPR), passed by the European Union in the spring of 2016, updates European privacy law with an array of provisions that better protect consumers and require organizations to focus on accounting for privacy in their business processes through "privacy-by-design" and "privacy by default" principles.[4] The National Privacy Research Strategy (NPRS), released by the Obama Administration's National Science and Technology Council in June 2016, makes several recommendations that reinforce the need for organizations to better protect data, particularly (1) "increasing the transparency of data collection, use, transfer, and retention," and (2) assuring "that information flows and use are consistent with privacy rules."[5]

Additionally, guidance from several government agencies in the last few years has encouraged a change in mindset away from viewing privacy strictly as a compliance mechanism and toward privacy as a method to improve business processes. The Office of Management and Budget released an updated version of Circular A-130 in July 2016 where it noted that "agencies manage information systems in a way that addresses and mitigates

---

3. *See, e.g.*, Tom Geller, *In Privacy Law, It's U.S. vs. the World*, COMM. ACM, Feb. 2016, at 21–23 (discussing data protection laws internationally).

4. Regulation 2016/679 of the European Parliament and of the Council of the European Union on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 2016 O.J. L. 119 [hereinafter GDPR], *available at* http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&qid=1490558317324&from=en (last visited Apr. 24, 2017) (on file with the Washington and Lee Law Review).

5. NAT'L SCI. & TECH. COUNCIL, NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT PROGRAM 1 (2016), https://www.nitrd.gov/PUBS/NationalPrivacyResearchStrategy.pdf [hereinafter NPRS].

security and privacy risks associated with new information technologies and new information processing capabilities."[6] As a result, organizations are advised to demonstrate that they adequately protect the personal data they collect, use, transfer, and retain, so as to maintain profitability and good relations with their customers.

In response to these rapid developments in privacy compliance, data flow mapping[7] has emerged as a valuable tool by (1) increasing the confidence of compliance with relevant privacy laws, (2) providing a record of how data flows through an organization for employees and customers, (3) identifying problems when there are issues with data processing, and (4) mitigating potential risks and challenges related to the handling of personal data. More specifically, data flow mapping, among other tasks:

- Depicts the flow of data through a system or process, often using visual representations of how the data moves from system-to-system and user-to-user,
- Enumerates specific data elements handled by the process or system being mapped, including notation for different types of sensitive information,
- Documents technical and procedures controls, such as whether data is encrypted,
- Identifies the risks at different stages of the data lifecycle, and
- Identifies unused and underutilized system sources allowing organizations to make systems more efficient and save costs.

This Article will accomplish several goals, in conjunction with showing how data flow mapping achieves the NPRS objectives listed above:

---

6. Tony Scott et al., *Managing Federal Information as a Strategic Resource*, WHITE HOUSE BLOG (July 27, 2016, 9:00 AM), https://www.whitehouse.gov/blog/2016/07/26/managing-federal-information-strategic-resource (last visited Apr. 24, 2017) (on file with the Washington and Lee Law Review).

7. *See generally* PRIVACY TECH. ASSISTANCE CTR., MAPPING DATA FLOWS – CHECKLIST (2015), http://ptac.ed.gov/sites/default/files/Mapping_Data_Flows_Checklist_final.pdf (providing an overview of data flow mapping and "recommend[ing] them as a best practice when confronted with complex data scenarios")

- Explains the critical features of a data flow map and how they contribute in advising organizations to understand the information collected.[8]
- Discusses how mapping may improve the transparency of the data lifecycle in an organization, and the challenges in receiving personally identifiable information (PII).[9]
- Recognizes the limitations in building out data flow maps and the difficulties of maintaining updated views of data flows and data lifecycles without a defined, tactical revision process.[10]
- Discusses how data flow mapping may support data collection, transfer, storage, and destruction practices pursuant to privacy regulations. It will discuss how mapping provides an efficient method for distinguishing anonymous and sensitive data, and showing what parts of the organization are responsible for processing it.[11]
- Shows a hypothetical case study where data flow mapping was used by an organization to stay complaint with privacy rules and improve the transparency of information flows. These examples will also demonstrate the scalability of data flow mapping, making it applicable for organizations within different industries and sizes.[12]

II. Features of a Data Flow Map

Data flow maps may follow a variety of models, forms and structures, include different degrees of textual and graphic representation, and may be developed manually (through interviews, questionnaires, and the manual drafting of diagrams), through technology (via the automated scanning of systems, databases, networks and other technology to identify data and

8. *Infra* Part II.
9. *Infra* Part III.
10. *Infra* Parts II.E; III.C.
11. *Infra* Part IV.
12. *Infra* Part V.

detail a business process), or a combination of the two.[13] Each method can be used to fully address an organization's goal and compliance requirements, as different solutions yield different benefits.[14] Regardless of the organization's adopted approach, effective data flow maps exhibit the following five features that enhance their overall impact and increase the likelihood of achieving the objectives defined above.

A. Feature 1: Tracking the Data Lifecycle

Maps should be aligned to the lifecycle of data. Organizations often confuse data flow maps with a system, asset, and data inventory. Such inventories identify and enumerate the systems, databases, and types of personal information processed, either in a graphical or text-based format. Usually, these inventories also show the relationships between systems and databases, and may show the sources of the personal information processed. This method often fails to account for unstructured data (typically free-form data that is difficult to categorize, such as e-mail, images, text files, presentations), leaving a gap between the documented inventory and the full scope of information handled.

Conversely, drafting data flow maps by data lifecycles allows organizations to document the collection, use, transfer, and archival/destruction activities of the in-scope data. Regardless of the sources of data (e.g., a free-form e-mail request from a customer), the storage methods (e.g., a filing cabinet), or various uses (e.g., presentations, research), tracking the lifecycle of data is likely to capture both structured and unstructured data, providing a more complete overview of the data processed within organizations.

---

13. *See, e.g.*, MAPPING DATA FLOWS – CHECKLIST, *supra* note 7, at 1–2 (providing an infographic recommending steps to ensure data is handled properly).

14. An organization should research the available methods (including consideration of varying skill sets amongst their staff), analyze the costs associated and select the option that adequately addresses their goals. Each method can be implemented to include the components discussed in this Part.

B. Feature 2: Tracking by Purpose

Data flow maps should track a given business process. As discussed further in Part IV,[15] one of the core benefits of data flow mapping is to clearly document the uses of data. Therefore, leading mapping practices typically include text-based narratives to explain the roles of given systems, users, and third parties. These narratives will enable non-technical audiences to read and understand a map, which will increase the usefulness of the mapping across the organization. For example, if a non-technical process owner understands a data flow map, she may be able to alert the map owner when processing activities change, decreasing the likelihood that the map becomes stale or obsolete. Further, improving accessibility by non-technical audiences allows for data flow maps to educate leadership responsible for funding security and privacy initiatives. Finally, if done correctly, these narratives may address a number of record-keeping requirements.[16]

C. Feature 3: Identify the Types of Data Handled

Data flow maps should note the types of data handled at each stage of activity, particularly when sensitive information (or other types of data necessitating heightened standards of care) is present. In some cases, business processes use sensitive information for only a part of a given process, meaning that the risk changes throughout. For example, assume a customer provides their contact information (e.g., name, billing address, shipping address) and credit card information for an online retail transaction. It is possible that the contact information may be used separately from the credit card information because, per Payment Card Industry guidance and leading security practices, credit card information should be used only as needed for the

---

15. *See infra* Part IV (discussing "requirements that address data presence, data flow, and data use).

16. *See* text accompanying *infra* note 36 (noting that " there are a number of US and international laws where data flow maps can help achieve compliance").

processing activity in question,[17] while other information, such as an email address, may be used for unrelated marketing purposes at a later date. Data flow maps that only track the information handled at the macro-level likely communicate that this entire retail transaction, from order creation through payment processing to shipment and delivery, uses credit card information, raising the risk level for all steps of the process (even those handling routine contact information only).

On the other hand, more granular data flow maps that specifically identify the types of data collected at order creation, versus the data used in payment processing and data used for shipment and delivery, will delineate the steps of the order that are specifically handling sensitive information (e.g., credit card information during payment) from the steps handling lower risk personal information (e.g., contact information to facilitate shipment and delivery). This allows organizations to tailor their understanding of the risk profile, compliance requirements, and leading security practices to the business process step-by-step, rather than applying the strictest standards to the entire process. Organizations may then implement an efficient and affordable path to compliance and protection based on their data flow maps.

D. Feature 4: Document Risks and Controls

In support of a more efficient path to compliance, the fourth element of effective data flow maps includes specific notation of privacy and security controls at each step of the business process. Maps should clearly show where technical controls, such as encryption, are implemented to protect data-in-use and data-at-rest. Further, data flow maps should be clear where required privacy controls (such as cross-border data transfer mechanisms) are implemented to support compliance with global regulatory obligations. Finally, internal and external resources, including people and systems, should be easily identifiable. For example, a cloud-based server hosted by a third-party vendor may appear in

17. *See* PCI SECURITY STANDARDS COUNCIL, PCI DATA STORAGE: UNDERSTANDING DO'S AND DON'TS 2 (2008), https://www.pcisecuritystandards.org/pdfs/pci_fs_data_storage.pdf ("Do not store cardholder data unless there is a legitimate business need.").

a map with a different color than internal servers so as to clearly and easily distinguish between resources for which the organization is responsible, and resources for which third-party activities and obligations should be carefully reviewed. Where such controls are required by leading practice or by legal requirement but are not present, the data flow map should also identify this as a risk, setting the stage for discussions around remediation or mitigation.

E. Feature 5: Proactively Develop a Maintenance Plan

Finally, data flow maps are prone to becoming outdated without a maintenance plan. Data flow maps can be time-intensive to properly and thoroughly document. When completed, however, ongoing maintenance can be simple and straightforward. Many organizations fail to make a plan for ongoing maintenance of the maps and, once recognized as being outdated, data flow maps may require more (or similar) effort to update as it would to simply develop a new map from scratch. To keep data flow maps current, many organizations make two designations at the time a map is created: a data flow map owner, and a process owner—sometimes the same person. The process owner is responsible for embedding flags in the process that notify the data flow map owner of changes to the flow of data within that process. The data flow map owner is then responsible for identifying the extent of changes, reviewing the changes with relevant subject matter specialists, and updating the data flow map to be consistent with the changes. In some cases, technology solutions may automate maintenance or, at a minimum, alert the data flow map owner that changes in the process or systems may require changes to the map.

F. Conclusion

The features discussed above demonstrate the critical role data flow maps serve in connecting organizations' information flows with their business processes. Parts III and IV will show how these same features of data flow maps can play a critical role in achieving NPRS objectives around transparency and

compliance.[18] Data flow mapping, by accomplishing both of these objectives, will also help organizations continue to use big data in new, innovative ways in the future.

III. NPRS Objective 3.4: Increase Transparency of Data Collection, Sharing, Use, and Retention

Data flow mapping can play an important role in increasing the transparency of an organization's data lifecycles, and addressing NPRS concerns regarding consumers' awareness of how their personal data is collected and processed.[19] The NPRS repeatedly emphasizes the importance of transparency for organizations to explain their information flows. This maintains the trust of their employees and customers by ensuring privacy rights are protected.[20] Fully understanding the data lifecycle in all facets of business also contributes to effective risk management.

A. Current Challenges Organizations Face in Understanding Data Lifecycles

The amount of data collected and processed has overwhelmed organizations' information systems in recent years.[21] This inundation of data is instigated by expansion of business as well

---

18. *See generally* NPRS, *supra* note 5 (discussing these objectives).

19. *See id.* at 14 (noting that individuals are often "unaware of when data about them is collected or for what purposes it will be used" and "often do not understand the extent to which data about them is shared with third parties").

20. *See id.* ("Research designed to increase transparency of data collection and use would enable individuals to better evaluate the privacy implications and potential benefits of their activities and would permit data collectors/users to develop data practices that respect and protect individuals' privacy desires.").

21. *See* Brian Lee, *Do You Really Have Big Data or Just Too Much Data*, INFO. WEEK (Sept. 12, 2016, 11:06 AM), http://www.informationweek.com/big-data/big-data-analytics/do-you-really-have-big-data-or-just-too-much-data/a/d-id/1326867 (last visited Apr. 24, 2017) ("There is more data available to organizations today than ever before. In 2015 alone, customers, employees, and other users created about 7.9 zettabytes of data globally—and that number is expected to reach 35 zettabytes in 2020.") (on file with the Washington and Lee Law Review).

as the avenues and methods for organizations to collect it.[22] Big data has become its own industry,[23] enabling organizations to gain new customers, analyze business trends, determine problems, and expand operations. As a result, organizations continue to collect more data through electronic means with the perspective that they can find a use that leads to commercial advantages.[24] Additionally, they engage more third parties to both collect and process data, resulting in a larger data ecosystem with more actors playing various roles.[25]

Further, the storage medium of data being handled presents unique challenges. Structured data is information contained in readable forms with pre-defined hierarchy and element attributes, such as that in a database.[26] Unstructured data, defined in Part II as free form data that difficult to categorize,[27] includes data found in documents, e-mail, and MP3 files. Studies show that unstructured data comprises about 80 percent of the information that organizations collect.[28] While organizations have

---

22. *See id.* ("Advances in technology, computer power, and analytics mean companies can collect and process data in almost real-time.").

23. *See* Sylvain Magdinier et al., *Too Much Information: How Big Data Is Changing Legal and Commercial Risk Management*, ACC DOCKET, September 2015, at 27 ("'Big data will be transformative in every sphere of life.' This is not a slogan promoting a Silicon Valley start-up, but the White House's assessment published in May 2014.").

24. *See id.* (observing that "information is becoming the customer's crown jewels—not just valuable or sensitive, but a core commercial asset").

25. *See* Nick Ismail, *Will Overconfidence Kill Big Data?*, INFO. AGE (Mar. 9, 2017), www.information-age.com/will-overconfidence-kill-big-data-123464912/ (last visited Apr. 24, 2017) ("61% of respondents reported that their organisation uses third-party consultants with big data expertise.") (on file with the Washington and Lee Law Review).

26. *See Introduction to Structured Data*, GOOGLE DEVELOPERS, https://developers.google.com/search/docs/guides/intro-structured-data (last updated Feb. 26, 2017) (last visited Apr. 24, 2017) ("*Structured data* refers to kinds of data with a high level of organization, such as information in a relational database.") (on file with the Washington and Lee Law Review).

27. *See supra* Part II (discussing unstructured data).

28. Seth Grimes, *Unstructured Data and the 80 Percent Rule*, BREAKTHROUGH ANALYSIS (Aug. 1, 2008), https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/ (last visited Apr. 24, 2017) ("It's a truism that 80 percent of business-relevant information originates in unstructured form, primarily text. . . . There are variations; Anant Jhingran of IBM Research, among others, cites an 85% figure. Whether 80 or 85 percent, the claim . . . has been repeated

an easier time accounting for structured data due to its pre-defined hierarchy, the growing amount of dynamic unstructured data collected makes it more difficult to inventory information, hindering efforts to understand the data an organization possesses, and how that data are used.

The increased appetite for information and its varying storage mediums has made the task of tracking personal data more challenging. Greater public attention to data breaches and more laws requiring organizations to understand where their data is coming from and how it is used further highlight this issue. Many organizations have policies publically stating that they strive to protect personal data collected.[29] Many policies include detailed language that dictates how the organization collects personal data, whether data subjects have the right to review and amend personal data, and notice/consent provisions detailing other rights regarding their personal data.[30] These policies, while important, reaffirm the need for data flow maps to graphically depict how these polices and controls are operationalized within an organization. Data flow maps can also show how these policies connect an organization's purpose for collecting data and how it is actually used on a day-to-day basis.

Organizations also face challenges implementing security controls that can enforce limits on which users have access to certain data. Given the responsibilities and tasks of different departments, some organizations may allow all employees access to any information they need, while other organizations may choose to limit access on an as-needed basis. Organizations may also differ with regard to who has access to stored data, either in physical or electronic form. They may also differ in terms of the justification required for employees to access stored data. Furthermore, third parties may have unfettered access to information beyond what is necessary for their responsibilities.

---

many thousands of times.") (on file with the Washington and Lee Law Review).

29. *See, e.g.*, *Data Policy*, FACEBOOK, https://www.facebook.com/full_data_use_policy (last updated Sept. 29, 2017) (last visited Apr. 24, 2017) ("We work hard to protect your account using teams of engineers, automated systems, and advanced technology such as encryption and machine learning.") (on file with the Washington and Lee Law Review).

30. *See, e.g.*, *id.* (informing users what information Facebook collects and how such information is used and shared).

These different requirements make it more challenging for employers to keep track of data within an organization, as well as for consumers to understand who may have access to their data, particularly years after it is collected.

Finally, many organizations struggle to maintain proper security controls to ensure data does not properly contain "data creep."[31] Data creep occurs when sensitive information, intended to be confined to a well-secured location, spreads to other databases and systems. While the initial database may have enhanced security, other systems may not and can potentially become a weak link in the chain.

B. How Data Flow Mapping Addresses Current Challenges

Data flow mapping plays a large role in helping organizations address the problems laid out above, and improving the transparency of data lifecycles. Data flow mapping addresses current challenges by: (1) identifying and categorizing the types of structured and unstructured data it collects, (2) explaining the content and purpose of how data is used, (3) providing information about which individuals such data is shared with, and (4) explaining the security controls that organizations use to protect the personal data in their environments.[32]

First, data flow mapping allows organizations to enhance their data governance tools by developing a comprehensive understanding of how personal data within their environment interacts with their data lifecycles. Data flow maps show how organizations collect, store, transfer, and destroy personal data, and can be directly leveraged to develop approaches to processing. Unlike data asset inventories, data flow maps will capture unstructured data, in addition to other analog uses. This provides organizations a greater holistic view of their data lifecycles, allowing them to better safeguard their personal data

---

31. *See* Joel R. Reidenberg, *Resolving Conflicting International Data Privacy Rules in Cyberspace*, 52 STAN. L. REV. 1315, 1371 n.32 (2000) (defining "data creep" as "the tendency to continually expand the scope of collection and use of personal information").

32. *See supra* Part II (explaining the goals and functions of data flow mapping).

and take action on where it may need to be better protected. It can also enable organizations to offer better insight to consumers on how their data is processed, particularly after the initial collection phase.

Data flow mapping also helps connect organizations' data elements with their business processes and objectives. It identifies the purpose and context of data that is collected and stored by organizations, explains how it is categorized, and explains how it is used by each entity of organizations. The categorization of data is particularly important for new legal requirements such as the GDPR which requires organizations to maintain records on how such information is processed.[33] Hence, mapping can play a strategic role in ensuring that legal and regulatory requirements are met, further lowering organizations' risks. It also helps organizations periodically assess that they are collecting information tailored to purposes for growing their business, while not putting themselves or their consumers at unnecessary risk.

Data flow mapping also identifies information that is shared with stakeholders, both internally and externally. This is particularly important with more organizations engaging in the use of third parties to process data.[34] Additionally, many US organizations are striving to comply with the EU-U.S. Privacy Shield so they can continue to easily transfer data between the United States and European entities.[35] Data flow mapping can show how this data is transferred, what parties are involved in

---

33. *See* GDPR, *supra* note 4 (providing that data controllers must maintain a record of processing activities including the purpose of the processing, a description of the categories of data, the categories of recipients of the data, whether the recipients are in other countries—and if necessary the safeguards in place—and the intended duration the data will be stored).

34. *See supra* note 25 and accompanying text (noting the increasing use of third-party data processors).

35. *See* Press Release, European Comm., Restoring Trust in Transatlantic Data Flows Through Strong Safeguards: European Commission Presents EU-U.S. Privacy Shield (Feb. 29, 2016) [hereinafter Privacy Shield], *available at* http://europa.eu/rapid/press-release_IP-16-433_en.htm (last visited Apr. 24, 2017) ("Once adopted, the Commission's adequacy finding establishes that the safeguards provided when data are transferred under the new EU-U.S. Privacy Shield are equivalent to data protection standards in the EU.") (on file with the Washington and Lee Law Review).

the transfer process, and the types of processing that third parties are conducting.

Finally, data flow mapping provides more transparency around the security controls used to protect data. Data flow maps permit stakeholders to identify risks in their systems based on data coming in and proper safeguards (or lack thereof) being applied, as well as controls to ensure that sensitive data does not unnecessarily flow into other areas. Subsequently, it permits stakeholders to prioritize risks to better manage problems around data. It can permit an organization's teams to efficiently collaborate in response to ongoing organizational and legal changes. For example, the release of a customer-facing mobile app may raise legal issues. However, legal review and approval will likely be expedited if all parties have a clear understanding of (1) how data collected and made available via the app is integrated into existing systems and (2) the security measures in place for such systems.

C. Limitations of Data Flow Mapping

In spite of the benefits discussed above, data flow mapping does have some limitations. Developing an original map is a lengthy process for an organization. Even with an extensive investment of time and resources, it still may not completely gain all required information on data processes. As a result, organizations have to weigh the costs of undertaking it. Past experience has shown that data flow maps are particularly useful for high risk processes and for tracking the most sensitive data. It will not always be possible for organizations to talk to everyone, or to resolve conflicts between departments on where and how personal data is collected. Organizations are also regularly changing the types of data they collect, the processes by which they collect, store, transfer, and destroy data, as well as the external parties they use to assist themselves with these tasks. Without processes in place that can periodically and automatically update information, maps can quickly become outdated.

Some organizations have employed the use of technology to create data flow maps. However, this technology often faces

challenges capturing unstructured information, the purpose/context for the processing, and the role(s) third parties plan in the data lifecycle.

D. Conclusion

Data flow mapping will continue to provide measurable benefits to organizations in the long-term, providing a better understanding of how data is processed. The advantages of mapping related to providing transparency within organizations and placing information flows in context with business processes will only make organizations more effective with integrating data in the long-term. Greater transparency around data flows can (1) improve trust with consumers and employees, and (2) allow organizations to continue to refine their data processing procedures. In spite of data flow mapping's limitations, organizations will continue to find new ways to establish mapping processes that address their own needs and help them avoid unnecessary risks.

IV. NPRS Objective 3.5: Assure that Information Flows and Use Are Consistent with Privacy Rules

With both impressive benefits to organizations, and considerable challenges and limitations, it can be difficult to decide whether data flow maps are a necessary initiative. Like other resource-intensive endeavors, organizations will often consider legal ramifications or requirements before choosing to undertake data flow mapping. While no laws explicitly require organizations to develop and maintain data flow maps, there are a number of US and international laws where data flow maps can help achieve compliance.[36] Together, the laws describe requirements that address data presence, data flow, and data use.

Some laws and codes presume that a certain level of data flow mapping has been performed. For example, the Federal

36. For examples of such regulations, see, e.g., GDPR, *supra* note 4; Privacy Shield, *supra* note 35.

Rules of Civil Procedure (Rules), which govern the procedures (not substantive law) for how civil cases are brought in the United States, require that attorneys have a familiarity with technical systems handling legal data.[37] The Rules discuss subjects such as pleadings,[38] motions,[39] trials,[40] and most relevant here, discovery.[41] Rule 26 requires the preservation and production of electronically stored information.[42] It is through this rule that organizations are obligated to search terabytes of information and sometimes produce millions of documents. With strict deadlines and attorneys' fees to consider, organizations with thorough data maps are better situated at responding to these requests. The "Advisory Committee" notes, which explain the rationale behind many of the Rules, states that attorneys should be familiar with their clients' systems.[43] The Oklahoma Bar Association is more specific, saying: "counsel is often required to exercise this working knowledge by discussing what data is in each system . . . counsel must become intimately familiar with a client's data creation and storage and be able to be conversant in the same. This could require looking at a map of each client's database for his or her company."[44] At the outset, one can see that the law encourages the use of data flow maps even in a context separate from privacy and data security. The remainder of this

---

37. *See infra* notes 38–43 (laying out these rules).

38. Fed. R. Civ. P. 7–16.

39. *Id.*

40. Fed. R. Civ. P. 38–53.

41. Fed. R. Civ. P. 26–37.

42. *See* Fed R. Civ. P. 26 (instructing parties to produce ESI); Barbara J. Rothstein et al., *Managing Discovery of Electronic Information: A Pocket Guide for Judges*, SN012 ALI-ABA 1617, 1637-38 (2007) (noting that "amended Rule 26(f) and the accompanying Committee Note direct parties to discuss issues regarding the preservation of discoverable information, particularly with respect to [electronically stored information] because of its dynamic, mutable nature").

43. *See* Fed. R. Civ. P. 26(f) advisory committee's note to 2006 amendment

> When a case involves discovery of electronically stored information, the issues to be addressed during the Rule 26(f) conference depend on the nature and extent of the contemplated discovery and of the parties' information systems. It may be important for the parties to discuss those systems, and accordingly important for counsel to become familiar with those systems before the conference.

44. Cody J Cooper, *E-Discovery Under Rule 26*, 84 OKLA. BAR J. 543, 543 (2013).

section provides an overview of various jurisdictional requirements that either directly or indirectly require data flow mapping.

A. US Laws

US laws with significant privacy components, such as the Health Insurance Portability and Accountability Act (HIPAA)[45] and the Gramm Leach Bliley Act (GLBA),[46] likewise do not explicitly require data maps. Nevertheless, organizations that use them can benefit in helping to comply with these laws. For example, HIPAA requires certain healthcare organizations to minimize their data collection.[47] A data flow map showing all of the points of data collection could help organizations comply with this requirement. Similarly, the GLBA requires, inter alia, financial institutions to provide consumers with privacy notices that: (1) explain what categories of nonpublic personal information they collect, (2) with which affiliates and nonaffiliated third parties it is shared, and (3) the policies maintained to protect the confidentiality and security of the information.[48] The first two requirements can be supported directly with data flow maps tailored to include these particular pieces of information. The third requirement, providing consumers with the policies used to protect their personal information, can be supported indirectly by a standard data flow map that identifies which systems and transfers of data are secured. On a higher level, the data flow maps may also be used

45. Health Insurance Portability and Accountability Act, Pub. L. No. 104-191, 110 Stat. 1936 (1996).

46. Gramm Leach Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999).

47. *See* 45 C.F.R. § 164.502(b) (2015) ("When using or disclosing protected health information or when requesting protected health information from another covered entity or business associate, a covered entity or business associate must make reasonable efforts to limit protected health information to the minimum necessary to accomplish the intended purpose of the use, disclosure, or request."); *id.* § 164.514(d) (providing the "minimum necessary requirements").

48. 15 U.S. C. § 6803 (2012).

as evidence to support the existence of an information security program as required by the GLBA Safeguards Rule.[49]

The Fair Credit Reporting Act (FCRA), as modified by the Fair and Accurate Credit Transactions Act of 2003 (FACTA),[50] also requires the destruction of certain types of information, namely, consumer reports.[51] FACTA's Disposal Rule, which applies to businesses and individuals that collect and use consumer reports,[52] permits the destruction method to vary based on the sensitivity of the data being destroyed.[53] Data flow maps can therefore be used to demonstrate that the destruction methods are reasonable given the sensitivity of the data.[54]

---

49. *See Financial Institutions and Customer Information: Complying with the Safeguards Rule*, FTC (Apr. 2006), https://www.ftc.gov/tips-advice/business-center/guidance/financial-institutions-customer-information-complying (last visited Apr. 24, 2017) ("Under the Safeguards Rule, financial institutions must protect the consumer information they collect.") (on file with the Washington and Lee Law Review).

50. 16 C.F.R. §§ 682.1–5.

51. *Id.* § 682.3(a) ("Any person who maintains or otherwise possesses consumer information for a business purpose must properly dispose of such information by taking reasonable measures to protect against unauthorized access to or use of the information in connection with its disposal.").

52. A "consumer report" means

> any written, oral, or other communication of any information by a consumer reporting agency bearing on a consumer's credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living which is used or expected to be used or collected in whole or in part for the purpose of serving as a factor in establishing the consumer's eligibility for [credit, insurance, or employment].

15 U.S.C. § 1681a.

53. *See* 16 C.F.R. §§ 682.3(b) (providing several examples of "[r]easonable measures to protect against unauthorized access to or use of consumer information in connection with its disposal").

54. The Federal Trade Commission supports this method in its tips to businesses. *See Disposing of Consumer Report Information? Rule Tells How*, FTC (June 2005), https://www.ftc.gov/tips-advice/business-center/guidance/disposing-consumer-report-information-rule-tells-how (last visited Apr. 24, 2017) ("The FTC says that financial institutions that are subject to both the Disposal Rule and the Gramm-Leach-Bliley (GLB) Safeguards Rule should incorporate practices dealing with the proper disposal of consumer information into the information security program that the Safeguards Rule requires.") (on file with the Washington and Lee Law Review); Press Release, FTC, FACTA Disposal Rule Goes into Effect June 1 (June 1, 2005), *available at* https://www.ftc.gov/news-events/press-releases/2005/06/facta-disposal-rule-goes-effect-june-1 (last visited Apr. 24, 2017) (same) (on file with the Washington and

Compliance with state laws can also be assisted with data flow mapping. Data flows not only include the transfer of data internally and to third parties, but also detail data retention procedures, including destruction and/or archival. Two-thirds of US states have enacted laws that require the disposal or destruction of personal data.[55] Some laws, for example New Jersey's Identity Theft Prevention Act,[56] are were drafted to combat the rise of identity theft. The rationale is simple: the more personal data stored and transmitted, the more opportunities for its loss and misuse. Requiring the destruction of personal data mitigates this risk and data maps provide evidence of processes to execute on destruction policies.

B. Illustrative International Laws

Data flow maps are particularly useful in documenting and understanding how data is transferred between different nations. The GDPR forbids the transfer of personally identifiable information (PII) to nations outside of the European Economic Area that do not have "adequate" privacy laws in place.[57] Organizations subject to this law can use data flow maps to efficiently ascertain whether they possess PII on EU-based data subjects and whether that data is transferred to an inadequate nation. They can likewise track where and when data has been anonymized to support permissible secondary processing of the data.[58] Even when data is not transferred across borders, the GDPR still imposes obligations on data controllers (e.g., to

---

Lee Law Review).

55. *See Security Breach Notification Laws*, NCSL (Feb. 24, 2017), http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx (last visited Apr. 24, 2017) (providing a list of which states have enacted security breach notification laws) (on file with the Washington and Lee Law Review).

56. N.J. Stat. Ann. § 56:8-161 (West).

57. *See* GDPR, *supra* note 4 ("A transfer of personal data to a third country or an international organisation may take place where the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection.").

58. Secondary processing refers to the processing of data for a purpose other than for which it was initially collected.

maintain records of processing activities) that may be satisfied in part through data flow mapping.[59]

Another example of a data flow map's utility in identifying cross-border data transfers is when evaluating compliance with the Russian data localization requirement. This law requires personal data of Russian citizens to be stored within the geographic boundaries of Russia.[60] Again, data maps that identify data collection points (to identify which data originates from Russian citizens), data types (to confirm personal data is collected), data storage locations and cross border data flows (to determine data is stored within Russia) can be used to ascertain compliance with this law.

C. Data Use Requirements

In addition to the benefits data maps provide for data flow regulations, they can also assist with laws dealing with data use. In many jurisdictions, data subjects must provide consent prior to the processing, which includes collection of personal health data. For advanced use cases, a tagging schema[61] may be used to track the details of user consent. Data flow maps can assist with documenting where in the process users provide consent and to what processing they have consented, in addition to the location

---

59. *See supra* note 33 and accompanying text (discussing the GDPR's recordkeeping requirements).

60. *See* IRINA TYMCZYSZYN & DAVID A. ZETOONY, BRYAN CAVE, RUSSIA DATA LOCALIZATION REQUIREMENT AT A GLANCE: PRACTICAL ASPECTS 1 (2015), bryancavedatamatters.com/wp-content/uploads/2015/05/Russia-Data-Localization-Requirement-at-a-Glance.pdf ("The most significant amendment introduced by Law No. 242 is the requirement that data operators must store personal data of Russian citizens on servers located within the territory of the Russian Federation.").

61. A tagging schema facilitates the addition of metadata to help identify the following aspects related to data collection and storage: (1) what is the legitimate ground for the data's collection, (2) is there a legal basis to support the collection of sensitive data, (3) for what purpose is the data collected, (4) for how long can the data be retained, (5) what controller/processor responsibilities exist, (6) what jurisdictions are in scope, and (7) to what extent is the data identifiable? The goal is for these attributes to be bound to the data itself rather than to a system which may share data and lose the original context of the data's collection and along with it the legal data processing requirements that pertain to that data.

of the user (to aid in determining what laws apply) and the context of the data transaction. Internationally, the GDPR also places restrictions on how data can be used, absent informed consent from a data subject.[62] In many cases, laws prohibit the use of personal data for secondary processing unless, inter alia, that data has been anonymized. Data flow maps can demonstrate that only anonymized data gets passed along to systems that perform secondary processing. Further, when laws require reasonable security measures to be put in place, data maps provide evidence of compliance to regulators and other interested parties.

D. Conclusion

Data flow maps are the bridge between security, legal, privacy, and business stakeholders. While stakeholders each have their own priorities and concerns, data flow maps can effectively provide a common language so that collaboration can take place and resources can be directed to areas in need.[63] With finite resources, stakeholders can leverage data flow maps to convey to sponsors why systems or processes require additional investment to enable data protection. It provides an understandable and highly visual demonstration of how a given system or process may be a weak link in the chain that could open the entire organization to a breach. Data flow maps showing where third parties have access to an organization's system(s) can help prioritize security infrastructure and training and avoid costly breaches caused by careless third parties. An organization's privacy office, legal department, and compliance group should collectively evaluate the risks they face and explore how data maps can mitigate those risks.

---

62. *See* GDPR, *supra* note 4 ("In order for processing to be lawful, personal data should be processed on the basis of the consent of the data subject concerned or some other legitimate basis, laid down by law, either in this Regulation or in other Union or Member State law as referred to in this Regulation.").

63. *See generally* IAN G. DIBERNARDO, STROOCK & STROOCK & LAVAN LLP, USING DATA MAPS AS AN EFFECTIVE TOOL FOR DATA SECURITY COLLABORATION (2013), http://www.stroock.com/siteFiles/Pub1391.pdf (advocating the use of data flow maps).

V. Case Study

Suppose that an undergraduate institution (the Institution) sought to understand potential data privacy-related risks in their application and admissions process. Colleges and universities, through the nature of the application process, tend to handle data of varying degrees of risk, including sensitive information such as tax identification numbers (e.g., social security numbers). Because these applications often come from data subjects around the world, the obligations on handling the data submitted can vary greatly as well. Additionally, their admissions process included a number of highly confidential review steps, metrics and other proprietary information that, if made public, could be used to the competitive benefit of other institutions or potentially by applicants to gain unfair advantages over other applicants.

To identify the types of data handled, develop a complete picture of the admissions business process, and document risks associated with its handling of personal and proprietary information, the Institution sought to develop data flow maps manually through interviews and questionnaires. This option was chosen because (1) the exercise could be initiated immediately without purchase and implementation of new technology, (2) the Institution was willing to hire a third party to support the manual process (as it tends to be more labor intensive), and (3) the limited scope of the data flow diagram exercises would make maintenance easy to track.

The process for developing data flow maps started with careful collection and review of published policies on the handling of information, as well as the gathering of their compliance obligations. The compliance obligations were rationalized (or harmonized) in a single framework, such that overlapping requirements were combined into the same testable control. Once these preparatory steps were complete, the Institution drafted a short list of easy-to-answer questions for distribution to the personnel in the admissions office. The questions solicited high-level responses on the types of data collected, sources of the data, whether certain protections were in place, requests for detail on the purpose of the information collected, etc. Once the questionnaires were gathered and analyzed, the Institution selected a subset of the questionnaire recipients to participate in

a follow-up workshop. The workshop involved review of the business process, collection activities, uses and purposes for the data, storage-related activities and systems, and ended with discussion on the types of data archived versus the data deleted, and the methods through which either activity was executed. The workshop was based on facilitated discussion and informal, white-board based mapping of the attendee's responses.

Upon completion of the workshop, the Institution converted questionnaire responses and workshop notes into a graphical diagram of the data lifecycle within the admissions process. The diagram included text-based narratives detailing the core information processing steps, clear identification of internal and external parties involved, and a notation of security and privacy mechanisms used to protect data. Further, the diagram also indicated areas of potential cyber and privacy compliance risks associated with the existing process. To conclude the actual data flow mapping process, the Institution scheduled a small number of follow-up interviews with various subject matter specialists and process owners for the admissions process to confirm the accuracy of the data flow map.

As an outcome of the data flow mapping exercise, the Institution was able to leverage the identified risks to personal and proprietary information to develop a set of remediation steps to address the risks. Since the data flow map was drafted to be accessible to a non-technical audience, the Institution was able to combine the data flow map and the associated remediation steps to create a clearly articulated business case for additional funding from Institution leadership. Once additional funding was granted and allocated, the Institution was able to implement the remediation steps to reduce the risks associated with its Admissions process. In pursuit of these remediation activities, the Institution was also able to provide transparency to internal teams seeking to aid the admissions office in protecting information, including Information Security, Compliance, and Legal.

VI. Conclusion

This Article has repeatedly emphasized that mapping provides consistent, cohesive, and comprehensive opportunities for organizations to understand how data flows relate to their businesses processes. While generally not a legal or regulatory requirement in either the United States or elsewhere, data flow mapping demonstrates significant value in helping organizations both stay compliant with security and privacy laws as well as improve the transparency of their data lifecycles. Meeting these objectives, as laid out by the NPRS,[64] is not only important for improving the short-term privacy posture of organizations, but also preparing for long-term big data and privacy trends. The use of clouds and data lakes provide unlimited less expensive ways for organizations to hold onto data for longer periods of time, whereas in prior times, organizations would face tangible limitations on enabling further use and analysis for business purposes. This permits organizations to use and scale data in new innovative ways, providing more context on day-to-day activities and customers' needs in a variety of industries.[65] Additionally, the growing adoption of sensors and other Internet of Things technology, provide new avenues for organizations to collect data relative to their work and purposes in real time, completely changing how they operate. These developments, besides opening the gates for data flood into organizations, will also present them with a new array of privacy challenges that will different between jurisdictions.[66] Both of these trend lines will only make the task

---

64. *See generally* NPRS, *supra* note 5 (discussing these objectives).

65. *See* Andy Haler, *Data Lake Concept Needs More Big Data Use Cases to Flourish*, TECHTARGET (June 2015), http://searchdatamanagement.techtarget.com/feature/Data-lake-concept-needs-more-big-data-use-cases-to-flourish (last visited Apr. 24, 2017) ("Consumer analytics offers further big data use cases.") (on file with the Washington and Lee Law Review).

66. *See* HEWLETT PACKARD ENTERPRISE, SECURING THE INTERNET OF THINGS 7 (2015), https://www.hpe.com/h20195/V2/GetPDF.aspx/4AA6-3369ENW.pdf

> Regulation will lag development—To support innovation, industry and governments must seek the right balance between free-market development and regulation. . . . There will also be increasing regulatory change, and companies will continue to struggle with it. Regulation will be characterized by the inconsistency of laws among countries, different levels of social responsibility, and business

of tracking and classifying data as well as monitoring data transfers more important. While data flow mapping can help organizations now, it will be critical in the future for helping them maintain their competitive edge.

---

competition.

See also Marcel, *The Big Picture of Big Data: Mapping the Internet of Everything*, DATAMASHUP (Dec. 16, 2016), http://www.datamashup.info/the-big-picture-of-big-data-mapping-the-internet-of-everything/ (last visited Apr. 24, 2017) ("Companies that wait to step into the big data stream risk losing customers to those that have already adopted real-time data technologies. Bringing geospatial technology to the vast Internet of Everything (IoE) opens opportunities that are only visible within a geographic context.") (on file with the Washington and Lee Law Review).