The Life of Data in Compliance Management

Nick Scope
College of Computing and Digital Media
DePaul University
Chicago, United States
Email: nscope52884@gmail.com

Alexander Rasin
College of Computing and Digital Media
DePaul University
Chicago, United States
Email: arasin@cdm.depaul.edu

Karen Heart
College of Computing and Digital Media
DePaul University
Chicago, United States
Email: kheart@cdm.depaul.edu

Ben Lenard
College of Computing and Digital Media
DePaul University
Chicago, United States
Email: blenard@anl.gov

James Wagner
Department of Computer Science
University of New Orleans
New Orleans, United States
Email: jwagner4@uno.edu

***Abstract*—Data privacy polices mandate requirements to protect the privacy of individuals, prevent fraud, and support audits. Organizations also implement their own internal data policies to minimize liabilities and protect user privacy. In practice, it is difficult (or impossible with most systems active today) to achieve the desired purpose of these policies due to technological limitations of storage systems. These limitations are ultimately caused by the lack of native database support for data privacy compliance. This paper surveys the principles of data compliance and analyzes the requirements imposed on organizations. We begin by defining data compliance terminology that must be shared between legal and technology domain experts; legislation and litigation examples provide real-world context and motivation for our analysis. Since the data life cycle model is universally accepted in data management, we next discuss how data compliance can be integrated into this model to fully support data management policies. Finally, we consider the open problems with current data storage systems and discuss the requirements for automated privacy regulation compliance.**

Keywords: Compliance Management; Privacy Regulations

I. INTRODUCTION

Data management by an organization is bound to data governance policies (e.g., internal requirements or government agency mandates) that define how the data must be stored and used. These policies include data retention (how long the data must be kept), data purging requirements (when the data must be destroyed), and data consent (whether the data can be used for a particular purpose). Failure to comply with these policies could result in large fines, a loss of customers, and an irrecoverable breach of customer data privacy.

Data policies set forth by legislation have been in place for decades. Some examples include the Health Insurance Portability and Accountability Act (HIPAA) of 1996 [8] (patient healthcare data) and the Gramm–Leach–Bliley Act of 1999 [32] (business records of financial institutions); additionally, new policies are continuously introduced, such as California's Proposition 24 of 2020 [7] (which expanded on the previous law "The California Consumer Privacy Act"). Furthermore, there are also instances where organizations must follow additional internal policies or policies of business contacts [10]. We further discuss these motivating examples and more in Section III.

A. Motivation

The data life cycle model is the state-of-the-art structure for organizations to understand and manage their data assets [25]. By examining the data life cycle phases, we can clearly see how compliance must be considered throughout an organization's processes. Since published literature presents several variations of the life cycle phases, we have abstracted the relevant phases from [4], [13], and [23] in Figure 1. Our phase definitions bridge the different goals of domain experts, legal departments, and other stakeholders associated with the data and legal requirements. These were developed to promote discussion on where policy compliance must be considered. For example, some data life cycles model "Usage" as multiple phases (e.g., analysis, reporting). However, for the purposes of privacy compliance, we believe that a single phase captures all necessary aspects of "Usage". To evaluate data lifecycle outside the scope of policy compliance, alternative models may be more applicable. For example, to analyze security considerations, the transition of data between phases would be modeled in further detail. Overall, we use our phases to guide a discussion on future research necessary to remedy the database software shortcomings and facilitate automated compliance management.

Figure 1 illustrates the connections between data life cycle management phases (we detail phase requirements in Section IV). For example, archival phase follows data usage phase; usage phase may also affect the decisions associated with the storage phase. Adding to the complexity, not all data goes through each phase. For example, data may be under an indefinite retention policy, staying in the archival phase and never proceeding to the destruction phase.

B. Contributions

Current systems are missing key functionality, which prohibits complete automated compliance; until these gaps are filled, consumer privacy will suffer. Due to the limitations

Copyright (c) IARIA, 2021. **ISBN: 978-1-61208-893-8**

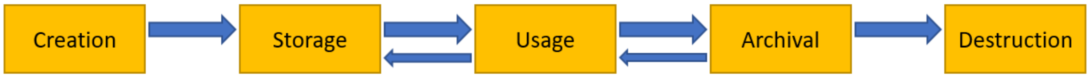

Figure 1. Data life cycle phases

in both research and technological implementations focusing on policy compliance, we believe our discussion provides the following contributions:

1) Surveys the current domain challenges and background information on data privacy compliance.
2) Analyzes how compliance must be considered at each phase of the data life cycle to satisfy legal requirements.
3) Discusses the current technological shortcomings that must be explored to automate policy compliance.

II. Domain Challenges

A. Concepts

Business Record: Organizational rules and requirements for data management are defined in units of business records. United States federal law refers to a business record broadly as any "memorandum, writing, entry, print, representation or combination thereof, of any act, transaction, occurrence, or event [that is] kept or recorded [by any] business institution, member of a profession or calling, or any department or agency of government [...] in the regular course of business or activity" [31]. In other words, business records describe any interaction or transaction resulting in new data.

Business records can be represented using different logical layouts. A business record may consist of a single document for an organization (e.g., an email message). In a database, a business record may span many combinations of rows across multiple tables (e.g., a purchase order consisting of a buyer, a product, and the purchase transaction from three different tables). The process of mapping business records to underlying data can vary depending on an organization and the data storage medium.

Policy: A data policy is any formally established rule for organizations dictating the requirements (i.e., how long data must be saved, when data access requires consent, and when data must be purged). Policies can originate from a variety of sources such as legislation or as a by-product of a court ruling (examples in Section III). Companies may also establish their own internal data retention policies to protect confidential data. In practice, database administrators work with domain experts and sometimes with legal counsel to define business records and retention requirements based on the written policy.

Policies can use a combination of time and external events as the criteria for data retention and destruction. For example, retaining employee data until employee termination plus 5 years illustrates a policy criteria that is based on a combination of an external event (employee termination) and time (5 years). The United States Department of Defense (DoD) "DoD 5015-02-STD" [2] outlines the minimum requirements and guidance for any record system related to the DoD, which includes how organizations must preserve and destroy data. Moreover, multiple US government agencies, such as the National Archives, use the same standards.

Policy compliance can be complex due to multiple overlapping policies or criteria for the same business record, or due to different data points belonging to different business records. For example, different rows or columns of a table belonging to an order purchase could be governed by different policies: purchase information (e.g., price) may fall under different retention policies versus customer information (e.g., address). Policy mapping must also consider the potential conflict between multiple policies with data retention and destruction requirements.

Verification: Data curators must be able to query the policies and the status of all business records in storage. Data storage systems must support a standard mechanism for defining the policies, listing or modifying current policies, and checking for potential conflicts (e.g., policies requiring retention and destruction of the same data) or overlap between different policies. For example, if an organization is unable to destroy data when requested by a customer, their refusal must be justified.

Enforcement: Enforcing policies includes archiving and deleting data as required as well as verifying consent when processing data. Enforcing a policy maintains an organization's compliance. Current database management systems do not incorporate automated robust data policy features; as a result, organizations are forced to develop manual solutions for policy compliance. Automated enforcement of policy requirements will both increase compliance and customer privacy.

B. Data Governance Topics

Retention: Retention defines the conditions when a business record must be preserved. Some organizations may choose to delete data once it is no longer needed to minimize liability (e.g, data theft or requested through legal discovery). Others may store the data longer than minimally required (in the gray area between "must be retained" and "must be destroyed"). DoD guidelines state that any storage system must support retention thresholds such as time or event (Section C2.2.2.7 of [2]). Some retention requirements, such as HIPAA with healthcare data, may require a complete historical log of any and all business record updates (e.g., current address and full history of address changes for a patient) [8]. Organizations subject to this level of retention must archive the complete business record before every update to ensure a complete audit trail history was preserved.

Consent: Per GDPR Recital 40, "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" [27]. Additionally, per Article 4(11), "Consent of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." The processing of these business records must be verified according to customer's consent (e.g., marketing versus order processing). Not all processing requires consent; data necessary to complete the business transaction for which data was collected does not require explicit consent.

Purging: In data retention, purging is the permanent and irreversible destruction of data in a business record [20]. Purging requirements establish when a business record must be destroyed. A business record purge can be accomplished by physically destroying the device which stored the data, encrypting and erasing the decryption key (although the ciphertext still exists, destroying the decryption key makes it inaccessible and irrecoverable), or by fully erasing data from all storage. If any part of a business record's data remains recoverable or accessible in some form, then the data purge is not considered to have been successfully completed. For example, if a file is deleted through a file system, but can still be recovered from the hard drive using a forensic tool, this does not qualify as a purge [20].

Some policies require an organization to completely purge business records either after the passage of time, at the expiration of a contract, or purely when the data is no longer needed. Additionally, there are an increasing number of regulations, such as the European Union's General Data Protection Regulation (GDPR) [27], which require organizations to purge business records at the request of a customer. Therefore, organizations must be prepared to comply with purging policies as well as ad-hoc requests.

III. Legal Precedent

Although it is beyond the scope of this paper to provide an overview of all data privacy legislation across different domains, we discuss some of the most impactful government regulations. Private organizations (such as in the financial industry) may set additional policies for any companies that do business with them [10]. Overall, we believe the following examples offer the most significant motivation to increase database support of data privacy compliance.

General Data Protection Regulation: In the European Union, the General Data Protection Regulation greatly expanded consumer power over personal data. This regulation was put into effect May 25, 2018, and is arguably "the toughest privacy and security law in the world" [27]. Any organization which is registered in the European Union, offers goods or services, or monitors behavior of EU residents must comply with GDPR requirements (regardless of whether the organization is based in the EU).

One significant addition to data privacy rights due to GDPR is the "Right to be Forgotten" [12], which allows individuals to request that companies delete all of their personal data. In Recital 65, Recital 66, and in Article 17 GDPR states: "The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay" [33].

Requesting data deletion does not guarantee that customer's data will be purged. When customers request their data to be deleted, organizations must check if they are legally permitted to do so. If an organization has a retention requirement on this data, they will not be able to purge it despite the request. Without automated verification, organizations must manually check for any applicable retention obligation when deleting data. The "Right to be Forgotten" requests require a response within a month. Thus, without an automated process, organizations must manually process requests within a deadline.

GDPR non-compliance carries significant penalties. On January 15, 2020, TIM was fined €27.8 million by the Italian data protection authority, the Garante, for violations of GDPR [16]. A large number of complaints were filed between January 1, 2017 until early 2019 due to TIM's unwanted promotional calls. Some customers (who did not consent) were contacted over 150 times in a single month. Because TIM did not enforce data consent, this eventually led to being fined due to failing to comply with GDPR.

California Consumer Privacy Act of 2018 & Proposition 24: California passed the California Consumer Privacy Act of 2018 (CCPA) [6], which was greatly inspired by GDPR and offers many similar privacy rights. CCPA went into effect on January 1, 2020 and California Attorney General began enforcing CCPA on July 1, 2020 [5]. Regardless, some privacy advocates believe CCPA did not go far enough to protect data privacy and are still pushing for additional regulations [11]. The California Secretary of State summarizes the position of the advocates for Proposition 24, who believe that this proposition will further increase consumer data privacy [7]. Data retention and management legislation is continuously evolving as Proposition 24 [19] just passed in November 2020.

Proposition 24 exemplifies the continuous battle of privacy advocates against those who are concerned about too much government regulation. Data retention is an evolving field where requirements are continuously changing; organizations must be able to quickly adapt to new and changing requirements. Because systems currently cannot easily add retention protections, any organization which manually added protections with CCPA now must manually update all of those protections due to Proposition 24.

Zubulake v. UBS Warburg: Between 2003 and 2004 in New York, the case of Zubulake v. UBS Warburg resulted in many additional electronic record keeping requirements [34]. According to Li at ABAJournal.com, "Companies were put on notice that they had a duty to preserve data once they reasonably anticipated they might be sued. [...] Otherwise, the consequences could be severe and a party could be hit with sanctions [which] could cripple its ability to mount a defense" [22]. In summary, this ruling clarified that organizations are required to retain all pertinent electronic data if they are aware of a forthcoming lawsuit, even before being explicitly requested to do so.

The plaintiff's attorneys argued that they were not given all of the relevant evidence to her case; the judge concluded that UBS Warburg did not hand over all relevant data. It was determined that the only copies of some relevant data was archived on tapes (which, in turn, made it expensive and difficult to acquire and review). This led to the judge's opinion that "anyone who anticipates being a party or is a party to a lawsuit must not destroy unique, relevant evidence that might be useful to an adversary" [34].

Health Insurance Portability and Accountability Act: In 1996, the United States passed a law that had a significant influence on the way privacy and retention of health care data was managed. HIPAA imposes requirements for both data retention as well as for purging. Per the United States Department of Health and Human Services, HIPAA increases healthcare recipients power over their own data in respect to both privacy as well as transparency [30]. According to research by Annas, HIPAA requires that "[...] a patient's entire medical record can seldom be lawfully disclosed without the patient's written authorization" [1]. HIPAA raised the minimum standards of privacy and medical data, empowering individuals to control their data.

For example, HIPAA Subpart D §164.504 "Uses and Disclosures: Organizational Requirements" [8], requires an individual's data to be destroyed at the expiration of a contract. "At termination of the contract, if feasible, return or destroy all protected health information received from, or created or received by the business associate on behalf of, the covered entity that the business associate still maintains in any form and retain no copies of such information[....]" Therefore, organizations must have a robust data purging process.

HIPAA also requires healthcare recipients in the United States to be clearly informed of their privacy rights. Per the United States Health and Human Services, "The HIPAA Privacy Rule requires health plans and covered health care providers to develop and distribute a notice that provides a clear, user friendly explanation of individuals rights with respect to their personal health information and the privacy practices of health plans and health care providers" [18]. Currently, there is no automatic process in deployed database systems to report all active retention policies and their relevant business records. Any request on which retention policies apply requires a manual lookup process.

IV. Privacy Requirements in the Data Life Cycle

A. Creation

In most organizations, every transaction creates data that could be stored and processed. Whether this data will move on to the storage phase of the life cycle depends on the requirements defined by the owner. In certain industries, such as the financial industry, all transactional data must be stored.

The first step of compliance is mapping the new data to policy requirements. This process typically involves domain experts working with the database administrators and legal professionals. Choosing protections placed on the data must be done immediately, lest retention compliance be violated before protections have been implemented.

Currently, many organizations have standardized processes which include data classification, but this classification is orthogonal to data policy compliance. For example, organizations may automatically label data generated from specific sources as "Internal Only" or "Highly Confidential". These labels indicate that the data must only be accessed by a specific audience but may not align to policy requirements. For example, "Highly Confidential" does not align to any specific requirement dictating when data must be purged.

B. Storage

After data is created, it enters a data storage management system. A number of considerations are factored into the data storage software choice. Organizations with large volume of transactions and a consistent (structured) data will typically deploy a relational database. For data which is less structured and more dynamically evolving, organizations may choose a NoSQL database [17]. Alternatively, keeping data in file documents in a simpler database may satisfy an organization's data storage requirements.

The type of storage used can greatly impact the difficulty in complying with data retention policies. When business records are stored in documents, this task will be simpler. When using advanced databases, the mapping of business records to the stored data is much more complex. When retention and purging requirements correspond to individual files, solutions such as Amazon S3 (which offers a file-level object life-cycle management) facilitate retention and purging compliance. The DoD's "Electronic Records Management Software Applications Design Criteria Standard" (DoD 5015-02-STD) requires systems support both time or event criteria.

Furthermore, most storage solutions require finer granularity than storing a file per business record. As discussed in Section II-A, businesses records may span multiple tuples across tables in relational databases. Prior research addressed retention [28, 3] and purging [29] in relational databases.

C. Usage

The use of data depends on the data owner's policies as well as the organizational need for the data. Common data uses include storing customer information, running statistical analysis to discover underlying trends, and documenting business transactions. Data may be continuously used for an extended duration and for multiple purposes. For example, customer records which are used for shipping orders may also be used for analyzing trends in customer purchase patterns.

Data consent support must be an inherent part of data privacy management. Databases neither offer functionality to define business records (with respect to consent) nor filter on consent for various processing uses (e.g., marketing). Because data privacy regulations require organizations to acquire customer consent when processing data for certain purposes, storage systems must guarantee verification of customer consent. Business records which have not been allowed to be processed must be excluded from the query output. Access control based on the identity of data analyst would not facilitate compliance.

On May 25, 2018 (the day GDPR took effect), Google was found in violation of GDPR [15], leading to a fine of

€50 million. Google was convicted for lack of transparency and failing to acquire user consent for data processing in an instance where consent was required.

D. Archival

HIPAA [9] requires medical data to be retained for at least 6 years. Therefore, organizations have an increased obligation to maintain archived data, even after the data is no longer needed for their operation. Business records that are no longer needed but must be preserved under a retention policy must be moved to an archive until the retention criteria has expired.

In order to reach the archival phase of the data life cycle, data from business records which are no longer needed in usage phase must be under a retention policy. Archived data is the data that has lost its primary relevance but is still required to be available in storage (e.g., historical or reporting purposes). Therefore, archived data does not require regular updates nor is expected to be actively used. Instead, it is stored in a separate repository until it is eligible for destruction (i.e., no longer requiring retention). As long as the data is subject to at least one retention policy, it must remain archived.

Archived business records do not require any updates nor should they be deleted while under retention. In rare situations, data in archive may be returned to active storage for usage (e.g., the result of a lawsuit). Any retention compliant system must purge business records from the archive once they no longer require retention and have a purge policy requirement.

E. Destruction

Once data is no longer needed and is not subject to retention requirements it may enter the destruction phase of its life cycle. Some organizations have data with a retention period “for the life of the company” meaning that it will never enter the destruction phase. On the other hand, some policies, including those from government regulation, explicitly require business records to be destroyed when no longer used nor requiring retention. HIPAA Subpart D §164.504 requires organizations to delete data at the end of a contract if there are no other applicable retention requirements [8]. The Children’s Online Privacy Protection Act states that personal information for children can be retained “for only as long as is reasonably necessary to fulfill the purpose for which the information was collected” (Section 312.10 of [14]).

Google has been repeatedly fined for violating GDPR’s “Right to be Forgotten” [24]. Google refused to delete customer data at their request despite having no legal basis for retaining the data. France, Sweden, and Belgium have all imposed fines for violations of failing to delete requested data.

To fully comply with purging requirements, systems must implement functionality that allows organizations to define business records and policies, which will automatically be enforced across active databases and backups [29]. This functionality must implement some form of secure deletion to render the required data permanently and irrecoverably destroyed. If any data belonging to a business record requiring purging is recoverable (whether the data exists in the active database, underlying database pages recoverable using forensic tools, or in a backup), the purge policy has not been properly enforced, and the system would not be considered compliant.

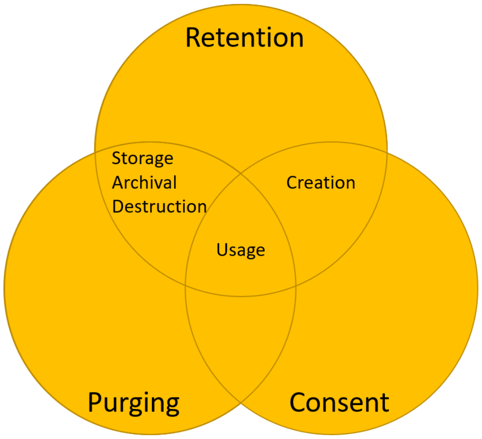

Figure 2. The relationship between life cycle phases and requirements.

V. Open Problems

The complexity caused by data policy requirements coming from a variety of sources (each with their own changes in requirements) means manually achieving compliance is an extremely difficult task. Throughout this paper, we outlined where in a data life cycle each type of compliance must be implemented. Manually achieving compliance requires each individual user to know which policies apply to which business records for each purpose. Therefore, facilitating comprehensive policy enforcement requires automated data policy compliance tools. Regardless of policy type, legacy systems will continue to be a difficult challenge. In this section, we discuss the requirements for the three main areas of data compliance and how they overlap with data life cycle phases.

Retention must be considered during each phase of the data life cycle. On the other hand, purging and consent must only be considered during some of the phases. For example, destruction does not require user consent (although users can request their records to be deleted, they cannot prevent required record purging). Figure 2 provides an overview on where each phase aligns to each governance requirement.

In practice, DBAs work with domain experts and legal counsel to define business records and retention requirements based on policies. Automated systems typically assume that data curators can express business records as a query or as a collection of files. The initial process of mapping the business records and retention policies to database tuples will always be a manual process; any automated system then will reference these definitions to enforce policy compliance.

A. Retention

Retention compliance is achieved by maintaining all relevant business records until some criteria (time or event) has been met. This can be achieved by either 1) blocking transactions which would delete or update protected data or 2) automatically archiving business records in a separate database before deleting or updating data. Either solution requires guaranteeing that update and delete operations automatically cross-reference defined policies and retention criterion.

Databases do not currently offer functionality to enforce retention and archival compliance. Currently, organizations build ad-hoc solutions manually. If any data targeted by a delete or update is protected by a policy, the automated system must either archive the entire business record as-is before executing the transaction [28] or block the transaction [3]. Systems must automatically cross-reference defined business records and retention requirements to archive data when deletes or updates would violate retention requirements. Systems proposed by Scope et al. [28] and Ataullah et al. [3] use triggers in relational databases to enforce retention policies.

Lawsuits may impose sudden and critical retention requirements on various business records. Zubulake v. UBS Warburg expanded on the precedent of organizations being required to retain any applicable business records for the duration of the case [34]. Organizations must be able to easily retain and archive all applicable records.

As shown in Figure 2, retention must be considered at each phase of the data life cycle. Retention must be immediately mapped at creation, and must protect data across storage, usage, archival, and destruction phases. If data prematurely enters the destruction phase from any of the other phases when retention is still required, compliance has been violated.

B. Consent

Regulations such as GDPR require user consent for certain types of data processing. Because the same analyst may process data for a variety of purposes, user-based permissions do not satisfy consent requirements. Therefore, research must implement automated filtering where business records require consent for an input purpose.

Consent must defined at data creation (although customers can revoke or give consent at any time). Additionally, this consent must be applied during data usage depending on whether consent is required. On the other hand, consent policies do not have to be considered during storage, archival, and destruction phases.

One common paradox are customers who demand that all of their data is deleted and to not be contacted in the future. Although the customer is revoking their consent to use their data, the organization cannot delete all of the data without risking contacting them in the future. Thus, organizations must maintain some data on a do-not-contact list, as long as the data maintained on this list is only referenced as a filter.

Although detailed usage requirements are beyond the scope of this paper, we also must note that consent is defined differently by different governing bodies. While some simply require customers to be remain anonymous (but still allowing their data to be used in aggregations), others do not allow any customer data to be included without their permission. Therefore, multiple independent solutions may be required to satisfy the different definitions and requirements.

C. Purging

Purging requires that organizations irreversibly and irrecoverably destroy their data after some criteria has been met. Database administrators must work with domain experts to guarantee that these are mapped as to not conflict with retention policies. If data is prematurely destroyed as the result of a user input or automated policies, this does not violate compliance (unless this violates retention).

Purging must remove data from both all backups (both accessible and inaccessible) and from active storage. If data is recoverable by forensic tools or by backup, compliance has not been achieved. Because data requiring purging is simultaneously stored in systems with data requiring retention, simply destroying the physical storage would satisfy purging compliance at the cost of retention compliance.

Multiple enhancements must be developed to achieve full compliance. First, automated systems must automatically delete all files or tuples in a database as necessary. Once the files or tuples are deleted, they may still be recoverable via forensic means. Therefore, the data must be deleted so that it is no longer accessible to forensic tools. Finally, data must be purged from all backups.

Reardon et al. [26] offered a comprehensive overview of secure deletion, which both provides various approaches and requirements for completely purging data from a storage systems. In their paper, the authors defined three user-level approaches to secure deletion: 1) execute a secure delete feature on the physical medium 2) overwrite the data before unlinking or 3) unlink the data to the OS and fill the empty capacity of the physical device's storage. For all three methods, one requirement is the ability to directly interact with the physical storage device. Therefore, these approaches are only applicable for physically accessible databases (which may not be possible for backups in storage).

Lenard et al. [21] provided an analysis on how long deleted or updated data remains in underlying database pages, which eventually leads to them being included in backups. Therefore, to fully purge data from all storage, it is necessary to both implement steps to remove the data from active storage as well as backups. Scope et al. [29] proposed using a form of cryptographic erasure to purge pertinent business records from backups of relational databases.

D. Performance Considerations

Throughout this paper we outlined the requirements and benefits of automated compliance. This automation does have an associated performance cost. Research by Scope et al. [28] and Ataullah et al. [3] performed experiments detailing the runtime overhead of their additional retention protections.

Balancing performance with automated enforcement is a difficult problem for future research. For example, financial organizations are heavily motivated by system performance. Automated trading system measure execution time in milliseconds. These systems use speed for a competitive trading advantage, but they are also subject to extreme regulations by both the exchanges as well as major government bodies. For these industries, implementing automated compliance will require optimization to minimize impact on system performance. For the necessary functionality enhancements to be widely adopted, these enhancements must both guarantee compliance and minimize system performance overhead.

VI. Conclusion

In this paper, we outlined the data life cycle model and the steps that must be Incorporated into the process to facilitate privacy compliance. Recognizing the data retention needs at each phase of the data life cycle provides a framework for where additional research must be prioritized to satisfy compliance management requirements. Current storage solutions do not have the necessary functionality to automatically enforce data governance policies. Until the necessary functionality is implemented, ad-hoc manual solutions will continue to risk violating privacy regulation requirements. Privacy compliance is only continuing to grow in importance; these issues must be addressed to increase user privacy protections.

Acknowledgment

This work was partially funded by US National Science Foundation Grant IIP-2016548 and CME Group. The authors thank Thamer Al-Johani for comments on the paper draft.

References

[1] G. J. Annas. *HIPAA regulations—a new era of medical-record privacy?* 2003. URL: https://pubmed.ncbi.nlm.nih.gov/12686707/.

[2] Assistant Secretary of Defense for Networks and Information Integration. *Electronic Records Management Software Applications Design Criteria Standard*. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodm/501502std.pdf. Accessed: Aug. 2020. 2007.

[3] A. A. Ataullah, A. Aboulnaga, and F. W. Tompa. "Records retention in relational database systems". In: *Proceedings of the 17th ACM conference on Information and knowledge management*. 2008, pp. 873–882.

[4] A. Ball. *Review of data management lifecycle models*. University of Bath, IDMRC, 2012.

[5] G. A. Brown. *CCPA Enforcement Begins Today*. https://www.natlawreview.com/article/ccpa-enforcement-begins-today. 2020.

[6] *California Consumer Privacy Act (CCPA)*. Accessed: Aug. 2020. 2020. URL: https://oag.ca.gov/privacy/ccpa.

[7] California Secretary of State. *Proposition 24: Official Voter Information Guide: California Secretary of State*. URL: https://voterguide.sos.ca.gov/propositions/24/.

[8] Centers for Medicare & Medicaid Services. *The Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. Accessed: Aug. 2021. 1996.

[9] Centers for Medicare & Medicaid Services and others. *The Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. Accessed: Sept. 2021. 1996. URL: http://www.cms.hhs.gov/hipaa.

[10] *CME Group Rule 536.B.1*. Accessed: Sept. 2021. 2020. URL: https://www.cmegroup.com/rulebook/files/cme-group-Rule-536-B.pdf.

[11] G. Edelman. *The fight over the fight Over CALIFORNIA'S Privacy Future*. 2020. URL: https://www.wired.com/story/california-prop-24-fight-over-privacy-future/.

[12] European Parliament and of the Council. *Regulation (EU) 2016/679 of the European Parliament and of the Council*. https://www.legislation.gov.uk/eur/2016/679. 2020.

[13] J. L. Faundeen et al. *The United States geological survey science data lifecycle model*. US Department of the Interior, US Geological Survey, 2013.

[14] Federal Trade Commission and others. *Children's online privacy protection act of 1998 (COPPA)*. Accessed: Sept. 2021. 1998.

[15] C. Fox. *Google hit with £44m GDPR fine over ads*. 2019. URL: https://www.bbc.com/news/technology-46944696.

[16] T. Garante. *Provvedimento correttivo e sanzionatorio nei confronti di TIM S.p.A. - 15 gennaio 2020 [9256486]*. 2020. URL: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9256486.

[17] J. Han, E. Haihong, G. Le, and J. Du. "Survey on NoSQL database". In: *2011 6th international conference on pervasive computing and applications*. IEEE. 2011, pp. 363–366.

[18] HHS Office of the Secretary and Office for Civil Rights. *Model Notices of Privacy Practices*. Accessed: Aug. 2021. 2013. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/model-notices-privacy-practices/index.html.

[19] A. Holmes. *California just passed a major privacy law that will make it harder for Facebook and Google to track people and gather data*. 2020. URL: https://www.businessinsider.com/prop-24-privacy-california-data-tracking-facebook-google-2020-11.

[20] International Data Sanitization Consortium. *Data Sanitization Terminology and Definitions*. Accessed: Feb. 2021. 2017. URL: https://www.datasanitization.org/data-sanitization-terminology/.

[21] B. Lenard, A. Rasin, N. Scope, and J. Wagner. "What is lurking in your backups?" In: Springer International Publishing, 2021, 401–415.

[22] V. Li. *Looking back on Zubulake, 10 years later*. 2015. URL: https://www.abajournal.com/magazine/article/looking_back_on_zubulake_10_years_later.

[23] N. L. of Medicine. *Date Lifecycle*. URL: https://nnlm.gov/data/thesaurus/data-lifecycle.

[24] A. Nicodemus. *Google fined $670K for violating GDPR's 'right to be forgotten'*. 2020. URL: https://www.complianceweek.com/gdpr/google-fined-670k-for-violating-gdprs-right-to-be-forgotten/29186.article.

[25] L. Pouchard. *Revisiting the data lifecycle with big data curation*. 2015. URL: https://www.researchgate.net/publication/305095078_Revisiting_the_Data_Lifecycle_with_Big_Data_Curation.

[26] J. Reardon, D. Basin, and S. Capkun. "Sok: Secure data deletion". In: *2013 IEEE symposium on security and privacy*. IEEE. 2013, pp. 301–315.

[27] *Regulation (EU) 2016/679 of the European Parliament and of the Council*. Accessed: Jun. 2021. 2020. URL: https://gdpr.eu/tag/gdpr/.

[28] N. Scope, A. Rasin, J. Wagner, B. Lenard, and K. Heart. "Database Framework for Supporting Retention Policies". In: *International Conference on Database and Expert Systems Applications*. (to appear). Springer. 2021.

[29] N. Scope, A. Rasin, J. Wagner, B. Lenard, and K. Heart. "Purging Data from Backups by Encryption". In: *International Conference on Database and Expert Systems Applications*. (to appear). Springer. 2021.

[30] Secretary, HHS Office of the and (OCR), Office for Civil Rights. *Your Rights Under HIPAA*. 2020. URL: https://www.hhs.gov/hipaa/for-individuals/guidance-materials-for-consumers/index.html.

[31] United States Congress. 28 U.S. Code §1732. Accessed: Aug. 2021. 1948. URL: https://www.law.cornell.edu/uscode/text/28/1732.

[32] United States Congress. *Gramm–Leach–Bliley Act, Financial Services Modernization Act of 1999*. Accessed: Aug. 2021. 1999.

[33] B. Wolford. *Everything you need to know about the "Right to be forgotten"*. 2020. URL: https://gdpr.eu/right-to-be-forgotten/.

[34] *Zubulake v. UBS WARBURG LLC*. 2005. URL: https://sosmt.gov/wp-content/uploads/attachments/E-ZubulakeV.pdf?dt=1519325634100.