Master Security Policy Research Paper: Data Retention

Lokesh Lankalapalli, Austin Rudolph

IFT 543: Security Policy and Compliance

Dr. Tatiana Walsh

December 1, 2024

Master Security Policy Research Paper: Data Retention

A Data Retention Policy defines how long an organization keeps its data, outlines the proper methods for disposal, and ensures compliance with legal and regulatory requirements. The purpose of this policy is to maintain a balance between retaining necessary information for business operations, legal obligations, or auditing while securely disposing of data that is no longer needed. This approach not only safeguards sensitive information but also optimizes storage and minimizes risks associated with retaining unnecessary data (Edwards, 2023). By addressing these aspects, a Data Retention Policy forms a critical part of an organization's overall data management and security framework.

The importance of a Data Retention Policy lies in its ability to protect organizational and customer data, maintain operational efficiency, and ensure compliance with industry standards and regulations. Effective data retention minimizes the risks of data breaches by securely disposing of outdated information and provides clear guidelines for retaining critical records. For example, the Sarbanes-Oxley Act (SOX) emphasizes the retention of financial records to ensure corporate accountability and prevent fraud (Stephens, 2005). Similarly, compliance with privacy-focused regulations like Health Insurance Portability and Accountability Act (HIPAA) and industry requirements such as PCI-DSS ensures data integrity, availability, and confidentiality across various sectors (Li et al., 2012). Implementing this policy helps the company not only meet these regulatory obligations but also foster trust with stakeholders.

Two key standards that come from a Data Retention Policy include the Data Classification Standard and the Secure Deletion Standard. The Data Classification Standard categorizes data based on its sensitivity, helping define retention periods for different types of information (National Institute of Standards and Technology [NIST], 2013). The Secure Deletion

Standard ensures that once data reaches the end of its retention period, it is permanently and irreversibly removed using methods such as encryption key destruction, as discussed by Li et al. (2012). Together, these standards provide a structured approach to handling data throughout its lifecycle, reducing the potential for breaches and ensuring compliance with legal requirements.

2.1 Demonstrating Mastery

A data retention policy ensures organizations manage their data responsibly by establishing clear guidelines on how long data should be stored and when it should be securely deleted. This policy is essential for protecting privacy, ensuring compliance, and improving operational efficiency. For instance, Cornell University's Policy 4.21 emphasizes the importance of retention practices that meet ethical and legal standards while reducing risks associated with data breaches or misuse (Cornell University, n.d.). Legal mandates like the Sarbanes-Oxley Act (SOX) require organizations to retain financial records for specific durations to ensure accountability, while frameworks such as NIST SP 800-53 advocate for secure data handling and disposal to minimize risks (NIST, 2020). These guidelines ensure a structured approach to data retention, aligning with industry best practices and legal.

The significance of a robust data retention policy lies in its capacity to mitigate legal and financial risks while fostering trust among stakeholders. Regulations such as HIPAA and industry requirements such as PCI DSS mandate the secure handling and disposal of sensitive information, including health and payment data, to protect privacy and maintain compliance. By encrypting stored data and conducting regular audits, as highlighted in various studies, organizations can reduce breaches and avoid penalties. A clear policy not only protects the organization but also enhances its reputation, demonstrating accountability and ethical responsibility in data management practices.

Data Lifecycle

To master a data retention policy, it is essential to understand its purpose, implications, and connection to the data lifecycle stages, which include Plan, Create, Manage, Use, Share, Collect/Reuse, and Destroy (University of Wisconsin, n.d.). Each stage of this lifecycle plays a vital role in ensuring that the organization's data handling practices are efficient, secure, and compliant with legal and regulatory requirements. By applying these stages with practical examples, businesses can develop a robust data retention policy that not only meets legal obligations but also supports operational efficiency and builds trust.

In the planning stage, organizations lay the groundwork for effective data retention by identifying the types of data they handle, setting retention schedules, and defining clear policies to guide the data's lifecycle. For example, a retail company might decide to retain customer purchase histories for five years to enhance marketing strategies while planning for the secure deletion of outdated data in compliance with GDPR. This stage ensures that data retention aligns with the organization's goals and legal requirements.

During the create stage, data is generated or collected through various means, such as customer transactions, patient records, or employee information. At this stage, it is crucial to classify the data based on its sensitivity and value. Each piece of data should have its own requirements attached to it. These requirements could also include legal ones. For instance, a hospital categorizing patient records as confidential ensures that these records are handled with heightened security measures, in line with HIPAA guidelines. Proper classification at this stage facilitates the appropriate application of retention and deletion rules later in the lifecycle.

In the manage stage, organizations store and organize the collected data, ensuring it is secure and accessible only to authorized individuals. This stage involves implementing access

controls, encryption, and backup strategies to protect data from breaches and ensure its integrity. For example, a financial institution managing customer transaction records ensures that these records are encrypted and stored in compliance with PCI-DSS standards, reducing the risk of unauthorized access.

The use stage involves leveraging data for its intended purpose, such as analyzing trends, making business decisions, or delivering services. For instance, a marketing team in a retail company might use customer purchase histories to create personalized campaigns. However, it is crucial to ensure that data usage adheres to privacy regulations and best practices, such as ensuring customers have consented to the use of their data for marketing purposes.

The share stage focuses on data transfer within or outside the organization. This could include sharing information with stakeholders, third-party vendors, or regulatory bodies. For example, a healthcare provider sharing patient data with a specialist must ensure that the transfer is encrypted and complies with HIPAA. This stage highlights the importance of maintaining data security and integrity during transmission.

In the collect/reuse stage, organizations gather insights from existing data to derive value. This may involve analyzing historical data to predict trends, improve services, or optimize operations. For instance, an e-commerce company might analyze past sales data to forecast demand for upcoming seasons. During this stage, organizations must ensure that data reuse is ethical, transparent, and compliant with legal requirements to maintain trust.

Finally, the destroy stage marks the end of the data lifecycle, where data that is no longer needed is securely and permanently deleted. This involves methods such as encryption key destruction, shredding physical documents, or overwriting digital files. Hospitals, for instance, might securely dispose of patient records after the legally mandated retention period to protect

patient privacy and comply with HIPAA. Proper destruction not only prevents unauthorized access but also demonstrates the organization's commitment to data security.

By following these lifecycle stages, organizations can implement a comprehensive data retention policy that supports compliance, operational efficiency, and cost optimization. Whether planning retention schedules, managing sensitive data, or securely destroying obsolete records, these stages ensure that data is handled responsibly, fulfilling its purpose while maintaining trust and security.

2.2 Arguing For Your Position

Cost Optimization

Cost optimization in a data retention policy is crucial for organizations to manage data storage expenses effectively while ensuring compliance and security. By implementing retention schedules, businesses can reduce the costs associated with storing unnecessary data, such as high-volume files that no longer serve a purpose. For example, cloud storage providers often charge based on the amount of data stored, and unnecessary retention of outdated data can lead to unnecessary costs. A study by Gartner found that 50% of enterprise data is unused, which could be better managed to reduce storage expenses (Ankush et al., 2020). Additionally, employing data classification standards to prioritize the retention of only critical data and applying secure deletion methods when data is no longer needed can streamline storage management and further lower costs. Thus, a well-structured data retention policy can significantly contribute to cost savings.

Imagine a healthcare organization that uses cloud storage to maintain patient records, billing information, and medical reports. Without a data retention policy, the organization decides to store all data indefinitely, regardless of its relevance or usefulness. Over the years, the

accumulation of outdated patient records, old billing details, and unnecessary reports causes their cloud storage costs to skyrocket. They find themselves paying significantly more each year for storing unused data that offers no real value to their operations. Recognizing the growing costs, the organization implements a data retention policy. They analyze the types of data they store and classify it based on legal, operational, and regulatory needs. For example, they decide to retain patient records for 10 years, complying with HIPAA requirements, while financial billing data is stored for seven years to meet audit standards. Reports and documents that are not legally required, like internal meeting notes or outdated project records, are scheduled for deletion after three years. Secure deletion methods, such as encryption key destruction, are used to ensure that data is erased permanently without compromising security. After putting the policy into action, the organization notices a significant reduction in its data storage requirements within the first year. This leads to considerable cost savings, which can now be reallocated to enhance cybersecurity measures, such as implementing advanced encryption systems or upgrading firewalls. The data retention policy not only helps the organization optimize costs but also ensures compliance with regulations and minimizes the risks associated with retaining unnecessary data. This proactive approach highlights the value of structured data management in achieving both financial efficiency and regulatory compliance.

Legal Benefits

A data retention policy becomes a necessity in the face of legal mandates for data retention. One such example is the Sarbanes-Oxley Act, which applies to all publicly traded companies in the United States. Sarbanes-Oxley has explicit requirements for data retention. It requires that companies keep documents related to their financial status retained for as long as they are relevant (Stephens, 2005, para. 8). Furthermore, audit records must be retained for seven

years (Stephens, 2005, para. 9). It is important for the organization to demonstrate that it has a proper data retention scheme in time for a Sarbanes-Oxley audit. This starts with the data retention policy. If these financial and audit documents are not retained, the organization would be breaking the law and could be subject to legal penalties. Other laws work in this manner as well. This includes HIPAA and "know your customer" laws such as the USA Patriot Act. Some laws require data deletion in certain circumstances such as the Gramm-Leach-Bliley Act (GLBA).The Federal Financial Institutions Examination Council, which enforces the GLBA, requires that financial institutions must have an information architecture that can remove or destroy data when it is no longer in use (Federal Financial Institutions Examination Council, 2021, p. 16). Without a data retention policy, the organization will not be able to support compliance with these regulations and can be subject to legal penalties. There are also similar requirements placed by certain industries that are not based on legislation. This includes the Payment Card Industry Data Security Standard (PCI DSS), which has specific retention and deletion requirements. For example, cardholder data is meant to be deleted after it is no longer necessary unless explicit authorization is given by the cardholder (PCI Security Standards Council, 2008). If these are not followed, the other companies in the payment card industry will not do business with the organization. This means a loss in partners and business. Legal and industry penalties will cost the company money and potentially employees as well. Therefore, there is a large risk associated with not implementing a data retention policy. Implementing a data retention policy will then allow the organization to support compliance.

Supporting Other Policies

The data retention policy has the ability to support other organizational security policies. For example, data retention supports data backups. When backups are made, the data retention

policy will be able to dictate how long that backup is meant to be retained and when it is scheduled for deletion. The team dictating the parameters for how long a backup will need to be stored will then translate into the categorization used by the data retention policy. With both of these policies in place, backups will be marked on how long they are meant to be kept. This then allows the data backup policy to operate while reaping the other benefits of the data retention policy such as cost optimization. In this case, the organization will not be keeping old, unnecessary backups, as they are not necessary for the functioning of the data backup policy and would actively cost money to maintain. This is especially the case when there are newer backups available. This also indirectly serves to help support the disaster recovery policy, which seeks to restore to a working backup—one that has been retained due to the data retention policy.

The data retention policy also supports the monitoring and logging policy. Retaining monitoring and logging data provides a history of the actions taken against a cyber asset. This history can then be used for forensic analysis. However, because the data retention policy also implies deletion, these two policies can conflict with one another, as mentioned by Lu et al. (2013). This will be discussed further in the next section.

2.3 Potential Objections

Users

One objection to the data retention policy is that data retention and deletion is not valued by users. Many users simply click through privacy policies and terms of service agreements without reading them. This could indicate that users do not care about how long their data is retained and if it is deleted. If this is true, it could mean that it is not worth it for the organization to expend resources to figure out which user data should be deleted and then delete it. It could also mean that the data could be reused in the future for other purposes.

Users might simply not be aware of the consequences of the lack of a defined data retention period. If user data were stored indefinitely, this means that a breach involving user data will be larger. Deleted data is not able to be breached provided it was destroyed securely. Users caught up in a data breach might be at increased risk of identity theft, which means that they will be affected negatively by the organization not implementing this policy. Any breach also reflects poorly on the organization in the public eye. Another aspect of user data is that some users truly do care about how much of their data is stored and how long it is kept. The GDPR in the European Union introduces a "right to be forgotten." This means that users can request what data is stored on them and to have it deleted. Blanchette & Johnson (2002) argue that proper data deletion has positive effects for society at large, allowing individuals to put a potentially negative past behind them. They accuse infinite retention periods as being panoptic and that policies should be in place to curtail this (para. 1). Therefore, implementing this policy creates an amount of organizational-level "forgetfulness" that can be seen as socially responsible even in jurisdictions outside of the purview of the GDPR. This policy can then serve to support any corporate social responsibility statements.

Distributed Computing

Distributed computing offers challenges for data deletion. If data is stratified across many different machines that are backed up or have redundancy in place, it can be difficult to know for sure whether or not data is actually deleted. Deleting a file could mean that there are more copies of it on other systems. Therefore, having a data retention period followed by data deletion could be very difficult and require more resources to pull off in the age of distributed computing.

While distributed computing does make deletion more difficult, there are still solutions available. One such example comes from Li et al. (2012). They came up with an example that

categorizes all data, encrypts each file with a different key, and then stores encrypted copies on an HDFS cluster. This cluster is then able to decrypt files when needed. The keys are managed and indexed centrally. When the file needs to be deleted, the key is deleted first. While the encrypted files will still remain, they will be unrecoverable because the key to decrypt them will have been deleted (Li et al., 2012). Centrally managing the keys sidesteps the issues associated with distributed computing. This is possible because these encryption keys are small compared to the data that ends up encrypted. Although this example uses HDFS and Apache Hadoop, these principles could be used in another type of system as well. Other ideas could be taken into account such as backup systems that periodically take the difference of the current storage on a live system and the current storage on a backup system, and then update accordingly in one direction only (from live to backup). This way, when data gets deleted, it eventually gets deleted from the backup as well. If it is not urgent that this data is deleted and it can persist in backup as plain text, the encryption scheme does not even need to be used for it. This policy will necessitate the drafting of procedures for exactly how to perform distributed, scaled deletion in the organization.

Valuable Data and Reuse

Deleting data could mean that valuable data is deleted as well. It is possible for the organization to decide to delete data that it maybe should not have. Additionally, data can often be reused for purposes other than the one(s) it was collected for. This includes service extensions or even sales. If this data were deleted, the organization could lose out on the value that this data holds, which could be realized by implementing new services or selling it directly.

The data retention policy will mean that data must be attached to a particular lifecycle based on the requirements for that data specifically. Requirements could be based on particular

legal requirements. For example, documents regarding the financial condition of the organization could be tagged with a lifecycle that implies Sarbanes-Oxley requirements. This reduces the likelihood of data being deleted by accident or while it is still useful. It may be difficult to know when data might stop being useful or if it could have use in the future that is currently unknown to the organization. However, it should be noted that storing data is not free. The organization can either reap the costs of data storage with the uncertainty of reaping the benefits of future data usability, or the organization can reap the benefits of data retention and prompt deletion with the uncertainty of reaping the costs of losing future data usability. The idea that data could be useful in the future is uncertain, so it is not worth spending extra on storage for it. Not to mention, storing data past the point of usefulness is not considered good practice. Using personal data for things such as "testing, research, and training increases the risk of unauthorized disclosure or misuse of the information" (NIST, 2013, p. J-15). This could also be applied to sales, which means that it is not advisable for the organization to sell data that has no longer become useful. Sold information getting breached or misused could reflect poorly on the organization.

Monitoring and Logging

Similar to the previous objection, monitoring and logging data could be deleted while it is still useful. Lu et al. (2013) states that "auditing the changes to a database is critical for identifying malicious behavior, maintaining data quality, and improving system performance" (para. 1), which could apply to more appliances than just databases. This auditing creates records that will be subject to a data retention policy. Deleting these records could restrict the ability of the organization to gather forensic evidence of malicious behavior, especially if that behavior initiated potentially even years prior to identification.

It is true that the monitoring and logging policy is somewhat in opposition to the data retention policy. It is also true that there is the potential for an attack to execute itself slowly over the course of many years, which could be longer than the logging data's retention period. However, this highlights the need for procedures to define exactly what the retention periods are for monitoring and logging data and why. These procedures should be drafted by subject matter experts and risk analysts. This way, the risks of storing data for too long and the risks of not being able to gather forensic evidence are mitigated the most that they can be. These procedures will get into specific methods for how to reduce data according to the lifecycle it is attached to. Lu et al. (2013) gives some examples specific to databases such as redaction and expunction (p. 210). Because this is similar to the previous objection, it also applies here that these logs ultimately have unknown value. The organization should not try to extract this unknown value by not implementing a proper data retention policy with deletion. There will still be a trade-off between these two policies and their goals, but ultimately the organization needs to reap as much of the benefits from both policies as it can.

3. Conclusion

In conclusion, a clear data retention policy is an important part of managing data in any organization. It helps businesses decide how long to keep different types of data, when to safely delete it, and how to do so. This is crucial for meeting legal requirements and reducing risks that come with holding onto unnecessary data. Following such policies also protects sensitive information and helps minimize the impact of data breaches that would compromise user data. Additionally, it supports the organization's compliance with privacy regulations like GDPR and GLBA which mandate data disposal.

Having a data retention policy also brings practical benefits, such as saving money. By keeping only the data that is necessary, companies can reduce storage costs and make it easier to manage important information. This approach not only saves money but also improves efficiency by ensuring that data is stored in a way that adds value to the business. When combined with other strategies like backup and cybersecurity measures, it helps strengthen the organization's overall security and operations.

Compliance requirements introduced by legislation such as HIPAA, Sarbanes-Oxley, etc. as well as by industries such as PCI DSS also necessitate this policy. Without proper data retention, the organization could be breaking the law. This puts the organization at risk of legal penalties or sanctions from partners (as would be the case for PCI DSS), which are unacceptable.

Being part of a holistic understanding of security, the data retention policy also supports other policies. This includes the data backup policy, which will require that backup data is to be retained for a specific period of time and then deleted. It also includes the monitoring and logging policy, which requires that logging data is retained for a specific period as well. Although it can be difficult to decide how long this period should be, implementing both of these policies is important in order to reap all of the benefits.

While creating and following a data retention policy can be challenging, especially when dealing with user habits and complex systems, it is still very important. Some employees may not fully understand or follow the rules, and deleting data from all systems can be tricky. However, these problems can be solved with proper training, clear communication, and the right tools. This policy will influence the drafting of various standards and procedures for exactly how data retention works within the organization, how employees that have the retain and delete data are to do their jobs, and what the process is for audits that ensure proper data retention and

deletion is taking place. A good data retention policy protects the company from legal penalties, saves money, and shows that the company values ethical data management.

References

Ankush, J., De Simoni, G., Thoo, E., Ronthal, A., Chien, M., Feinberg, D., Zaidi, E., Parker, S., Walker, S., Hawker, M. (2020, June 22). *Cost optimization is crucial for modern data management programs*. Gartner. https://www.gartner.com/en/documents/3986583.

Blanchette, J.-F., & Johnson, D. G. (2002). Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness. *The Information Society*, *18*(1), 33–45. https://doi.org/10.1080/01972240252818216

Cornell University. (2022). *Policy 4.21: Research data retention*. Cornell University Policy Office. https://policy.cornell.edu/policy-4-21-research-data-retention

Edwards, J. (2023, August 22). *How to build a strong and effective data retention Policy*. InformationWeek. https://www.informationweek.com/data-management/how-to-build-a-strong-and-effective-data-retention-policy

Federal Financial Institutions Examination Council (2021, June). *Architecture, infrastructure, and operations*. FFIEC Information Technology Examination Handbook. https://ithandbook.ffiec.gov/media/ywfm2ftz/ffiec_itbooklet_aio.pdf

Li, J., Singhal, S., Swaminathan, R., & Karp, A. H. (2012). Managing Data Retention Policies at Scale. *IEEE eTransactions on Network and Service Management*, *9*(4), 393–406. https://doi.org/10.1109/TNSM.2012.101612.110203

Lu, W., Miklau, G., & Immerman, N. (2013). Auditing a database under retention policies. *The VLDB Journal*, *22*(2), 203–228. https://doi.org/10.1007/s00778-012-0282-x

PCI Security Standards Council (2008). *PCI Data Storage Do's and Don'ts*. PCI Security Standards Council. https://listings.pcisecuritystandards.org/pdfs/pci_fs_data_storage.pdf

National Institute of Standards and Technology (2013, April). *Security and Privacy Controls for Federal Information Systems and Organizations* (Revision 4). NIST. http://dx.doi.org/10.6028/NIST.SP.800-53r4

National Institute of Standards and Technology (2020, September). *Security and privacy controls for information systems and organizations* (Revision 5). NIST. https://doi.org/10.6028/nist.sp.800-53r5

Stephens, D. O. (2005). The Sarbanes-Oxley Act: Records management implications. *Records Management Journal, 15*(2), 98-103. https://doi.org/10.1108/09565690510614247

University of Wisconsin. (n.d.) *Introduction to the data lifecycle*.. Data, Academic Planning & Institutional Research. https://data.wisc.edu/data-literacy/lifecycle/#:~:text=Data%20lifecycle%20stages,Collect%2FReuse%2C%20and%20Destroy.