data

MDPI

Article

Understanding Data Breach from a Global Perspective: Incident Visualization and Data Protection Law Review

Gabriel Arquelau Pimenta Rodrigues [1], André Luiz Marques Serrano [1], Amanda Nunes Lopes Espiñeira Lemos [2,3], Edna Dias Canedo [1], Fábio Lúcio Lopes de Mendonça [1], Robson de Oliveira Albuquerque [1,4], Ana Lucila Sandoval Orozco [1,4] and Luis Javier García Villalba [4,*]

1 Professional Post-Graduate Program in Electrical Engineering (PPEE), Department of Electrical Engineering (ENE), University of Brasília (UnB), Brasília 70910-900, Brazil; gabriel.arquelau@redes.unb.br (G.A.P.R.); andrelms@unb.br (A.L.M.S.); ednacanedo@unb.br (E.D.C.); fabio.mendonca@redes.unb.br (F.L.L.d.M.); asandov@ucm.es (A.L.S.O.)
2 Graduate Program in Law (PPGD), Law School, University of Brasilia (UnB), Brasília 70910-900, Brazil; amandaespineira@ccom.unb.br
3 School of Law, University of Minho (EDUM), Campus de Gualtar, 4710-057 Braga, Portugal
4 Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), Faculty of Computer Science and Engineering, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, Spain
* Correspondence: javiergv@fdi.ucm.es

Abstract: Data breaches result in data loss, including personal, health, and financial information that are crucial, sensitive, and private. The breach is a security incident in which personal and sensitive data are exposed to unauthorized individuals, with the potential to incur several privacy concerns. As an example, the French newspaper Le Figaro breached approximately 7.4 billion records that included full names, passwords, and e-mail and physical addresses. To reduce the likelihood and impact of such breaches, it is fundamental to strengthen the security efforts against this type of incident and, for that, it is first necessary to identify patterns of its occurrence, primarily related to the number of data records leaked, the affected geographical region, and its regulatory aspects. To advance the discussion in this regard, we study a dataset comprising 428 worldwide data breaches between 2018 and 2019, providing a visualization of the related statistics, such as the most affected countries, the predominant economic sector targeted in different countries, and the median number of records leaked per incident in different countries, regions, and sectors. We then discuss the data protection regulation in effect in each country comprised in the dataset, correlating key elements of the legislation with the statistical findings. As a result, we have identified an extensive disclosure of medical records in India and government data in Brazil in the time range. Based on the analysis and visualization, we find some interesting insights that researchers seldom focus on before, and it is apparent that the real dangers of data leaks are beyond the ordinary imagination. Finally, this paper contributes to the discussion regarding data protection laws and compliance regarding data breaches, supporting, for example, the decision process of data storage location in the cloud.

Keywords: compliance; data breach; data protection regulation; information security; privacy

Citation: Pimenta Rodrigues, G.A.; Marques Serrano, A.L.; Lopes Espiñeira Lemos, A.N.; Canedo, E.D.; Mendonça, F.L.L.d.; de Oliveira Albuquerque, R.; Sandoval Orozco, A.L.; García Villalba, L.J. Understanding Data Breach from a Global Perspective: Incident Visualization and Data Protection Law Review. *Data* **2023**, *9*, 27. https://doi.org/10.3390/data9020027

Academic Editor: Kesheng (John) Wu

Received: 11 December 2023
Revised: 3 January 2024
Accepted: 23 January 2024
Published: 31 January 2024

1. Introduction

In an information-rich world, data have become a valuable asset for businesses across all industries [1]. If effectively collected, analyzed, and utilized, data can enhance decision making, optimize operations, and propel business growth, empowering companies to understand customer behavior, personalize marketing campaigns, and develop innovative products and services [2]. This applies to both private entities and the public administration.

While the benefits of data are evident, they also conceive associated implications, particularly regarding security and data subject privacy [3]. For the data subjects, unauthorized access to personal data, such as social security numbers, dates of birth, and addresses,

may result in identity theft and physical safety [4]. Depending on the type of data stored, additional concerns emerge, such as credit card information leakage promoting financial frauds [5] and breach of medical records disclosing sensitive diagnosis [6].

Likewise, for companies and public entities, unintentional data disclosure may imply reputational damages [7], lawsuits [8], and direct financial losses [9]. Due to these impacts, the average cost of a data breach in organizations with high-security skills shortages is USD 5.36 million [10].

Because of these repercussions for both the data subjects and the data owner, businesses and governments must deploy security controls to safeguard sensitive information from unauthorized access, breaches, or misuse [11]. Implementing robust cybersecurity measures, such as encryption, access controls, and regular security audits, is essential to protect both the company's data and the privacy of its customers [12].

Organizations took 204 days as the mean time to identify a data compromise in 2023, representing only marginal changes from the numbers of previous years [10], which emphasizes the need for a more efficient deployment of detection mechanisms. For that, an improved comprehension of trends in data breaches is necessary. To advance this discussion, we conduct a visualization approach for enhancing the understanding of data leak patterns worldwide, which advances the planning for security control employment. This work provides a comprehensive overview of data breaches and offers valuable insights into preventing and mitigating these incidents. Additionally, the overview of data protection laws associated with the observed statistics contributes to the enrichment of the discussion regarding the regulatory landscape. This paper's contributions can help organizations and governments of all sizes and industries to protect their data and their customers.

A geographical-oriented discussion and statistics visualization related to data compromises is also relevant as the data on the cloud may be stored abroad, and incident trends and regulatory analysis may be suitable for protecting the data appropriately [13]. As the foundation of this study, we use a dataset of data breaches with at least 30,000 affected records that occurred globally between 2018 and 2019 [14].

The remainder of this work is structured as follows: Section 2 reviews the related literature, Section 3 provides a visualization of the statistics regarding the data breaches present in the studied dataset, Section 4 discusses some incidents specifically, Section 5 reviews the applicable data protection laws of the affected countries in the dataset, and Section 6 presents some of the threats to the validity of the study. Section 7 concludes the paper.

2. Literature Review

Data protection frameworks may not be adequately structured to regulate the collection and processing of user data using Internet of Things (IoT) devices. For instance, the European General Data Protection Regulation (GDPR) may not properly protect the data processed by this technology [15]. This raises security concerns and users' privacy issues, especially considering the low-security awareness of IoT consumers and the possibility of data breaches [16].

Considering the applicability of IoT devices in the healthcare environment [17], medical records may be leaked, exposing sensitive data such as medical treatments, vaccines taken, diagnoses, and exam results. Due to its sensitivity, several authors have studied breaches related to these types of data and their implications in the regulatory realm [18,19].

While continuing the analysis of medical privacy violation, statistical visualization assists in uncovering key insights and trends within these incidents. As an example, it has been observed that healthcare organizations are mainly breached due to hacking activities, indicating a weakness in the technological countermeasures deployed by companies in this sector [20]. Alfawzan et al. [21] studied 23 mobile women's health apps, and evaluated them on their privacy and data collection policies, remarking that some of them collect the user's data without consent and that most of them share the data with third parties. These

practices inflict consent requirements usually present in data protection laws, including the General Data Protection Regulation [22].

Strupczewski [23] performed an exploratory data analysis on a data breach dataset focused on the United States, enabling an observation of the influence of factors such as the economy sector of the company, operating geographical region, and the cause of the incidents on its severity. The authors have also remarked that, in the time frame of the studied dataset, comprehended between 2005 and 2016, more than half of the data compromises affected the healthcare sector, and negligent data exposure happens twice more frequently than malicious ones.

In addition to incident pattern identification, threat intelligence methods also aim to gather information on the attacker intentions, trends, and procedures, enabling the organization to better protect its informational assets [24]. However, the effective implementation of threat intelligence faces some challenges, such as the complexity of the threat landscape, the big data scenario, and the lack of transparency and visibility [25]. As an example of a threat intelligence framework for data breaches, Noor et al. [26] proposed a machine learning model that achieved a 92% detection rate, with an average detection time of 0.15 s.

Data disclosure may negatively affect the stock prices of the breached company, and its recovery may depend on factors such as its age, size, profitability, and brand familiarity [27]. A tendency of short sellers to exploit prior knowledge of a data breach occurrence [28] has also been observed. To reduce the frequency and repercussions of a data leak, countermeasures must be effectively adopted and tested [11]. Machine learning algorithms are also applicable to prevent and deter these incidents [29].

It is also relevant to study the changes in the behavior of data subjects after a breach. It is observed that, after an incident of this type, users of a website seeking extramarital affairs tend to decrease the messaging frequency and delete their posted pictures, returning to normal usage after three weeks [30]. In a specific data exposure case that affected the U.S. Office of Personnel Management (OPM) in 2015, Twitter (now X) users expressed anxiety, anger, and sadness when commenting about the event, indicating a situational awareness of the incident [31]. Companies also tend to report a data breach in complicated language and in lengthy paragraphs [32], which may compromise its readability and understanding by the affected users.

On the data protection regulation and beach incidents, Alazab et al. [33] studied the Australia's Notifiable Data Breach scheme, referring to a greater responsibility given to data subjects while data protection entities are treated with more leeway.

Focusing on the United States, Kesari [34] concluded that the California data breach notification 2016 law reduced identity theft crime cases, but suggested that more studies should be conducted to confirm the findings.

Authors have also studied privacy and data protection laws within the jurisdiction of some of the countries mentioned in this work, such as in China [35], in India [36], in Brazil [37], in Singapore [38], in Israel [39], and in the European Union [40].

3. Dataset Visualization

The dataset studied in this paper is provided by Neto et al. [14]. For the recognition of patterns of occurrences of data breaches, we present visualizations of the dataset according to their geographical and economic sector incidence and to their sizes.

3.1. Data Collection Methodology and Considerations

According to Neto et al. [14], the database creation was based on publicly available sources from government entities, security research groups, research entities, and media reports in several languages. An incident was only included in the dataset if confirmed by multiple sources or if a source provided evidence of its occurrence. The incidents were also filtered to include exclusively data breaches reported between January 2018 and December 2019 that exposed at least 30,000 records. This resulted in a dataset comprising 428 incidents,

which can be accessed through a web page[1]. For the visualization, we use Python and RAWGraphs [41] to generate the charts.

For each incident, the dataset informs the year of occurrence, the affected company, and its sector, country, geographical region, the number of records leaked, and the source of the information. The dataset presents, in total, 37 affected countries situated in North America, South America, the Caribbean, Europe, Asia Pacific, and Africa. The sectors of the organizations are education, government, and military; medical and healthcare; business; entertainment; technology; and banking, credit, and financial. The statistics of the number of records disclosed for the complete dataset are presented in Table 1.

Table 1. Descriptive statistics of the number of records leaked in the dataset

	Breach Size
count	428
mean	61,673,880
std	400,573,400
min	30,000
25%	74,375
50%	422,548
75%	6,000,000
max	7,400,000,000

The incident that exposed the most data in the dataset targeted the French newspaper, Le Figaro, exposing 7.4 billion records. This disclosure is almost as big as one of the biggest reported data breaches that affected Cam4 in 2020, exposing 10.8 billion records [42]. Also, the sum of all disclosed records, 22 billion, is greater than the world population in 2019, 7.7 billion people. This may be due to breaching records of deceased people and the same person having their data leaked multiple times [14].

3.2. Records Leaked per Country and Region

When a company adopts the cloud model, for example, an important aspect to be considered before hiring the provider is where the data will be located [43]. This computing model enables data to be stored across the borders of the customer country, and new challenges emerge, especially related to security and regulation [44].

Cloud data breaches are a rising concern [45] and, therefore, the choice of the location of the stored cloud data should consider the occurrence of this type of incident. To promote the discussion regarding cloud and related location decisions, we present statistics regarding geographical occurrences of data leaks. The majority of breaches occurred in the United States, as shown in Figure 1, which displays the ten countries with the highest number of incidents, disregarding the number of records breached.

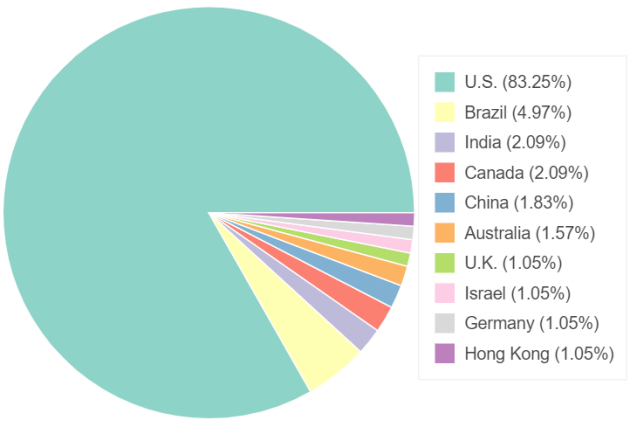

Figure 1. The ten most frequently breached countries.

Despite Brazil appearing in the second position in Figure 1, it is not a significant country when considering the size of the breaches. Figure 2 depicts the total amount of records breached for every country in the dataset, demonstrating that, regarding this metric, France and the United States were the most expressive countries, followed by China and India.

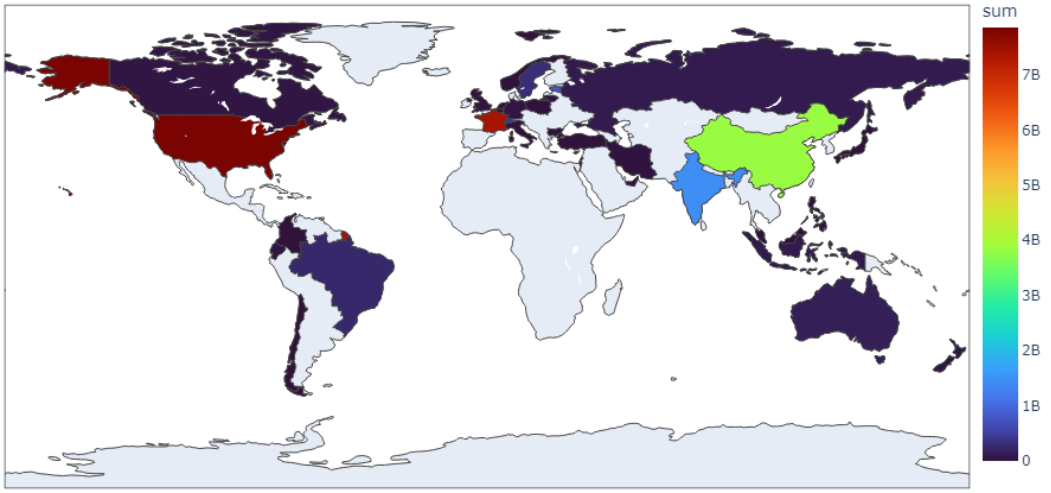

Figure 2. Sum of breached records per country.

The fact that France is among the countries with the highest number of records leaked (Figure 2) while not being in the group of the ten most breached companies per incident count (Figure 1) suggests that it had a fewer number of incidents with higher severity, breaching a greater amount of data per incident.

This conclusion is confirmed by Figure 3, which indicates that France had the highest median in the size of records breached per incident among all countries in the dataset. Countries like Estonia, Switzerland, and Sweden, although breaching a relatively high number of records, were only breached once. Oppositely, the median of the number of records breached in the United States is lower than that of Australia, for example, which did not present a relevant sum of breached records in Figure 2. This is due to the fact that the U.S. reported a great number of incidents, contributing to a higher sum of records leaked.

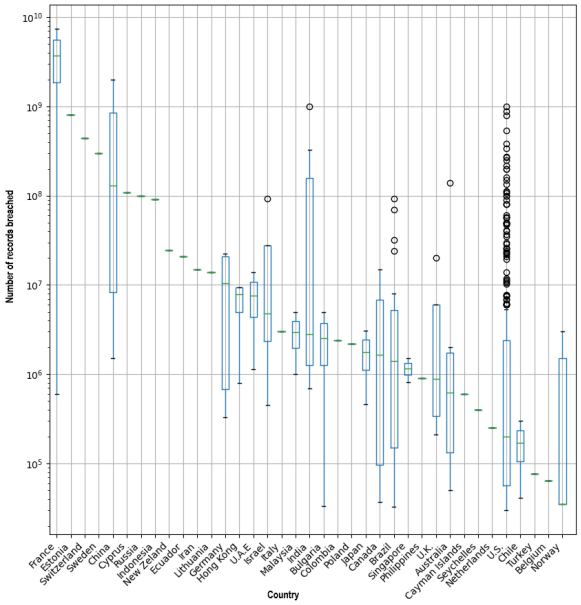

Figure 3. Boxplot of breached records per country.

For an improved view of the proportions of the sum of leaked records for each country in its region, Figure 4 indicates that the European region had the highest number of records leaked, mainly led by France.

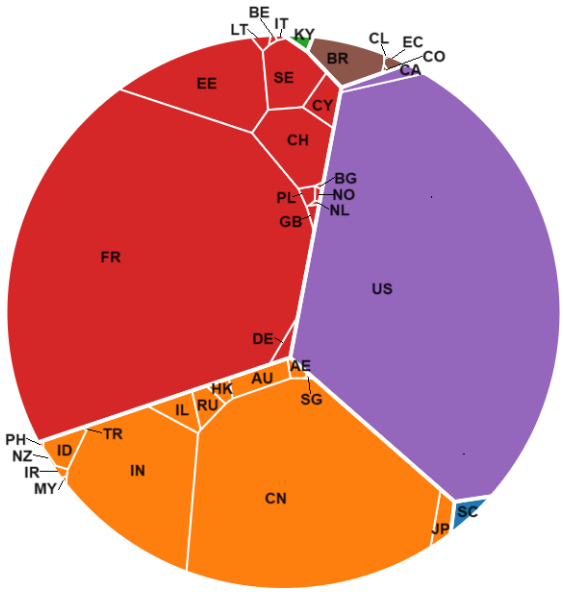

Figure 4. Proportion of sum of records breached per country in regions.

Likewise, North America, the second most frequently breached region, is led by the United States, with a few occurrences in Canada. In the Asia Pacific region, China and India represent the countries with the most records breached, with some other countries contributing with smaller amounts. South America and Africa mostly comprise data leaked in Brazil and Seychelles, respectively, while the Caribbean region is represented exclusively by the Cayman Islands.

Asia Pacific, despite being only the third in the sum of information breached, is the region with the highest median in size of breaches, as noted in Figure 5. North America was the region with the lowest median, mainly due to the United States statistics.

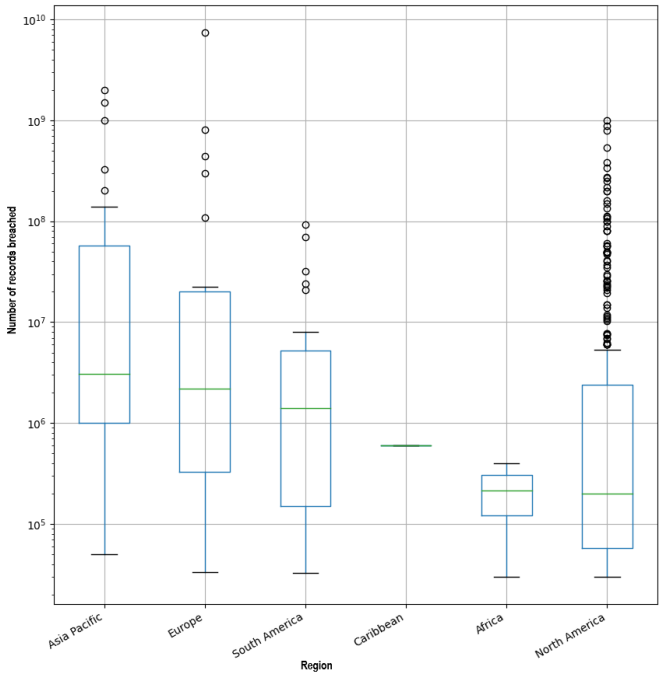

Figure 5. Boxplot of breached records per region.

3.3. Records Leaked per Sector

Different types of disclosed data impact the lives of breached subjects differently. For example, healthcare data disclosure may impact confidential medical information [46], and military breaches possibly uncover a country's armed forces' sensitive data [47].

Furthermore, the nature of the stored data also determines regulatory aspects that the company must comply with [48]. The legal aspects of data breaches are more deeply discussed in Section 5. As shown in Figure 6, the technology sector presents the highest median in the number of records leaked per incident, followed by government/military. The most voluminous breach in the dataset, however, occurred in the business sector. Education had the lowest median. However, understanding the geographical and temporal distributions of these leaked records may reveal trends and patterns in the occurrence of this type of security incident. Figure 7 depicts the sum of records leaked per region for each sector in 2018 and 2019.

The figure shows that the majority of incidents took place in 2019. As stated by [14], this may be due to the GDPR entering into force in 2018 and companies being compliant with breach notification requirements throughout 2019. An exception to this observation is the healthcare sector, with the prevalence of records breached in 2018. However, the sector with the highest volume of information leaked in 2018 was technology, surpassed by the business sector in 2019.

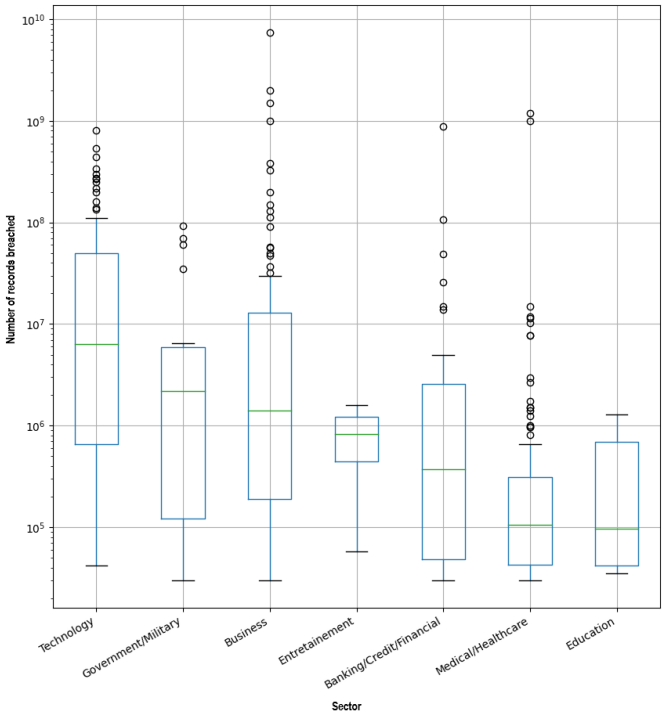

Figure 6. Boxplot of breached records per sector.

In addition, business was the economic sector with the most records breached when considering both years, despite being the third highest median of records breached per incident, indicating that organizations in this branch suffered more breaches that leaked fewer records than technology and government, for example.

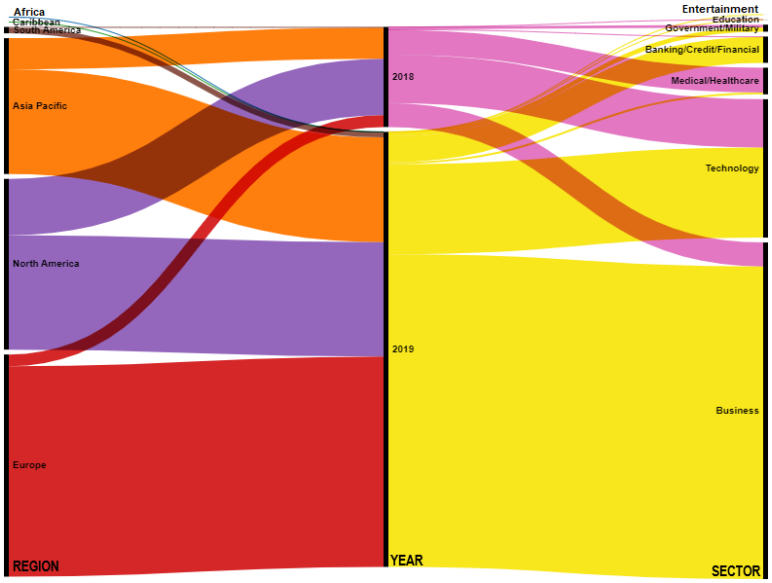

Figure 7. Sum of records breached per region per year per sector.

It is relevant, however, to interpret the most targeted sector in each country, as it may evidence technical weaknesses in different areas for different regions and emphasize the necessity for stronger regulation and employment of more effective security controls.

In that regard, Figure 8 presents the relationship between the economic sectors in which the breached companies operate and the countries where the incident occurred. For this analysis, we considered exclusively the ten most frequently breached countries (Figure 1), adding France, as it represents a significant portion of breach sizes (see Figure 4).

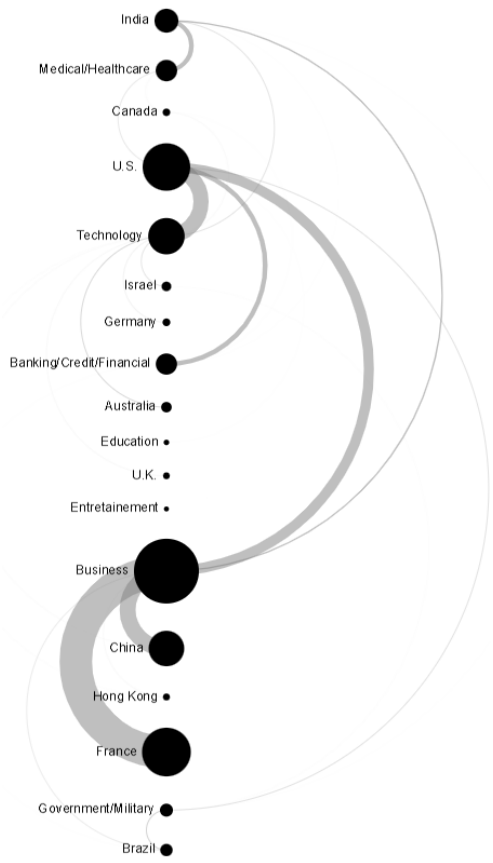

Figure 8. Most explored sector in the top 10 breached countries.

From Figure 8, it is observed, for example, that France and China were mostly breached in the business sector, which also comprises a significant amount of incidents from the United States. Notwithstanding, the most significant breaches in the U.S. affected the technology sector, which may also be a consequence of the significant number of technology companies based in the country. It is also observed that the banking, credit, and financial sector were mostly breached in the United States, which demonstrates the significance of the Gramm–Leach–Bliley Act. This pattern demonstrates a trend change as, from 2010 to 2017, the U.S. registered mostly healthcare breaches [49].

Additionally, healthcare incidents affected mostly Indian organizations. Churi et al. [50] acknowledges some privacy issues related to the healthcare sector in India, namely lack of technology and infrastructure, absence of trust in the relationship between doctor and patient, medical data being stored in the cloud with privacy concerns, weak security controls deployed, data shared without subject's consent, inadequacy of security policies, and cultural aspects. This may impact the trust and acceptance of healthcare technologies of citizens from this country [51].

Government and military disclosed predominantly records from Brazil and the United States. In Brazil, Ferrão et al. [52] indicated a lack of compliance to the General Data Protection Law (LGPD) and immaturity in the area, with many organizations that had not yet established a Data Protection Officer (DPO). Specifically to the military and defense sector, Brazil has demonstrated a lack of attention toward national security policy making [53], which may have effects on cyber security incidents in this sector.

Government data breaches may also emerge as a consequence of geopolitical conflicts. As examples, Shires [54] mentions four cases affecting political people and public figures in the U.S. in hack-and-leak operations.

For a more globally comprehensive reasoning of this relationship, Figure 9 presents the number of records breached per sector in each region. The business sector suffered significantly voluminous breaches in Europe, North America, Asia Pacific, and South America, and technology was also relevant in Europe and North America. In the Asia Pacific, medical and healthcare incidents also correspond to significant portions, mainly due to the incidents in India, as seen in Figure 8. Government and military is the second main sector affected in South America, with a great contribution from Brazil.

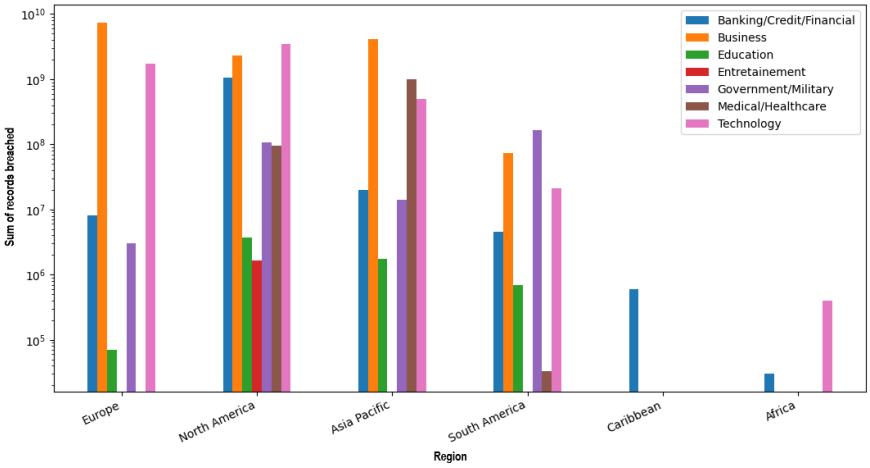

Figure 9. Distribution of sum of records breached by sectors per region.

The only incident that took place in the Caribbean inflicted the Cayman National Bank, and African countries suffered breaches in the financial and technology areas. Additionally, entertainment breaches were reported in North America exclusively, related to the incidents involving AMC Networks and MoviePass. Europe, despite being the region with the most breaches records, did not publicly register any data leakage incident in the healthcare sector and was also the region with the fewest education records disclosed among the localities that were affected in this area.

The size of data breaches in the technology and business sectors are more evenly distributed, especially in the range between 10^5 and 10^9 records breached per incident, as seen in Figure 10. Medical data violations, differently, are more concentrated at around 10^5, with few occurrences of 10^9 records leaked and none around 10^8.

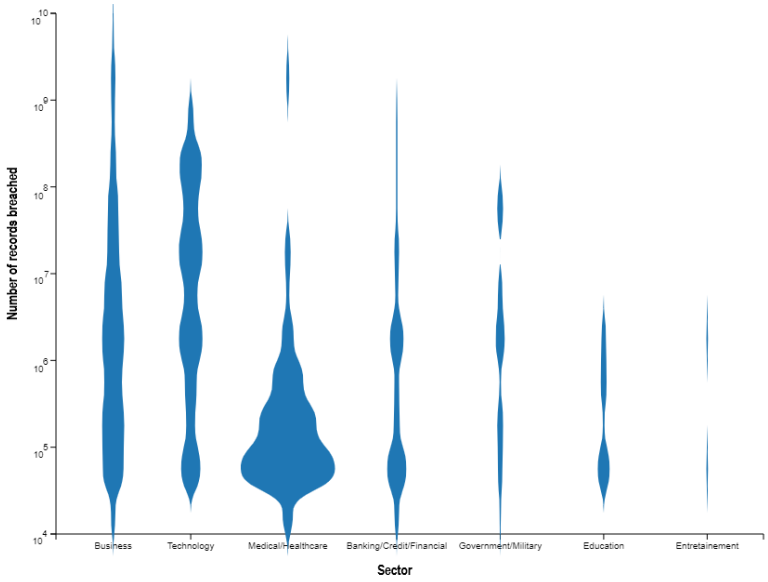

Figure 10. Distribution of breach sizes per sector.

4. Study Cases and Possible Mitigation

For a more thorough discussion on the data protection regulation on the affected countries, implications of data exposures, and applicable security controls used to mitigate them, we study some cases that are present in the dataset.

In Europe, British Airways was fined USD 229 million by the UK Information Commissioner's Office (ICO), corresponding to 1.5% to 2% of the annual revenue of the company at the time [55]. This significant financial impact, along with the privacy issues faced by the data subjects, emphasizes the importance of complying to data protection laws and implementing security controls.

One of such controls is an effective cryptographic mechanism. The biggest incident reported in India affected Aadhaar and exposed one billion records in January 2018. This breach is categorized as in the medical sector by Neto et al. [14], possibly due to its relation to health services [56], including vaccines [57], despite Aadhaar being a biometric database also used for demographic, financial, and welfare policy entitlement data [58]. According to the reference linked to the incident in the dataset, Aadhaar actually suffered multiple breaches between 2017 and 2018, which may be due to known vulnerabilities, such as cryptographic issues observed in the system [59]. As countermeasures, besides a stronger encryption scheme, Tyagi et al. [60] also suggest the adoption of security testing, the creation of a Computer Emergency Response Team (CERT), and a more effective integration with the private sector and their standards. Following the implementation of these controls, the authors also recommend initiating open security challenges.

There were two reported incidents in Japan, both within the business sector. The largest involved Toyota, which experienced a breach affecting over 3 million records in 2019. This case draws attention to the expanding field of self-driving cars, which require big data for effectiveness and penalization [61]. A lack of protection on these data may disclose geolocation trends of the users, raising safety concerns. Previous data leaks in the automotive industry occurred as a consequence of vulnerabilities in products provided by third party manufacturers, which could have been prevented with supply chain management [62].

In Brazil, a relevant data breach refers to the incident involving the Unified Health System (SUS - Sistema Único de Saúde), which is Brazil's public healthcare system. According to the official portal of the Brazilian Ministry of Health[2], the incident occurred due to the leakage of an access credential to the system. This compromised personal data such as the individual taxpayer registry number, name, mother's name, gender, race, date of birth, blood type, nationality, and date of death. The sensitiveness of this data, along with the the attack vector on this case, reinforces the importance of a secure credential storage with the use of strong hashing and salting algorithms [63], multi-factor authentication [64], and a password breach alerting system [65].

Additionally, to enhance identity protection, the use of common passwords should be avoided. The Zynga breach in 2019, in the United States, exposed 218 million records, and revealed that the most used passwords included strings such as `password`, `12345`, `changeme`, and `qwerty` [66]. The enforcement of a password policy, requiring more secure passwords, should be employed. Additionally, security awareness programs may also improve the security on both the user's and the employee's end, reducing the use of weak passwords and the likelihood of successful phishing campaigns [67].

Adversarial Tactics and Techniques Framework

For the application of more tailored security controls and for more efficiently mitigating breach risks, a threat intelligence program may be necessary [68]. To assist in this task, the MITRE ATT&CK, a knowledge base of adversary tactics and techniques, is a valid framework for threat modeling and the efficient employment of data breach countermeasures considering the observed malicious behavior [69].

This structure organizes the security knowledge into hierarchies. It presents several adversarial techniques categorized into 14 tactics, which are namely reconnaissance, resource development, initial access, execution, persistence, privilege escalation, defense evasion, credential access, discovery, lateral movement, collection, command and control, exfiltration, and impact. Each of these tactics comprise different techniques that may be used by attackers to achieve their goals.

In the initial access tactic, for example, examples of techniques include phishing (T1566), replication through removable media (T1091), and supply chain compromise (T1195). The security analyst may use the knowledge of these common tactics and techniques as guidelines for implementing security countermeasures as prevention of data breaches and also for identifying adversarial groups, who commonly use a specific set of techniques.

5. Data Protection Regulation

Given the sensitiveness of personal data, its collection, processing, storage, use, and destruction must be carefully planned. For enforcing this planning, data protection laws regulate the privacy of the data subjects [70].

5.1. Regulatory Measures for Mitigating Data Breaches

A data protection legislation aims to assist in mitigating data leaks by means that mandate the adoption of preventive security mechanisms. It also aims to reduce the impact caused by leaks through timely notifications to regulatory authorities, enabling appropriate actions to be taken [71]. In this latter scenario, established laws and regulatory frameworks assist in risk mitigation for both companies and the public sector. Regulations are commonly rooted in principles such as prevention, accountability, and transparency, providing measures to prevent breaches and means to identify and account the responsible parties. Thus, principled interpretations should already aid in reducing incidents or mitigating risks, even if a norm is not robust and lacks explicit provisions for data protection, processing, or breach notifications.

Conversely, even legislation with numerous extensive provisions, if not appropriately applied, will not be as effective as expected. Enforcement and the role of the regulatory

authority, in awareness, monitoring, and sanctioning, are fundamental for norm compliance and effectiveness. Each scenario involves a distinct interpretation and adaptation to the types of leaks occurring, making certain tools more suitable for certain countries. However, elements like notification and the impact report should be common guidelines across all regulations concerning data breaches.

In some countries, specific regulations exist for specific types of data. In the United States, for example, the Sarbanes–Oxley Act (SOX) regulates data security in publicly listed companies [72], the Gramm–Leach–Bliley Act (GLBA) in financial organizations, [73], the Federal Information Security Modernization Act (FISMA) in the U.S. federal government agencies [74], the Family Educational Rights and Privacy Act (FERPA) in education organizations [75], and the Children's Online Privacy Protection Act (COPPA) related to children [76]. However, this work focuses on federal-level data protection laws, and we do not discuss sectoral regulations such as the aforementioned.

As a reiteration of the importance of data protection legislation, Table 2 compares the distribution of data breach costs in lowly and highly regulated environments. The financial impacts of a security breach typically oscillate over time after the incident due to costs at different stages of the response, such as identification, containment, recovery, and repair [10]. A more evenly distributed cost, observed in highly regulated areas, will dilute the financial damage over time, facilitating the subsistence of the organization. In environments with low regulations, the cost is presented as a burst in the first year with a decreasing tendency.

Table 2. Comparison of the distribution of costs associated with data breaches in low- and high-regulation regions [10].

Years Since Breach	2023 Average	Low Regulation	High Regulation
1st	51%	64%	42%
2nd	31%	32%	37%
2+	18%	4%	21%

For a company located in a highly regulated region, compliance demonstration is of fundamental importance [36]. After a data compromise, in addition to the aforementioned costs, a company may incur sanctions if it fails to demonstrate compliance with the pertinent regulation. For example, 20% of breached companies paid at least USD 250,000 in fines [10].

5.2. Regulation Levels in the Breached Countries

To identify the regions that are lowly and highly regulated regarding data protection and privacy, we briefly compare aspects of the pertinent legislation of all the countries present within the dataset. Table 3 summarizes this comparison, in which the rows are sorted descending by the sum of records leaked. The column 'Regulation and enforcement' presents the level of regulation of the country, according to [77], in which (++) represents a heavy regulation, (+) a robust regulation, (−) moderate, and (−−) limited.

A possible analysis for revealing whether legislation is effective in incident mitigation is a comparison between the frequency and costs of data breaches before and after the law's enforcement, or even examining how the data protection authority acts in cases of occurring leaks, the existence of easy notification channels, assessing the impact that these leaks had on individuals and whether affected people were informed, and adhering to transparency and accountability duties. However, the dataset studied in this work is too short in time and does not provide enough information to conduct these analyses.

Therefore, our approach used to assess the effectiveness of these regulations is to compare key elements of the data protection norms scenario established in countries of different regulation and enforcement levels, namely heavy, robust, moderate, and limited.

The General Data Protection Regulation is in effect in Europe and considered a heavy regulation (++). It is considered a global standard [78] due to, for example, its enhanced

rights for individuals, stricter consent requirements, substantial penalties, and requirement of compliance demonstration.

It sets a maximum deadline of 72 h for reporting a personal data breach to the supervisory authority and to the data subject, as per Articles 33 and 34. The GDPR Recital 85 also outlines the necessity of adopting risk mitigation measures for physical, material, or immaterial damages to individuals, such as a loss of control over their personal data, limitation of their rights, discrimination, theft or usurpation of identity, financial losses, unauthorized reversal of pseudonymization, reputation damage, loss of confidentiality of personal data protected by professional secrecy, or any other significant economic or social disadvantage to individuals. Article 4, subparagraph 12, of the GDPR defines a data breach as a breach of security leading to accidental or unlawful destruction, loss, alteration, unauthorized disclosure, or access to transmitted, stored, or otherwise processed personal data. Therefore, under EU law, a leak would constitute a breach by definition.

Table 3. Summary of the data protection landscape in the breached countries [77], sorted descending by sum of breach sizes divided by population sizes in 2019.

Country	Regulation Level	Data Protection Law	Law Approved	Defines Personal Data	DPA	Requires Registration	Requires DPO	Breach Notification
CH	++	FADP	2020	✓	FDPIC	✗	✗	✓
EU	++	GDPR	2016	✓	EDPB	✗	✓	72 h
US	++	✗	–	✓	✗	✗	✗	✓
IL	++	PPL	1981	✓	PPA	✓	✓	immediately
KY	–	DPA	2017	✓	Ombudsman	✗	✗	5 days
AU	++	PA & APP	1988	✓	OAIC	✗	✗	72 h
NZ	+	PA	2020	✓	Privacy Commissioner	✗	✓	✓
SC	++	DPA	2003	✓	✗	✓	✗	✗
HK	++	PDPO	1995	✓	PCPD	✗	✗	✗
CN	++	PIPL	2021	✓	CAC	✗	✓	✓
AE	–	PDPL	2021	✓	✗	✗	✓	immediately
EC	–	PDPL	2021	✓	✗	✓	✗	5 days
BR	–	LGPD	2018	✓	ANPD	✗	✓	2 working days
IN	––	DPDP	2023	✓	✗	✗	✓	✓
CA	++	PIPEDA	2000	✓	OPC	✗	✓	✓
RU	–	DPA	2006	✓	Roskomnadzor	✓	✓	24 h
SG	++	PDPA	2012	✓	PDPC	✗	✓	3 calendar days
ID	+	PDP	2022	✓	PDP Agency	✗	✗	72 h
GB	++	UKGDPR	2018	✓	ICO	✓	✓	72 h
MY	+	PDPA	2010	✓	PDPC	✓	✗	✗
IR	––	✗	-	✗	✗	✗	✗	✗
CO	–	Law 1581	2012	✓	SIC & SOF	✓	✗	15 business days
JP	+	APPI	2003	✓	PPC	✗	✗	✓
CL	–	PDPL	1999	✓	✗	✗	✗	✗
PH	–	DPA	2012	✓	NPC	✓	✓	72 h
TR	–	LPPD	2016	✓	KVKK	✓	✗	72 h

The Japanese Law has been selected as an example of robust regulation (+). The Act on the Protection of Personal Information (APPI) was initially enacted in 2003 but was amended in 2017. In 2020, a bill was passed to further amend the APPI, which came into effect in 2022, integrating the public and private sectors, previously separate [77]. The established authority is the Personal Information Protection Commission (PPC). Under the amended APPI, commercial operators must report data breach incidents to the Personal Information Protection Commission and affected data subjects if the breach incident data could harm individuals' rights and interests. The PPC established a concrete threshold for communication obligations, where the business operator needs to report it to the PPC and notify the affected individuals. The defining elements of a breach are: (i) sensitive personal information leaked or likely to have leaked; (ii) personal information causing financial harm due to unauthorized use leaked or likely to have leaked; (iii) data leak due to malicious intent occurred or likely to have occurred; and (iv) a data leak incident involving over 1000 data subjects occurred or likely to have occurred. Additionally, the PPC's guidelines suggest that business operators conduct necessary investigations, take preventive measures, and disclose the nature of the breach and the corrective actions taken if appropriate and necessary.

The Brazilian General Data Protection Law (LGPD), considered a moderate regulation (−), deals with responsibility in Articles 31 and 32 and also with damages and reparation in Articles 42 to 45. The law addresses the reparation by the controller or processor that, due to the processing of personal data, causes patrimonial, moral, individual, or collective damages in violation of the legislation. Thus, non-compliance with good practices, security, and prevention as envisaged by the law would necessitate damage reparation in the event of a data leak. Article 48 specifically addresses the notification by the controller of security incidents to the national authority (ANPD), which may entail risks or relevant harm to data subjects. The provision outlines the minimum requirements: (i) description of the nature of the affected personal data; (ii) information about the involved data subjects; (iii) indication of technical and security measures used to protect the data, observing trade and industrial secrets; (iv) risks related to the incident; (v) reasons for delay if the communication was not immediate; and (vi) measures taken or to be taken to reverse or mitigate the damage effects. There is no legal deadline for incident notifications to the ANPD. However, the authority published guidelines in 2021 stating that communication should be made within two business days from the date of becoming aware of the incident. The institutional website contains instructions for this notification[3], with updated guidelines on violations. In the case of significant risk or harm to data subjects, individuals may also need to be notified. The notification can be sent by the Data Protection Officer (DPO) or the legal representative, along with the corresponding documentation or authorization. An additional recommendation, not legally required, is the implementation of contractual clauses establishing notification obligations between controllers and processors to expedite assessment and minimize risks to data subjects. Although it is not necessary to provide a list of affected data subjects, the ANPD may request the data controller to present a copy of the notification to the data subjects about the breach. This notification to the data subject should be made individually, whenever possible, and can be carried out by any means, such as e-mail, letter, or electronic message.

The Indian Law Personal Data Protection Bill in 2019 (PDP), still undergoing updates in 2022 and considered as a limited regulation (–), was enacted after the Indian Supreme Court in 2017 recognized privacy as a fundamental right, enshrined in Article 21 of the Constitution [77]. There is still no established data protection authority, which might compromise the law's applicability in data breach construction and mitigation.

As observed, both European and Japanese regulations encompass data protection measures concerning prevention, audit, and notification. They also ensure compliance and oversight of legal requirements through national data protection authorities. In contrast, Brazil, despite modeling its national law after the GDPR and anticipating similar provisions, remains in a relatively immature regulatory landscape. This is exemplified by the previous attachment of the national data protection authority to the Presidency of the Republic, which led to increased subordination and limited autonomy in law enforcement [79]. Only in 2022 did the ANPD gain more autonomy by attaching to the Brazilian Ministry of Justice and Public Security, albeit without subordination. Conversely, India's data protection law was enacted in 2023 and is yet to be fully implemented [80]. Moreover, the absence of a data protection authority underscores its significant immaturity. Consequently, despite the existence of laws and security requirements, the lack of effective enforcement, monitoring, and penalization mechanisms impedes the effective enforcement of data protection. These maturity issues are among the possible causes for the classification of Brazil's and India's data protection regulations at lower levels when compared to Europe's and Japan's, for example. The United States, despite also not having an enacted federal-level data protection law nor a data protection authority, is considered a country of heavy regulation; this may be due to the enactment of the sector-specific laws mentioned in Section 5 and the passing of state-level data protection laws [81], which demonstrate a greater maturity of the data protection regulation in the United States.

These differences in regulation levels in different regions may also arise from cultural differences. As an example, it has been observed that people from North America are more

willing to give up privacy than those from Europe, who are also more concerned about data breaches and transparency on how the data are used [82].

A comparison of the maturity levels between the geographical regions, expressed by the regulation and enforcement levels, is then presented in Figure 11. It is observed that all countries in Europe, North America, and Africa apply heavy regulations, while all countries in South America and the Caribbean are moderately regulated. Asia Pacific is the most diverse region, as it also comprehends a greater number of countries in the dataset.

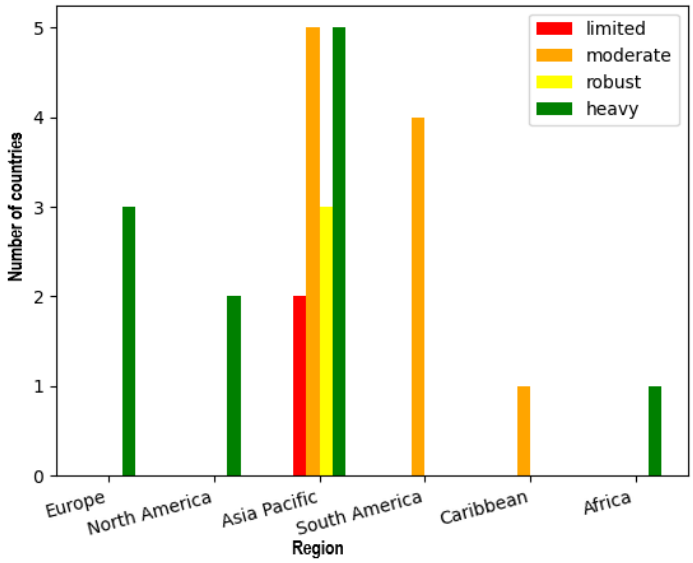

Figure 11. Classification level of regulation and enforcement for countries per region [77].

It is also noted that the Asia Pacific, the only region to have a country with limited regulation in the dataset, is also the region with a greater median of records breached per incident, according to Figure 5, with a close value to the median in Europe, a strongly regulated region. The North America, which is also strongly regulated, presents the lowest median value. When analyzing the median values in countries, from Figure 3, it is observed that the top countries are heavily regulated. From this, it may be inferred that regulation levels and breach sizes are not strongly correlated. In fact, the amount of data leaked is more strongly related to the amount of data stored by the organization. The nature of the stored data may also influence the breach size, as organizations may apply more efforts to secure more sensitive data, which could be more severely penalized by data protection laws. As an example, medical-related data presented the second lowest median values, as seen in Figure 6.

Following an analysis of Table 3, which is sorted by the sum of breach sizes divided by the populations sizes in 2019, it is observed that countries with greater regulation and enforcement levels are more concentrated in the superior lines, which suggests that these countries breach a greater number of records per inhabitant than those of lower levels. This may be due to several factors. One of them is that developed countries, like the United States and those in Europe, are more likely to offer more data-related services, such as cloud storage, even for non-residents, which increases the chances of breaches. It may also be a direct consequence of a more significant enforcement of breach notification. Nonetheless, it still evidences the need for more effective security controls to mitigate these occurrences.

5.3. Comparison of Regulatory Aspects

Data protection laws are regulations designed to safeguard the privacy and security of individuals' data. Whereas these laws differ in some aspects for different countries, general principles and legal basis include consent, purpose limitation, integrity, confidentiality, accountability, and transparency in data handling [83].

Although the majority of the countries within the scope of this study have enacted a data protection law, the United States and Iran, at the time of this writing, do not [77]. These countries do, however, have a pending bill or law initiative [84]. It is noteworthy that while Iran is classified as having limited regulation and enforcement regarding data protection and privacy, the United States is categorized as a heavily regulated country, possibly due to the reasons stated in Section 5.2.

Moreover, some Member States of the European Union had a national data protection law in effect before the General Data Protection Regulation, some of which still coexist with the GDPR. Examples include France's Loi Informatique et Libertés [85] and Germany's Bundesdatenschutzgesetz [86]. These Member States, however, abide by GDPR, and, for simplicity, we have grouped them by region in Table 3. The GDPR has influenced data protection regulations in regions other than Europe as a consequence of the Brussels effect [87]. A timeline of the years in which the data protection laws were enacted is presented in Figure 12. The years of enactment of the laws in the countries present in Table 3 are following [77], while for the countries that constitute the European Union, the dates are consonant with the Council of Europe[4].

Figure 12. Timeline of enactment of data protection laws.

5.3.1. Personal Data Definition

A clear definition of personal and sensitive data is necessary to ensure that organizations apply the appropriate legal safeguards to the relevant data and that individuals comprehend what data are protected, enhancing enforcement and transparency [88]. As an example, the GDPR defines personal data as any information relating to an identified or identifiable natural person, such as name, date of birth, email address, and billing address [77]. Brazil's LGPD does not classify anonymized data as personal information unless it can be reversed by applying reasonable efforts [77].

Sensitive data may also be defined. Japan's APPI, for example, defines them as any information that might cause the person to be discriminated against, such as race, medical history, and criminal record [77]. As seen in Table 3, Iran is the only country that does not define personal or sensitive information.

5.3.2. Data Protection Authority and Data Protection Officers

A Data Protection Authority (DPA) is a public entity responsible for supervising and enforcing data protection regulations and may also provide guidelines and raise awareness for data protection [89]. In Europe, the European Data Protection Board (EDPB), at the EU level, standardizes the data protection at each Member State, which also institutes federal DPAs, such as Italian Garante [90].

Some laws may require that any data controller intending to process personnel notify the public competent authority. In Table 3, this information is presented in the column 'requires registration'. Russia, for example, requires that the registration mentions, for instance, the full name and address of the data controller, the purpose of the processing, the categories of data being processed, protection measures being deployed, and the occurrence of the cross-border transfer of personal data [77].

The distribution of countries that have a DPA and require data controller registration among the regions is shown in Figure 13. Countries that appear as 'no' in the DPA requirement but 'yes' as registration have a legal prediction for establishing a Data Protection Authority and require registration but have not yet constituted it.

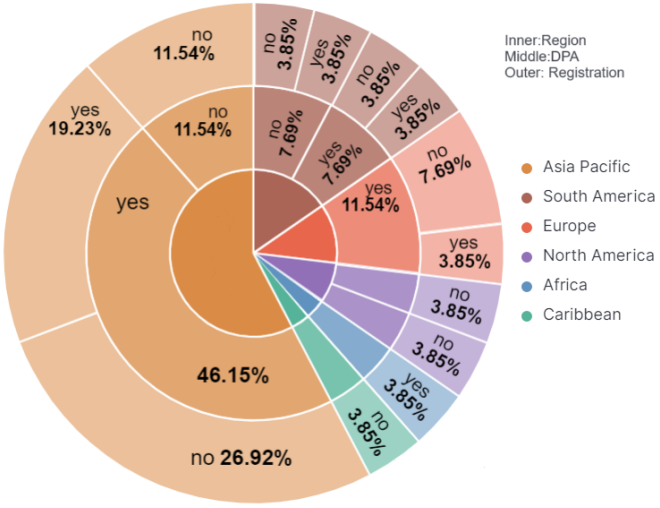

Figure 13. Distribution of countries that have a DPA and require registration.

On the organization's end, it may be required that companies indicate a point of contact for matters related to data protection. This role is identified as the Data Protection Officer, and some of its assignments include compliance monitoring and employer advisement [91]. In Canada, this position is by default occupied by the highest authority within the organization, and its responsibilities also include responding to and reporting security breaches [77]. The distribution of countries that require a DPO in companies is presented in Figure 14.

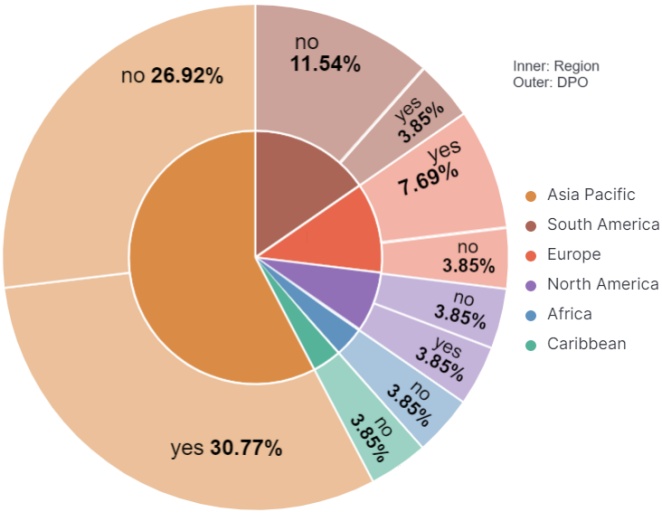

Figure 14. Distribution of countries that require a DPO.

5.3.3. Data Breaches Notification

The definition of a data breach may vary according to the laws. As an example, New Zealand defines it as any unauthorized or accidental access to, or disclosure, alteration, loss, or destruction of, personal information, or any action that prevents the agency from accessing the information on either a temporary or permanent basis [77].

The Indonesian law requires that any breach must be notified in written form within 72 h of becoming aware of the incident, and the notification must be directed to both the national authority and the affected users, including information such as the description of the breached data, when and how the incident occurred, and the efforts undertaken to mitigate it [77].

In Table 3, the column 'breach notification' specifies the notification requirements for each country. In it, the ✓ value indicates a law that requires notification but does

not specify a time limit. Figure 15 shows the distribution of countries that require data breach notification and, among those who do, whether they specify a time limit or not. It is also noteworthy that countries like Chile and Seychelles, although being among the three countries with the fewest records breached, do not require notification and, therefore, might be under-reporting incidents [92].

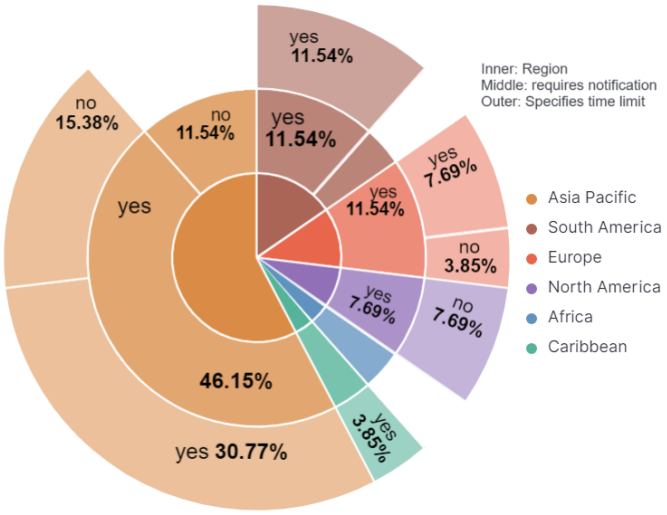

Figure 15. Distribution of countries that require breach notification and specify a time limit.

5.3.4. Other Aspects

Data protection is a complex matter, and we do not intend to present an exhaustive comparison of regulation in these countries but rather discuss some key elements of them. Hereof, the transfer requirements determine the legal aspects that an organization must comply with when on-shoring and offshoring data. Especially when transferring data abroad, additional obligations must be fulfilled. In this scenario, Israeli law mandates that the laws of the destination country provide a level of data protection no less stringent than that afforded by Israeli law [77]. If that is not the case, at least one of the other criteria must be met, such as the data subject's consent, whether the transfer is vital to public safety, or others [77].

For enforcement, the public competent entity may apply sanctions to non-compliant organizations, such as stopping data collection, destruction of personal data collected, and financial penalties in Singapore [77]. The Turkish Criminal Code also punishes a person who illegally collects, transfers, publishes, or deletes data with imprisonment [77].

Data protection laws may also apply to electronic marketing. One example of such an application is Malaysia's PDPA, which states that any data subject may require their data to cease or not to begin processing for direct marketing purposes [77].

6. Threats to Validity

The conclusions of this study face threats regarding their validity, which may be classified into four types [93].

External validity. The dataset may have been generated focusing on incidents that happened in specific regions, such as countries with greater economies or that may be more representative in the media. Less publicized breaches that meet the dataset criteria could have been omitted. Although the dataset incorporates sources from several languages [14], the language barrier also poses a potential external threat to validity. This limitation might result in less comprehensive coverage of less frequently spoken languages, leading to potential under-representation of breaches in those linguistic scenarios.

Internal validity. The remarked patterns might be specifically related to the limited time range encompassed by the dataset and not represent the general scenario of data breaches in the studied countries. Changes in data protection laws may also change patterns

of these incidents over time. Additionally, the casual relationship between regulation level in a country and the frequency and impact of a data leak within its jurisdiction is not always clear, as other factors, such as population size, may also influence the metrics on these incidents.

Construct validity. Countries have different definitions for personal data and what constitutes a data breach. This threatens the consistency of the incidents between regions. Moreover, a data protection regulation is rich in details and nuances that were not deeply explored, potentially affecting the comparability between the discussed regulatory aspects and the related data violations.

Conclusion validity. The dataset may suffer from under-reporting, especially in countries where data breach notification was not legally required between 2018 and 2019.

Despite these threats, this study provides valuable insights and promotes the discussion regarding compliance and data protection in the affected countries during the dataset time range.

7. Conclusions and Future Work

This research aimed to visualize worldwide data breaches, especially related to their geographical and economic sector occurrences. For that, we used a dataset comprising this type of incident that happened between 2018 and 2019 that leaked 30,000 records or more [14]. In addition to that, we provided a regulatory discussion in the scope of the affected countries.

This study explores the correlation between data protection enforcement levels and data breach occurrences and impacts, and also compares key elements between different levels of regulations, potentially guiding policy enhancements for improving global data protection. Furthermore, these findings can offer valuable insights for organizations adopting data storage solutions, such as cloud computing models, by aiding in informed decisions about where to store sensitive data, considering the nature of the data, the location of the storage, and its data protection regulation level and history of data confidentiality issues.

This investigation identified several strong relationships between countries and economic sectors affected by data breaches, such as India, and the medical or healthcare sector; and Brazil, and the government or military sector. We also observed that leaks affecting education records affected North America exclusively in this time frame. A correlation with the regulatory aspects in these countries and regions may promote the recognition of root causes for data violations and mitigation strategies. It is relevant to mention, however, that these findings heavily rely on the dataset used. Despite the seemingly robust methodology in data collection, the potential threats to its validity must be considered. These limitations might hinder the representation of reality.

Given that these patterns are subject to change over time, we propose future research to extend this study to encompass a broader time frame to validate the aforementioned discoveries. Future work may also enrich the dataset used, including a field that informs the cause of the data breach, which would enable a more in-depth discussion regarding mitigation techniques and the identification of patterns of adversarial procedures per geographical region. This enriched dataset could subsequently be fed into machine learning algorithms to uncover deeper correlations and predict future occurrences of data leaks. Future research may also focus on the countermeasures, providing more extensive discussion about security controls that may mitigate risks associated to data exposures. On the regulatory aspects, future work may also compare the impact of the enactment of data protection laws on security incidents by analyzing the occurrence of incidents before and after they started effects.

Overall, the research provides a valuable contribution to the understanding of data breaches and their implications for organizations and policymakers. The identified patterns and correlations can inform data security practices and regulatory frameworks, while the proposed future research can further advance our knowledge of this critical issue.

Author Contributions: G.A.P.R., A.L.M.S. and R.d.O.A. conceived the general composition of the data breach study; A.N.L.E.L. conducted the analysis on the data protection laws; E.D.C. and F.L.L.d.M. indicated the methods for the data exploratory analysis; A.L.S.O. and L.J.G.V. conceived and designed the experiment. All authors contributed equally to performing the conceptualization, formal analysis, investigation, validation, writing—original draft, and writing—review and editing the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the European Commission under the Horizon Europe Programme, as part of the project LAZARUS (Grant Agreement no. 101070303). The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: A publicly available dataset, published by [14], was analyzed in this study.

Acknowledgments: R.d.O.A. thanks: the Office of the General Attorney of the National Treasury of Brazil (PGFN 23106.148934/2019-67); the Union General Attorney Office of Brazil (AGU 697.935/2019); the Brazilian Federal Police (PF 03/2020); the CNPq–National Council for Scientific and Technological Development (PQ-2 312180/2019-5 in Cybersecurity and 465741/2014-2); the support of PPEE (calls N0430 and N044–Research Support | Process 23106.118956/2023-89); the Research Support Foundation of the Federal District–FAPDF (call Tech Learning–grant n.º 519/2023 and call Gov Learning–03/2022-Projectum Project).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANPD	National Data Protection Authority
APP	Australian Privacy Principles
APPI	Act on the Protection of Personal Information
CAC	Cyberspace Administration of China
DPA	Data Privacy Act / Data Protection Authority
DPDP	Digital Personal Data Protection
DPO	Data Protection Office
EDPB	European Data Protection Board
FADP	Federal Act on Data Protection
FDPIC	Federal Data Protection and Information Commissioner
ICO	Information Commissioner's Office
IoT	Internet of Things
GDPR	General Data Protection Regulation
KVKK	Kişisel Verileri Koruma Kurumu (Personal Data Protection Authority)
LGPD	General Data Protection Law
LPPD	Law on Protection of Personal Data
NPC	National Privacy Commission
OAIC	Office of the Australian Information Commissioner
OPC	Office of the Privacy Commissioner
PA	Privacy Act
PCPD	Privacy Commissioner for Personal Data
PDP	Personal Data Protection
PDPA	Personal Data Protection Act
PDPC	Personal Data Protection Commission
PDPL	Personal Data Protection Law
PDPO	Personal Data Privacy Ordinance
PIPEDA	Personal Information Protection and Electronic Documents Act
PIPL	Personal Information Protection Law
PPA	Privacy Protection Authority

PPC	Personal Information Protection Commission
PPL	Protection of Privacy Law
SIC	Superintendence of Industry and Commerce
SOF	Superintendence of Finance
UKGDPR	United Kingdom General Data Protection Regulation

Notes

[1] www.databreachdb.com/ accessed on 26 January 2024

[2] www.gov.br/saude/pt-br/acesso-a-informacao/lgpd/registro-de-incidentes-com-dados-pessoais accessed on 26 January 2024

[3] www.gov.br/anpd/pt-br/canais_atendimento/agente-de-tratamento/comunicado-de-incidente-de-seguranca-cis accessed on 26 January 2024

[4] www.coe.int/en/web/data-protection/ accessed on 26 January 2024

References

1. Diez-Olivan, A.; Del Ser, J.; Galar, D.; Sierra, B. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Inf. Fusion* **2019**, *50*, 92–111. [CrossRef]
2. Kovacova, M.; Machova, V.; Bennett, D. Immersive extended reality technologies, data visualization tools, and customer behavior analytics in the metaverse commerce. *J.-Self-Gov. Manag. Econ.* **2022**, *10*, 7–21.
3. Ogbuke, N.J.; Yusuf, Y.Y.; Dharma, K.; Mercangoz, B.A. Big data supply chain analytics: Ethical, privacy and security challenges posed to business, industries and society. *Prod. Plan. Control.* **2022**, *33*, 123–137. [CrossRef]
4. Bani Issa, W.; Al Akour, I.; Ibrahim, A.; Almarzouqi, A.; Abbas, S.; Hisham, F.; Griffiths, J. Privacy, confidentiality, security and patient safety concerns about electronic health records. *Int. Nurs. Rev.* **2020**, *67*, 218–230. [CrossRef] [PubMed]
5. Ileberi, E.; Sun, Y.; Wang, Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. *J. Big Data* **2022**, *9*, 1–17. [CrossRef]
6. Raghupathi, W.; Raghupathi, V.; Saharia, A. Analyzing Health Data Breaches: A Visual Analytics Approach. *AppliedMath* **2023**, *3*, 175–199. [CrossRef]
7. Perera, S.; Jin, X.; Maurushat, A.; Opoku, D.G.J. Factors affecting reputational damage to organisations due to cyberattacks. *Informatics* **2022**, *9*, 28. [CrossRef]
8. Duggineni, S. Impact of Controls on Data Integrity and Information Systems. *Sci. Technol.* **2023**, *13*, 29–35.
9. Foerderer, J.; Schuetz, S.W. Data breach announcements and stock market reactions: A matter of timing? *Manag. Sci.* **2022**, *68*, 7298–7322. [CrossRef]
10. IBM. *Cost of a Data Breach Report*; Technical Report; IBM Security: Armonk, NY, USA, 2023.
11. Zhang, X.; Yadollahi, M.M.; Dadkhah, S.; Isah, H.; Le, D.P.; Ghorbani, A.A. Data breach: Analysis, countermeasures and challenges. *Int. J. Inf. Comput. Secur.* **2022**, *19*, 402–442. [CrossRef]
12. Xue, Y.; Xue, K.; Gai, N.; Hong, J.; Wei, D.S.; Hong, P. An attribute-based controlled collaborative access control scheme for public cloud storage. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 2927–2942. [CrossRef]
13. Farsi, M.; Ali, M.; Shah, R.A.; Wagan, A.A.; Kharabsheh, R. Cloud computing and data security threats taxonomy: A review. *J. Intell. Fuzzy Syst.* **2020**, *38*, 2517–2527. [CrossRef]
14. Neto, N.N.; Madnick, S.; Paula, A.M.G.D.; Borges, N.M. Developing a global data breach database and the challenges encountered. *J. Data Inf. Qual. (JDIQ)* **2021**, *13*, 1–33. [CrossRef]
15. Vojković, G.; Milenković, M.; Katulić, T. IoT and Smart Home Data Breach Risks from the Perspective of Data Protection and Information Security Law. *Bus. Syst. Res. Int. J. Soc. Adv. Innov. Res. Econ.* **2020**, *11*, 167–185. [CrossRef]
16. Nemec Zlatolas, L.; Feher, N.; Hölbl, M. Security perception of IoT devices in smart homes. *J. Cybersecur. Priv.* **2022**, *2*, 65–73. [CrossRef]
17. Rejeb, A.; Rejeb, K.; Treiblmaier, H.; Appolloni, A.; Alghamdi, S.; Alhasawi, Y.; Iranmanesh, M. The Internet of Things (IoT) in healthcare: Taking stock and moving forward. *Internet Things* **2023**, *22*, 100721. [CrossRef]
18. Kiel, J.M. Data privacy and security in the US: HIPAA, hitech and beyond. In *Nursing Informatics: A Health Informatics, Interprofessional and Global Perspective*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 427–435.
19. Shahid, J.; Ahmad, R.; Kiani, A.K.; Ahmad, T.; Saeed, S.; Almuhaideb, A.M. Data protection and privacy of the internet of healthcare things (IoHTs). *Appl. Sci.* **2022**, *12*, 1927. [CrossRef]
20. Ugwu, A.O.; Gao, X.; Ugwu, J.O.; Chang, V. Ethical Implications of AI in Healthcare Data: A Case Study Using Healthcare Data Breaches from the US Department of Health and Human Services Breach Portal between 2009–2021. In Proceedings of the 2022 International Conference on Industrial IoT, Big Data and Supply Chain (IIoTBDSC), Beijing, China, 23–25 September 2022; IEEE: Piscataway NJ, USA, 2022; pp. 343–349.
21. Alfawzan, N.; Christen, M.; Spitale, G.; Biller-Andorno, N. Privacy, data sharing, and data security policies of women's mhealth apps: Scoping review and content analysis. *JMIR Mhealth Uhealth* **2022**, *10*, e33735. [CrossRef]

22. Utz, C.; Degeling, M.; Fahl, S.; Schaub, F.; Holz, T. (Un) informed consent: Studying GDPR consent notices in the field. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, London, UK, 11–15 November 2019; pp. 973–990.
23. Strupczewski, G. What Do We Know About Data Breaches? Empirical Evidence from the United States. In Proceedings of the Eurasian Economic Perspectives: Proceedings of the 23rd Eurasia Business and Economics Society Conference, Madrid, Spain, 27–29 September 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 281–299.
24. Saxena, R.; Gayathri, E. Cyber threat intelligence challenges: Leveraging blockchain intelligence with possible solution. *Mater. Today Proc.* **2022**, *51*, 682–689. [CrossRef]
25. Ibrahim, A.; Thiruvady, D.; Schneider, J.G.; Abdelrazek, M. The challenges of leveraging threat intelligence to stop data breaches. *Front. Comput. Sci.* **2020**, *2*, 36. [CrossRef]
26. Noor, U.; Anwar, Z.; Malik, A.W.; Khan, S.; Saleem, S. A machine learning framework for investigating data breaches based on semantic analysis of adversary's attack patterns in threat intelligence repositories. *Future Gener. Comput. Syst.* **2019**, *95*, 467–487. [CrossRef]
27. Rasoulian, S.; Grégoire, Y.; Legoux, R.; Sénécal, S. The effects of service crises and recovery resources on market reactions: An event study analysis on data breach announcements. *J. Serv. Res.* **2023**, *26*, 44–63. [CrossRef]
28. Wang, H.E.; Wang, Q.E.; Wu, W. Short selling surrounding data breach announcements. *Financ. Res. Lett.* **2022**, *47*, 102690. [CrossRef]
29. Adharsh, C.; Vijayalakshmi, S. Prevention of Data Breach by Machine Learning Techniques. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 28–29 April 2022; IEEE: Piscataway NJ, USA, 2022; pp. 1819–1823.
30. Turjeman, D.; Feinberg, F.M. When the data are out: Measuring behavioral changes following a data breach. *Mark. Sci.* **2023**, *ahead of print*.
31. Bachura, E.; Valecha, R.; Chen, R.; Rao, H.R. The Opm Data Breach: An Investigation of Shared Emotional Reactions on Twitter. *MIS Q.* **2022**, *46*, pp. 881–910 . [CrossRef]
32. Zou, Y.; Danino, S.; Sun, K.; Schaub, F. YouMight'Be Affected: An Empirical Analysis of Readability and Usability Issues in Data Breach Notifications. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glassglow, UK, 4–9 May 2019; pp. 1–14.
33. Alazab, M.; Hong, S.H.; Ng, J. Louder bark with no bite: Privacy protection through the regulation of mandatory data breach notification in Australia. *Future Gener. Comput. Syst.* **2021**, *116*, 22–29. [CrossRef]
34. Kesari, A. Do data breach notification laws reduce medical identity theft? Evidence from consumer complaints data. *J. Empir. Leg. Stud.* **2022**, *19*, 1222–1252. [CrossRef]
35. Pernot-Leplay, E. China's approach on data privacy law: A third way between the US and the EU? *Penn St. JL Int'l Aff.* **2020**, *8*, 49.
36. Chatterjee, C.; Sokol, D.D. Data security, data breaches, and compliance. In *Cambridge Handbook on Compliance*; Daniel Sokol, D., Rooij, B.v., Eds.; Cambridge University Press: Cambridge, UK, 2019.
37. Silva, J.; Calegari, N.; Gomes, E. After Brazil's general data protection law: Authorization in decentralized web applications. In Proceedings of the Companion Proceedings of the 2019 World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 819–822.
38. Ong, E.I. Singapore report: Data protection in the Internet. In *Data Protection in the Internet*; Springer: Cham, Switzerland, 2020; pp. 309–347.
39. Haber, E.; Tamò-Larrieux, A. Privacy and security by design: Comparing the EU and Israeli approaches to embedding privacy and security. *Comput. Law Secur. Rev.* **2020**, *37*, 105409. [CrossRef]
40. Yuan, B.; Li, J. The policy effect of the general data protection regulation (GDPR) on the digital public health sector in the european union: An empirical investigation. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1070. [CrossRef]
41. Mauri, M.; Elli, T.; Caviglia, G.; Uboldi, G.; Azzi, M. RAWGraphs: A visualisation platform to create open outputs. In Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter, Cagliari, Italy, 18–20 September 2017; pp. 1–5.
42. Granova, V.; Mashatan, A.; Turetken, O. Changing Hearts and Minds: The Role of Cybersecurity Champion Programs in Cybersecurity Culture. In Proceedings of the International Conference on Human-Computer Interaction, Copenhegen, Denmark, 23–28 July 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 416–428.
43. Wu, E. Sovereignty and data localization. In *Belfer Center for Science and International Affairs*; Harvard Kennedy School: Cambridge, MA, USA, 2021.
44. George, D.A.S.; George, A.H. Potential Risk: Hosting Cloud Services Outside the Country. *Int. J. Adv. Res. Comput. Commun. Eng.* **2022**, *11*, 5–11.
45. Sampson, D.; Chowdhury, M.M. The growing security concerns of cloud computing. In Proceedings of the 2021 IEEE International Conference on Electro Information Technology (EIT), Mt. Pleasant, MI, USA, 14–15 May 2021; IEEE: Piscataway NJ, USA, 2022; pp. 050–055.
46. Seh, A.H.; Zarour, M.; Alenezi, M.; Sarkar, A.K.; Agrawal, A.; Kumar, R.; Ahmad Khan, R. Healthcare data breaches: Insights and implications. *Healthcare* **2020**, *8*, 133. [CrossRef] [PubMed]
47. Koch, R. Hidden in the shadow: The dark web-a growing risk for military operations? In Proceedings of the 2019 11th International Conference on Cyber Conflict (CyCon), Tallinn, Estonia, 28–31 May 2019; IEEE: Piscataway NJ, USA, 2022; Volume 900, pp. 1–24.

48. Haber, M.J.; Chappell, B.; Hills, C. Regulatory compliance. In *Cloud Attack Vectors: Building Effective Cyber-Defense Strategies to Protect Cloud Resources*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 297–373.
49. McCoy, T.H.; Perlis, R.H. Temporal trends and characteristics of reportable health data breaches, 2010–2017. *JAMA* **2018**, *320*, 1282–1284. [CrossRef]
50. Churi, P.; Pawar, A.; Moreno-Guerrero, A.J. A comprehensive survey on data utility and privacy: Taking Indian healthcare system as a potential case study. *Inventions* **2021**, *6*, 45. [CrossRef]
51. Dhagarra, D.; Goswami, M.; Kumar, G. Impact of trust and privacy concerns on technology acceptance in healthcare: An Indian perspective. *Int. J. Med. Inform.* **2020**, *141*, 104164. [CrossRef] [PubMed]
52. Ferrão, S.É.R.; Carvalho, A.P.; Canedo, E.D.; Mota, A.P.B.; Costa, P.H.T.; Cerqueira, A.J. Diagnostic of data processing by brazilian organizations—a low compliance issue. *Information* **2021**, *12*, 168. [CrossRef]
53. Lima, R.C.; Silva, P.F.; Rudzit, G. No power vacuum: National security neglect and the defence sector in Brazil. *Def. Stud.* **2021**, *21*, 84–106. [CrossRef]
54. Shires, J. The Simulation of Scandal: Hack-and-Leak Operations, the Gulf States, and US Politics (Fall 2020). *Tex. Natl. Secur. Rev.* **2020**, *3*, 10–29.
55. Islam, R. *The Impact of Data Breaches on Stock Performance*; Glucksman Institute for Research in Securities Markets, Leonard N. Stern School of Business, New York University: New York, NY, USA, 2020.
56. Gopichandran, V.; Ganeshkumar, P.; Dash, S.; Ramasamy, A. Ethical challenges of digital health technologies: Aadhaar, India. *Bull. World Health Organ.* **2020**, *98*, 277. [CrossRef]
57. Bondre, A.; Pathare, S.; Naslund, J.A. Protecting mental health data privacy in India: The case of data linkage with Aadhaar. *Glob. Heal. Sci. Pract.* **2021**, *9*, 467–480. [CrossRef]
58. Mali, N.V.; Avila-Maravilla, M.A. Convergence or Conflict? Digital Identities vs. Citizenship Rights: Case Study of Unique Identification Number, Aadhaar, in India. In Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance, Galway, Ireland, 4–6 April 2018; pp. 443–448.
59. Tiwari, P.R.; Agarwal, D.; Jain, P.; Dasgupta, S.; Datta, P.; Reddy, V.; Gupta, D. India's "Aadhaar" Biometric ID: Structure, Security, and Vulnerabilities. In Proceedings of the International Conference on Financial Cryptography and Data Security, Grenada, Spain, 2–6 May 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 672–693.
60. Tyagi, A.K.; Rekha, G.; Sreenath, N. Is your privacy safe with Aadhaar?: An open discussion. In Proceedings of the 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan, India, 20–22 December 2018; IEEE: Piscataway NJ, USA, 2022; pp. 318–323.
61. Bella, G.; Biondi, P.; Tudisco, G. A double assessment of privacy risks aboard top-selling cars. *Automot. Innov.* **2023**, *6*, 146–163. [CrossRef]
62. Peacher, H.B. Regulating Data Privacy of Connected Vehicles: How Automotive Giants Can Protect Themselves and Their Golden Goose. *Alb. LJ Sci. Tech.* **2020**, *30*, 74.
63. Song, Y.; Xu, C.; Zhang, Y.; Li, S. Hardening Password-Based Credential Databases. *IEEE Trans. Inf. Forensics Secur.* **2023**, *19*, 469–484. [CrossRef]
64. Ometov, A.; Petrov, V.; Bezzateev, S.; Andreev, S.; Koucheryavy, Y.; Gerla, M. Challenges of multi-factor authentication for securing advanced IoT applications. *IEEE Netw.* **2019**, *33*, 82–88. [CrossRef]
65. Thomas, K.; Pullman, J.; Yeo, K.; Raghunathan, A.; Kelley, P.G.; Invernizzi, L.; Benko, B.; Pietraszek, T.; Patel, S.; Boneh, D.; et al. Protecting accounts from credential stuffing with password breach alerting. In Proceedings of the 28th USENIX Security Symposium (USENIX Security 19), Santa Clara, CA, USA, 14–16 August 2019; pp. 1556–1571.
66. Trautman, L.J.; Hussein, M.T.; Opara, E.U.; Molesky, M.J.; Rahman, S. Posted: No Phishing. *Emory Corp. Gov. Account. Rev.* **2021**, *8*, 41–74. [CrossRef]
67. Zwilling, M.; Klien, G.; Lesjak, D.; Wiechetek, Ł.; Cetin, F.; Basim, H.N. Cyber security awareness, knowledge and behavior: A comparative study. *J. Comput. Inf. Syst.* **2022**, *62*, 82–97.
68. Wagner, T.D.; Mahbub, K.; Palomar, E.; Abdallah, A.E. Cyber threat intelligence sharing: Survey and research directions. *Comput. Secur.* **2019**, *87*, 101589. [CrossRef]
69. Xiong, W.; Legrand, E.; Åberg, O.; Lagerström, R. Cyber security threat modeling based on the MITRE Enterprise ATT&CK Matrix. *Softw. Syst. Model.* **2022**, *21*, 157–177.
70. Hoofnagle, C.J.; Van Der Sloot, B.; Borgesius, F.Z. The European Union general data protection regulation: What it is and what it means. *Inf. Commun. Technol. Law* **2019**, *28*, 65–98. [CrossRef]
71. Shastri, S.; Wasserman, M.; Chidambaram, V. The seven sins of {Personal-Data} processing systems under {GDPR}. In Proceedings of the 11th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 19), Renton, WA, USA, 8 July 2019.
72. Sebastian, G. Could incorporating cybersecurity reporting into SOX have prevented most data breaches at US publicly traded companies? An exploratory study. *Int. Cybersecur. Law Rev.* **2022**, *3*, 367–383. [CrossRef]
73. Pang, M.S.; Tanriverdi, H. Strategic roles of IT modernization and cloud migration in reducing cybersecurity risks of organizations: The case of US federal government. *J. Strateg. Inf. Syst.* **2022**, *31*, 101707. [CrossRef]
74. Ryle, P.; Yan, J.; Gardiner, L.R. Gramm-Leach-Bliley gets a systems upgrade: What the ftc's proposed safeguards rule changes mean for small and medium american financial institutions. *EDPACS* **2022**, *65*, 6–17. [CrossRef]

75. Cohen, B.; Hu, A.; Patino, D.; Coffman, J. Educational Data in the Cloud Legal Implications and Technical Recommendations. In Proceedings of the 2022 IEEE/ACM 15th International Conference on Utility and Cloud Computing (UCC), Vancouver, WA, USA, 6–9 December 2022; IEEE: Piscataway NJ, USA, 2022; pp. 181–182.
76. Skowronski, D.S. Coppa and educational technologies: The need for additional online privacy protections for students. *Ga. State Univ. Law Rev.* **2022**, *38*, 12.
77. Piper, D. *Data Protection Laws of the World Full Handbook*; Technical Report; DLA Piper, London, UK, 2023.
78. Mantelero, A. The future of data protection: Gold standard vs. global standard. *Comput. Law Secur. Rev.* **2021**, *40*, 105500. [CrossRef]
79. Bezerra Sales Sarlet, G.; Piñeiro Rodriguez, D. Alternatives for an adequate structuring of the national data protection authority (ANPD) in its independent profile: Proposals to overcome the technological challenges in the age of digital governance. *Int. Cybersecur. Law Rev.* **2023**, *4*, 197–211. [CrossRef] [PubMed]
80. Srinivasan, S.; Sinha, V.; Modi, S. Drafting a pro-antitrust and data protection regulatory framework. *Indian Public Policy Rev.* **2023**, *4*, 35–56. [CrossRef]
81. Hartzog, W.; Richards, N. Privacy's constitutional moment and the limits of data protection. *BCL Rev.* **2020**, *61*, 1687.
82. Sheth, S.; Kaiser, G.; Maalej, W. Us and them: A study of privacy requirements across North America, Asia, and Europe. In Proceedings of the 36th International Conference on Software Engineering, Hyderabad, India, 31 May–7 June 2014; pp. 859–870.
83. Demetzou, K.; Zanfir-Fortuna, G.; Vale, S.B. The thin red line: Refocusing data protection law on ADM, a global perspective with lessons from case-law. *Comput. Law Secur. Rev.* **2023**, *49*, 105806. [CrossRef]
84. Banisar, D. National Comprehensive Data Protection/Privacy Laws and Bills 2023. *Priv. Laws Bills* **2023**. [CrossRef]
85. Demotes-Mainard, J.; Cornu, C.; Guerin, A.; Bertoye, P.H.; Boidin, R.; Bureau, S.; Chrétien, J.M.; Delval, C.; Deplanque, D.; Dubray, C.; et al. How the new European data protection regulation affects clinical research and recommendations? *Therapies* **2019**, *74*, 31–42. [CrossRef]
86. Etteldorf, C. Germany Revisited: The Second Data Protection Adaption and Implementation Act. *Eur. Data Prot. L. Rev.* **2019**, *5*, 397. [CrossRef]
87. Mahieu, R.; Asghari, H.; Parsons, C.; van Hoboken, J.; Crete-Nishihata, M.; Hilts, A.; Anstis, S. Measuring the Brussels Effect through Access Requests: Has the European General Data Protection Regulation Influenced the Data Protection Rights of Canadian Citizens? *J. Inf. Policy* **2021**, *11*, 301–349. [CrossRef]
88. Finck, M.; Pallas, F. They who must not be identified—Distinguishing personal from non-personal data under the GDPR. *Int. Data Priv. Law* **2020**, *10*, 11–36. [CrossRef]
89. Sevinç, İ.; Karabulut, N. A review on the personal data protection authority of turkey. *Akad. Hassasiyetler* **2020**, *7*, 449–472.
90. Botta, M.; Wiedemann, K. The interaction of EU competition, consumer, and data protection law in the digital economy: The regulatory dilemma in the Facebook odyssey. *Antitrust Bull.* **2019**, *64*, 428–446. [CrossRef]
91. Ciclosi, F.; Massacci, F. The data protection officer: A ubiquitous role that no one really knows. *IEEE Secur. Priv.* **2022**, *21*, 66–77. [CrossRef]
92. Amir, E.; Levi, S.; Livne, T. Do firms underreport information on cyber-attacks? Evidence from capital markets. *Rev. Account. Stud.* **2018**, *23*, 1177–1206. [CrossRef]
93. Wohlin, C.; Runeson, P.; Höst, M.; Ohlsson, M.C.; Regnell, B.; Wesslén, A. *Experimentation in Software Engineering*; Springer Science & Business Media: Cham, Switzerland, 2012.